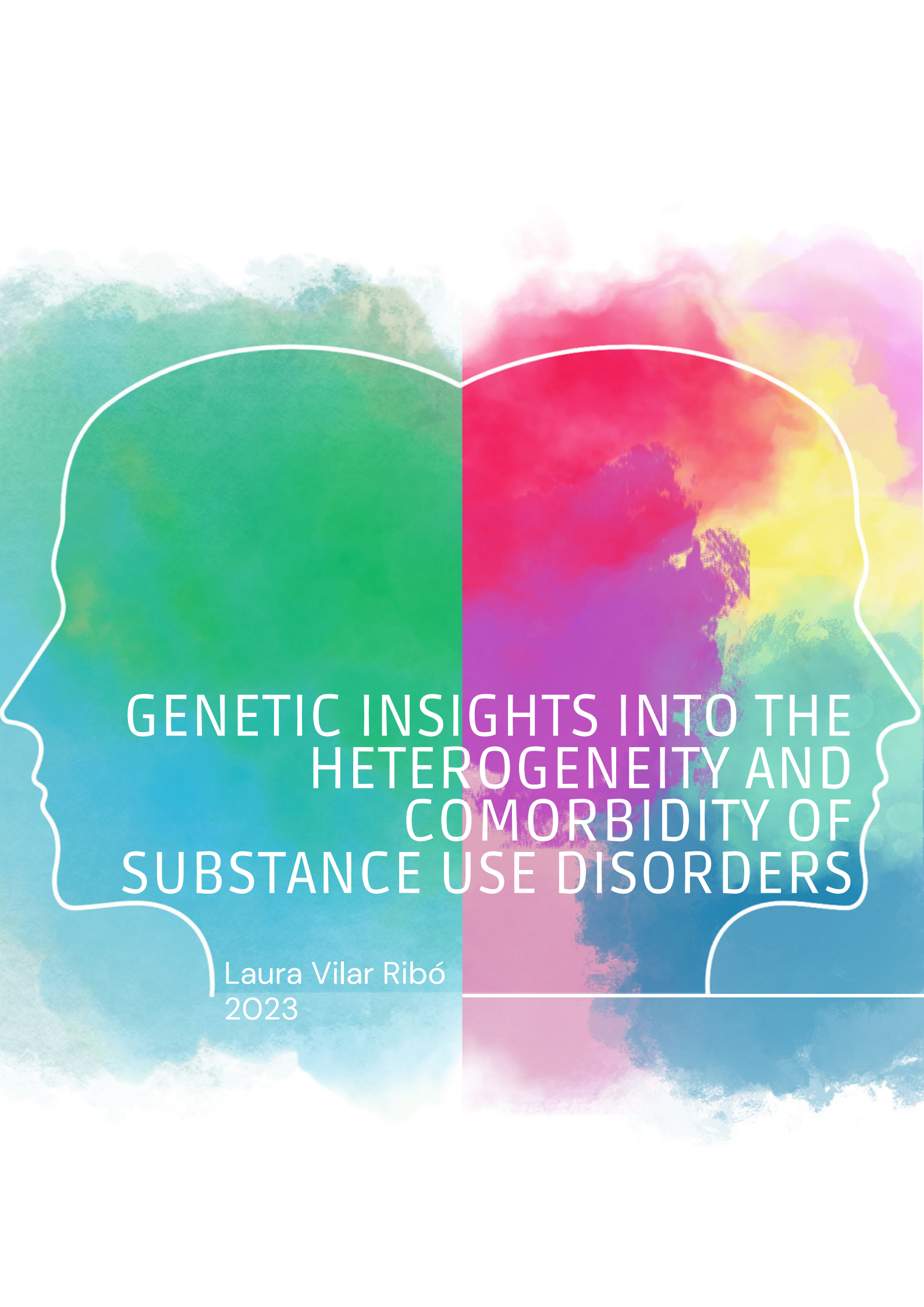
UNIVERSITAT DE BARCELONA

Genetic Insights into the Heterogeneity and Comorbidity of Substance Use Disorders

Laura Vilar Ribó



GENETIC INSIGHTS INTO THE HETEROGENEITY AND COMORBIDITY OF SUBSTANCE USE DISORDERS

Laura Vilar Ribó
2023

Genetic Insights into the Heterogeneity and Comorbidity of Substance Use Disorders

Memòria presentada per **Laura Vilar Ribó**

Per optar al grau de

Doctora per la Universitat de Barcelona

Programa de Doctorat en Genètica

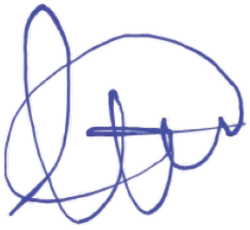
Tesi realitzada al grup de Psiquiatria, Salut mental i Addiccions de l'Institut de Recerca Vall D'Hebron, sota la direcció de la Dra. Marta Ribasés Haro i la Dra. Maria Soler Artigas, i la tutoria del Dr. Bru Cormand Rifà.

Dra. Marta Ribasés Haro
Codirectora

Dra. María Soler Artigas
Codirectora

Dr. Bru Cormand Rifà
Tutor

Laura Vilar Ribó
Doctoranda



Barcelona, Setembre del 2023

If you change the way you look at things,
the things you look at change.

Wayne Dyer

ACKNOWLEDGMENTS

Aquests línies representen una oportunitat única per expressar la meva profunda gratitud cap a les persones que han jugat un paper essencial en el recorregut acadèmic i personal d'una de les etapes mes significatives de la meva vida. Avui, quan m'apropo al final, m'adono que aquest viatge ha estat ple de reptes i triomfs, i que no hauria estat possible sense el suport, la inspiració i la dedicació de moltes persones excepcionals. A través d'aquestes paraules espero poder transmetre el meu profund agraïment, ja que tots i cadascú de vosaltres representa una part d'aquesta historia, i la vostra influencia i amor han donat forma a la persona que sóc avui.

En primer lloc, vull expressar el meu agraïment cap a tots els participants dels estudis. La vostre generositat i compromís al compartir les vostres dades és la base que promou tota la nostra investigació, sense vosaltres no hagués estat possible. A més, al darrera d'aquesta feina hi ha un equip excepcional de psiquiatres, psicòlegs, psicòlogues, infermers i infermeres que són clau en el procés de reclutament de pacients i recollida de mostres. La vostre dedicació i col·laboració amb la recerca és profundament apreciada, i entre tots formem un equip integra. A Constanza, por ponerle tanto corazón y dedicación. Gracias por resolver todas mis dudas, por estar horas repasando la base de datos interminable conmigo, tu duro trabajo ha sido indispensable. Al Josep Antoni Ramos-Quiroga, cap del departament i mentor. Gràcies per sempre promoure la investigació amb passió i determinació. El teu compromís amb en la recerca científica ha portat a l'excel·lència el departament de psiquiatria de Vall d'Hebron. Ha sigut un honor formar part de teu equip. Al Bru, gràcies per sempre implicar-te en la meva feina, per donar-me suport i paraules d'ànim. T'admiro molt i ha sigut un plaer tenir-te com a tutor.

A les meves directores de tesis, la Marta i la María. A la Marta, gràcies per la teva confiança en mi i per obrir-me les portes al teu grup. Has estat una guia inspiradora des del moment que vaig decidir llençar-me a la piscina de la recerca. La teva calidesa, naturalitat i determinació que em vas mostrar des del moment en què et vaig conèixer, i que m'ha acompanyat al llarg d'aquests anys, van confirmar que fer la tesi amb tu era una bona decisió. Gràcies per preguntar-me "estàs contenta?" al final de les reunions, i que, semblarà una pregunta simple, però reflecteix el teu interès genuí de que tothom es sentí còmode i satisfet. Gràcies per ensenyar-me la importància de la resiliència i de mai tirar la tovallola quan les coses es torcen. Gràcies per sempre estar disposada a buscar un pla i a ajudar-me a afrontar els obstacles amb determinació. M'has donat la confiança per arribar fins aquí i m'has mostrat que podia comptar amb tu en tot moment. Gràcies també per la teva alegria contagiosa i la teva capacitat de celebrar al màxim cada èxit, perquè amb brownie o lemon pie els triomfs encara són més dolços.

A María, gracias porque contigo he podido aprender lo que no se aprende en los libros. Me has enseñado que siempre hay que hacer pruebecitas antes de dar un gran paso. "Pruébalo con 10 SNPs, ahora con 100, después con 1000. Perfecto, ya puedes ir

con todo". Me has enseñado a ser meticulosa, a investigar con conciencia y con propósito. De ti he aprendido que las cosas bien hechas se tienen que hacer con calma. Gracias por saber escuchar, por saberte poner en la piel de los demás y por saber que decir para subirme el ánimo en momentos difíciles. Gracias por tu infinita paciencia en enseñarme R, mendelian y, literalmente, todo lo demás. Uno de los grandes retos de esta tesis ha sido tener que aprender lenguaje de programación, y sin tu ayuda hubiera sido imposible. Lo he dicho antes, lo vuelvo a decir ahora y lo repetiré siempre: todos los laboratorios necesitan a una María.

A més, he tingut la gran sort de treballar rodejada de persones intel·ligents, treballadores i perseverants. Algunes ja no hi són, i algunes just acaben d'arribar, però de totes m'emporto un record molt bonic. A la Cris, perquè tot i que el nostre temps al laboratori juntes va ser curt em va encantar contagiar-me de la teva incansable energia. La teva dedicació cap a totes les activitats que envoltaven el laboratori era realment admirable, i la teva absència es va notar des del dia què vas marxar. A Lorena, porque entramos en el grupo prácticamente a la vez y aprendimos de la mano. Tantos años juntas me han mostrado que eres una persona con un corazón enorme, que te preocupas por los demás y que estas siempre dispuesta a echar una mano. Eres una persona llena de inquietudes, con ganas incansables de aprender y de crecer. Espero que tengas mucha suerte en tu nueva aventura y que siempre sigas luchando por todo aquello que anhelas. A la Pauli, quina sort haver-te conegut. Des de que vaig arribar al laboratori, sense tu saber-ho, em vas motivar a ser la millor versió de mi mateixa, a interessar-me, a aprendre i a demostrar el que era capaç de fer. És innegable que vaig aprendre molt de tu dins del laboratori, però el que m'emporto de veritat i el que és únic és el que vaig aprendre de tu fora de la feina. Com tu mateixa vas dir, vam treballar juntes però ens vam conèixer més tard. Tens una llum que és única, una energia que atrau i que fa que la gent vulgui estar al teu costat, perquè il·lumines. La passió i dedicació que li poses a tot el que et proposes és una virtut espectacular i no tinc cap dubte que aconseguiràs tot al que li posis el teu cor en aquesta vida. Desitjo de veritat que, sigui com sigui que avancin les nostres vides, sempre seguim connectades. Gràcies per ser una font d'inspiració, la meva yogi personal y per creure en mi.

A la Natalia, perquè treballar al teu costat és una alegria. Gràcies per ajudar-me sempre que ho necessitava, encara que et fes anar justa per agafar el bus, i per sempre tenir una actitud positiva i emprenedora. Admiro moltíssim la teva manera de treballar, al·lucino amb tot el que ets capaç de fer amb R i no tinc cap dubte que tindràs una tesis espectacular. A la Silvia i la Judit, gràcies per inspirar-me cada dia amb la vostre passió per la investigació. Gràcies per ensenyar-me que l'èxit requereix esforç, determinació i també cor. Les vostres incansables ganes de continuar aprenent i millorant m'han servit de guia per saber quin tipus de postdoc vull arribar a ser. Gràcies també per sempre mostrar interès en la meva feina, perquè el vostre coneixement m'ha ajudat a millorar i perquè he pogut aprendre moltíssim amb vosaltres. Al Pau i a la Valeria, les

incorporacions estel·lars al laboratori. Pau, ser el primer noi en un laboratori ple de noies ja és un gran èxit, però el èxit és nostre de poder comptar amb tu en el grup. Has vingut una energia i unes ganes de treballar que et portaran molt i molt lluny. Valeria, me ha encantado poder coincidir contigo aunque haya sido breve. Llenas el laboratorio con tu sonrisa y te deseo lo mejor allí donde decidas llevar tu vida.

A la Maite, gràcies per ser una de les millors sorpreses que m'ha portat la etapa en el VHIR. La nostre amistat es va fer forta de cop i m'encanta saber que encara ens queden moltes moments per viure juntes. Gràcies per deixar-me entrar a la teva vida, per ser la meva *california girl* i per compartir moments amb mi que mai oblidaré. No puc expressar les ganes que tinc de tornar a prop teu i d'explorar juntes a on ens portarà la vida. A la Laura, la Sandra, la Paula a l'Anabel i a la Ingrid, gràcies per fer que els dies post-pandèmia fossin molt més portables, perquè inclús quan el VHIR estava buit, fèieu que valgues la pena venir per compartir dinars, riures i penes.

A la Sandra, gràcies infinites per confiar cegament amb mi i no dubtar en donar-me la oportunitat de treballar amb tu. Perquè des del primer moment et vas implicar al 100% en el nostre projecte juntes, em vas fer sentir valorada i em vas incloure com una més dins del teu equip. Gràcies per contagiar-me de la teva energia i de les teves ganes de treballar. Gràcies per convidar-me a casa teva quan vaig arribar, per presentar-me a la teva família, perquè això em va fer saber que, passes el que passes a tants quilometres de casa, no estaria sola. Gràcies per donar-me la oportunitat i la confiança de continuar treballant amb tu, m'emociona saber que, al tancar aquesta etapa, en començaré una de nova amb molta força i energia en el teu equip.

Gràcies a les amistats que sempre han estat presents, de manera incondicional, a les amistats que si han marxat ho han fet per tornar més fortes que mai, i a les amistats noves que, sense jo saber-ho, eren exactament el que necessitava. Al càmping, per ser el nostre petit paradís, perquè sense aquest lloc i aquesta gent estic segura que jo no seria qui sóc avui en dia. Berta, Ari, Lidia, Marta, Carles, Sergi, Joan, gràcies per ser l'alegria dels meus estius, per omplir-me d'amor i felicitat sempre que us tinc al costat i per haver construït junts una família. Us porto a la pell, i allà on vagi us portaré amb mi. A la Leire, perquè em llegeixes com si fos un llibre obert i perquè tens un cor enorme, gràcies per fer-me saber que sempre puc comptar amb tu. Al guillem, perquè ets alegria i llum, gràcies per fer-me veure el món des del teus ulls, perquè ho fas tot molt més bonic. A la Sesca, perquè em sento afortunada de poder dir que soc la teva amiga i perquè t'estimo incondicionalment, gràcies per ser la font d'energia que m'ha ajudat a tirar endavant en els moments més difícils. A Estixu, porque si existe amistad a primera vista, está fuiste tú, gracias por ser la tata que siempre había querido. A la Marta, la Sara, les Moniques, la Laura i l'Ari, perquè després de tants anys i tantes aventures no us heu cansat de mi, i això diu moltíssim de lo forta que és la nostre amistat. Gràcies perquè estar amb vosaltres sempre em treu un somriure, sou un regal. A Bailén 149, perquè aquest pis té vida i

tothom que hi entra ho percep. A Meri y a Pietro, porqué poder compartir estas cuatro paredes con vosotros durante el final de esta etapa me ha mantenido un poco más cuerda. Porqué incluso cuando los días son malos, teneros en casa y poder hablar, o mejor dicho cotillear, con vosotros, me levanta el ánimo. A tots vosaltres, perquè al·lucino amb lo extraordinaris que sou i amb la sort que tinc de poder dir que sou els meus amics, gràcies, us estimo.

A San Diego, gràcies per arribar just quan ho necessitava i per donar-me molt més del que podria haver imaginat. Qui m'ho hagués dit que, al perdrem, aconseguiria trobar-me de nou en un lloc tant lluny però que, a la vegada, es va tornar casa tant ràpid. To all the wonderful people that I met there, because you are a family and you made a spot for me, and I could never thank you enough. Because my heart broke a little when I left and now I am counting the days until I come back. Thank you for turning San Diego my new home.

A la meva família, perquè no tinc paraules per expressar lo molt que us estimo. Perquè no exagero quan dic que no hagués arribat aquí si no fos per vosaltres. A la mama i al papa, m'heu ensenyat a ser forta, a ser valenta, a lluitar i a no tirar la tovallola, però també a ser pacient, a parar i a pensar, i sempre actuar des del cor. Gràcies per creure en mi, sobretot quan jo he deixat de fer-ho. Gràcies per dir-me sempre, no el que vull, sinó el que necessito sentir, encara que no sigui bonic. Gràcies per lluitar per mi, per estimar-me i per ensenyar-me a estimar. Gràcies per deixar-me se qui soc sense jutjar-me, però sempre guiant-me com la llum d'un far quan m'he perdut. Gràcies per demostrar-me que els meus pares també poden ser els meus amics, perquè us ho puc explicar tot, lo bo, lo dolent, lo vergonyós y lo dolorós. Gràcies per donar-me ales i deixar que voli, perquè se que no és fàcil tenir que veure'm marxar. Al meu germà, Pau, gracies per ser la persona més indispensable de la meva vida, per sempre sempre sempre donar-me la teva mà quan he necessitat ajuda, per dir-me sempre les veritats més dures, encara que fessin mal. Gràcies per deixar-me ser la teva confident, per mostrar-te vulnerable i deixar que jo també sigui la teva roca. Sempre has sigut la meva guia, perquè t'estimo, però sobretot, perquè t'admiro moltíssim.

No puc evitar que em caiguin les llàgrimes mentre escric aquestes pàgines, però si ho fan és de pura felicitat, perquè em sento increïblement afortunada, i és gràcies a tots vosaltres. Recordo que algú, ja buscaré el nom, va dir una vegada que tothom hauria d'escriure un llibre al llarg de la teva vida. Doncs bé, aquí teniu el meu.

ABSTRACT

Substance use disorders (SUDs) are psychiatric disorders characterized by a recurring desire to continue taking a substance regardless of its destructive consequences. The etiology of SUDs is complex and multifactorial, where both genetic and environmental factors have an impact on the disease development. In addition, SUDs often co-occur at high prevalence with other psychiatric conditions, significantly impacting life expectancy, disease severity and societal burden. Over the past decade, genome-wide association studies (GWASs) have identified various risk loci for substance-specific SUD, as well as a shared genetic vulnerability for addiction. In addition, post-GWAS analyses have helped unravel the complex genetic architecture of SUDs, which can also involve an interplay of gene-environment interactions, and its relationship with comorbid mental health conditions. Current research in this field is making collective efforts to provide deeper and clearer knowledge into the genetic and environmental factors involved into the co-occurrence of SUDs and psychiatric disorders, which may be partially driving the high heterogeneity observed in SUDs, and the biological mechanisms driving these relationships.

The present thesis comprises two studies that leverage in-house clinical cohorts, with both phenotypical and genetic data available, and state-of-the-art genomic techniques to investigate the shared genetic liability between SUDs and co-occurring traits, and to shed light into the genetic underpinnings of SUDs heterogeneity.

The first study particularly focused on the relationship between SUDs and attention-deficit and hyperactivity disorder (ADHD). In this study, we tested whether the genetic liability to five SUD-related phenotypes share a common genetic background in both the general population and clinically diagnosed ADHD individuals, using an in-house sample of 989 subjects and polygenic scores (PGSs) analyses. We further explored the genetic overlap and the causal relationship between ADHD and SUDs using genetic correlation and Mendelian randomization analyses. Our results confirmed a significant genetic correlation between ADHD and SUDs and supported the current literature on the causal effect of the genetic liability to ADHD on the risk for SUDs. We provided novel findings on the effect of the genetic liability to lifetime cannabis use on an increased risk

for ADHD and found evidence of a shared genetic background underlying SUDs between general population and ADHD, at least for lifetime cannabis use, alcohol dependence and smoking initiation.

The second study aimed to disentangle SUDs heterogeneity using multidimensional data from a deeply phenotyped SUDs cohort of 1,427 individuals and PGSs for comorbid psychiatric disorders, behavioral and other related traits. We systematically explored the associations between the PGSs and 39 SUD-related phenotypes, and performed PGSs-environment interaction analyses using information on lifetime emotional, physical and/or sexual abuse. Our results revealed different patterns of associations between the genetic liability for mental health-related traits and SUD-related phenotypes, which may help explain part of the heterogeneity observed in SUDs. We also found evidence of a PGS-environment interaction showing that genetic liability for suicide attempt worsened the psychiatric status in SUDs individuals with a history of emotional physical and/or sexual abuse.

Overall, the results of the present thesis provide new insights into the genetic overlap and causal relationships between SUDs and ADHD and contribute to a better understanding of the role of the genetic liability for psychiatric disorders and related traits, as well as its interaction with adverse life experiences, in the complexity of SUD heterogeneity. Lastly, this thesis provides a general discussion of the findings, which offers an extensive interpretation of the results in the context of existing literature, discusses the main methodological implications and outlines prospective directions for advancing in this line of research.

RESUM

Els trastorns per l'ús de substàncies (TUS) són trastorns psiquiàtrics caracteritzats per un desig recurrent de continuar prenent una o diverses substàncies, independentment de les seves conseqüències destructives. L'etiologia dels TUS és complexa i multifactorial, on tant factors genètics com ambientals tenen un impacte en el desenvolupament de la malaltia. A més, els TUS sovint es presenten simultàniament amb altres trastorns psiquiàtrics, afectant significativament la severitat de la malaltia, l'esperança de vida i la càrrega en la societat. Durant l'última dècada, els estudis d'associació del genoma complet (GWASs) han identificat diverses variants genètiques de risc per a TUS de substàncies específiques, així com una vulnerabilitat genètica compartida per a l'addicció. A més, les anàlisis post-GWAS han ajudat a desxifrar l'arquitectura genètica complexa dels TUS, que també pot implicar la interacció entre gens i ambient, i la seva relació amb trastorns de salut mental comòrbids. La recerca actual en aquest camp està focalitzada en profunditzar en el coneixement sobre els factors genètics i ambientals involucrats en la coexistència del TUS i trastorns psiquiàtrics, el qual pot ser parcialment responsable de l'alta heterogeneïtat observada en el TUS, i els mecanismes biològics implicats.

La present tesi està composta per dos estudis que utilitzen cohorts clíniques, amb dades fenotípiques i genètiques disponibles, i tècniques genòmiques actuals per explorar la carga genètica compartida entre els TUS i els trets comòrbids, i per investigar la heterogeneïtat dels TUS des del punt de vista genètic.

El primer estudi es centra particularment en la relació entre els TUS i el trastorn per dèficit d'atenció i hiperactivitat (TDAH). En aquest estudi, vam testar si la càrrega genètica per a cinc fenotips de TUS comparteixen una base genètica comuna en la població general i en individus amb TDAH, fent servir un mostra interna de 989 individus i anàlisis de puntuacions poligèniques (PGSs). Seguidament, vam explorar el solapament genètic i la relació causal entre el TDAH i els TUS utilitzant anàlisis de correlació genètica i de randomització mendeliana. Els nostres resultats confirmen una base genètic comuna entre el TDAH i els TUS i donen suport a la literatura actual sobre l'efecte causal de la càrrega genètica pel TDAH en el risc de TUS. A més, descrivim per primera vegada l'efecte causal de la càrrega genètica per a l'ús de cànnabis en el risc de TDAH i trobem evidències

d'un component genètic compartit subjacent als TUS en la població general i en els individus amb TDAH, almenys per a l'ús de cànnabis, la dependència a l'alcohol i l'inici del consum de tabac.

El segon estudi té com a objectiu desxifrar la heterogeneïtat dels TUS utilitzant dades multidimensionals d'una cohort de TUS de 1,427 individus dels quals es disposa una àmplia informació fenotípica, i PGSs per a trastorns psiquiàtrics comòrbids, trets del comportament i altres trets relacionats. Vam explorar les associacions entre els PGSs i 39 fenotips de TUS, i vam portar a terme anàlisis d'interacció PGS-ambient utilitzant informació sobre abús emocional, físic i/o sexual al llarg de la vida. Els nostres resultats revelen diferents patrons d'associacions entre la càrrega genètica per a trets relacionats amb la salut mental i fenotips de TUS, el que pot ajudar a explicar part de la heterogeneïtat observada en els TUS. També trobem evidència d'una interacció PGS-ambient que mostra que la càrrega genètica per a intents de suïcidi empitjora l'estat psiquiàtric en individus amb TUS que han patit abús emocional, físic i/o sexual.

En conjunt, els resultats de la present tesi aporten noves perspectives sobre el solapament genètic i les relacions causals entre els TUS i el TDAH i contribueixen a una millor comprensió del paper de la càrrega genètica pels trastorns psiquiàtrics i trets relacionats, així com la seva interacció amb experiències adverses al llarg de la vida, en la complexitat de la heterogeneïtat dels TUS. Finalment, aquesta tesi ofereix una discussió general, la qual proporciona una extensa interpretació dels resultats en el context de la literatura existent, discuteix les principals implicacions metodològiques i detalla les futures direccions per avançar en aquesta línia de investigació.

TABLE OF CONTENTS

LIST OF TABLES, FIGUES AND BOXES

Tables

Figures

Boxes

ABBREVIATIONS AND ACRONYMS

1KG	1000 Genomes
Addiction-rf	The Addiction-Risk-Factor
ADH1B	Alcohol Dehydrogenase 1B
ADHD	Attention-Deficit/Hyperactivity Disorder
ALDH2	Aldehyde Dehydrogenase 2
ATP2C2	Calcium-transporting ATPase
AUDIT	Alcohol Use Disorders Identification Test
BEND4	BEN Domain Containing 4
CADM2	Cell Adhesion Molecule 2
CAUSE	Causal Analysis Using Summary Effect Estimates
CHRNA2	Cholinergic Receptor Nicotinic α2 Subunit
CHRNA3– CHRNA5–CHRNB4	Neuronal Acetylcholine Receptor gene cluster
CNIH3	Cornichon Family AMPA Receptor Auxiliary Protein 3
CNNM2	Cyclin and CBS Domain Divalent Metal Cation Transport Mediator 2
CNV	Copy Number variants
CSMD1	CUB and Sushi Multiple Domains 1
DBH	Dopamine β -Hydroxylase
DNMT3B	DNA Methyltransferase
DRD2	Dopamine Receptor D2
DSM	Diagnostic and Statistical Manual of Mental Disorders
EPHX2	Epoxide Hydrolase 2
EuropASI	European version of the Addiction Severity Index
FAM53B	Sequence Similarity 53, Member B
FOXP2	Forkhead Box Protein P2
FTND	Fagerström Test for Nicotine Dependence
FURIN	Paired Basic Amino Acid Cleaving Enzyme
GABA	γ-aminobutyric acid
GBD	Global Burden Of Disease
GCKR	Glucokinase Regulator

GSCAN	GWAS and Sequencing Consortium of Alcohol and Nicotine use
gSEM	Genomic Structural Equation Modeling
GWAS	Genome-Wide Association Studies
GWASMA	GWAS Meta-Analysis
GWS	Genome-Wide Significance
GxE	Gene-environment interaction
h^2	Heritability
h^2_{SNP}	SNP-based Heritability
HRC	Haplotype Reference Consortium
HWE	Hardy-Weinberg Equilibrium
ICD	International Classification Of Diseases
JCAD	Junctional Cadherin 5 Associated
KLB	β-Klotho
LCV	Latent Causal Variant
LD	Linkage Disequilibrium
LPHN2	Latrophilin 2
MAF	Minor Allele Frequency
MR	Mendelian Randomization
MVP	Million Veteran Program
NCAM1	Neural Cell Adhesion Molecule 1
NEGR1	Neuronal Growth Regulator 1
OCD	Obsessive Compulsive Disorder
PCA	Principal Component Analysis
PDE4B	Phosphodiesterase 4B
PGC	Psychiatric Genomics Consortium
PGS	Genome-Wide Polygenic Score
PheWAS	Phenome-Wide Association Studies
PPP1R1B	Phosphatase 1 Regulatory Subunit 1B
PRS	Polygenic Risk Score
PTSD	Post-Traumatic Stress Disorder
RABEPK	Rab9 Effector Protein With Kelch Motifs

rg	Genetic correlation
rGE	Gene-environment correlation
RGMA	Repulsive Guidance Molecule BMP Coreceptor A
SAMHSA	Substance Abuse and Mental Health Services Administration
SLC30A9	Solute Carrier Family 30 Member 9
SLC39A13	Solute Carrier Family 39 Member 13
SLC39A8	Solute Carrier Family 39 Member 8
SLC6A4	Serotonin Transporter
SNP	Single-Nucleotide Polymorphism
SSADDA	Semi-structured Assessment for Drug Dependence and Alcoholism
TENM2	Teneurin Transmembrane Protein 2
TMEM51	Transmembrane Protein 51
TRAK2	Trafficking Kinesin-Binding Protein 2
UNODC	United Nations Office on Drugs and Crime
WDR	The World Drug Report
WHH	World Mental Health

GLOSSARY

Glossary from section 1. Introduction to SUDs

Allostasis. Process by which the body responds to challenges to maintain apparent homeostasis through changes in brain reward and stress mechanisms.

Compulsivity. Preservative, repetitive actions that are excessive and inappropriate. Compulsivity dominates at later stages of drug addiction through the emergence of negative emotional states in the withdrawal/negative affect stage and anticipation of the drug in the preoccupation/anticipation stage.

Impulsivity. Predisposition toward rapid, unplanned reactions to internal or external stimuli without regard for negative consequences of these reactions to one-self or others. Impulsivity often dominates at the early stages of drug addiction through repeated binge/intoxication and positive reinforcement.

Negative reinforcement. Process by which removal of an aversive stimulus or state increases the probability of a response.

Physiological dependence. Is a state that develops as a result of the adaptation (tolerance) produced by a resetting of homeostatic mechanisms in response to repeated substance use.

Positive reinforcement. Process by which presentation of a stimulus increases the probability of a response.

Tolerance. A state in which a substance produces a diminishing biological or behavioral response in an individual, which occurs when the drug is used repeatedly and the body adapts to the continued presence of the drug.

Withdrawal. Physical, cognitive, and affective symptoms that occur when the chronic use of a substance is reduced abruptly or stopped among individuals who have developed tolerance to a drug. These symptoms are characteristic for a given category of drugs and tend to be the opposite to the original effects produced by the drug before tolerance development

Glossary form section 3. Etiology of SUDs

Bonferroni correction. A method to adjust for multiple testing that consists of dividing the significance threshold by the number of independent tests carried out.

Causal genetic variant. Genetic variant that carries the functional allele influencing disease susceptibility and explaining the observed association with the phenotype of interest

Collider bias. A bias that occurs when two variables (A and B) both influence a third variable (C), and the third variable is used to condition on. This can induce spurious correlations between variables A and B.

Confounding factor. A confounding variable is a third variable that influences the two variables of interest in a model assessing the causal effect of an exposure on an outcome.

Discovery sample (in the context of PGS studies). Group or population used in the initial phase of the polygenic score calculation workflow to identify the genetic variants associated with a particular trait or disease.

Genetic correlation (rg). Refers to the proportion of variance that two traits share due to genetic causes.

Genome-wide association study (GWAS). Research approach used to identify genomic variants that are statistically associated with a risk for a disease or a particular trait.

Genome-wide polygenic score (PGS). Quantitative metric of an individual's inherited liability for a trait based on the cumulative impact of many common genetic variants.

Haplotype. A set of genetic variants (or SNPs) that tend to be inherited together. A haplotype can refer to a combination of alleles or to a set of genetic variants found on the same chromosome.

Hardy-Weinberg Equilibrium (HWE). The HWE principle states that in a large, randomly mating population with no evolutionary forces acting upon it (such as mutation, migration, selection, or genetic drift), the frequencies of alleles and genotypes will remain constant from one generation to the next. If a population is in HWE, the observed genotype frequencies will match the expected genotype frequencies based on the allele frequencies.

Heritability (h_2). The proportion of variation in a phenotype due to genetic factors; traditionally measured using family or twin studies. It is a population-level measure that can vary across time and environments.

Heterozygous. It refers to the presence of two different alleles at a given SNP within an individual.

Liability (for a disease). Refers to a non-observable and continuous variable that reflects an individual's genetic and environmental susceptibility to a certain disease.

Linkage disequilibrium (LD). The phenomenon wherein nearby genetic variants are inherited non-independently of each other. LD measures degree to which an allele of one SNP is observed with an allele of another SNP within a population.

Loci. Specific positions or locations on the genome where alleles of different genetic variants are situated

Mendelian randomization. Research method that provides evidence about putative causal relations between risk factors and outcomes, using genetic variants as instrument variables.

Principal component analysis (PCA). Statistical method commonly used in population genetics to identify structure in the distribution of genetic variation across geographical location and ethnic background.

Summary statistics. Output file form a GWAS containing association data and p-values for every variant analyzed.

Minor Allele Frequency (MAF). The frequency of the less common of two alleles for a genetic variant in a population (with two alleles carried by each person), ranging from >0 to ≤0.5. For example, a genetic variant with a minor allele (G) frequency of 0.4 implies that 40% of that population has the G allele versus the more common allele (the major allele), which is found in 60% of the population.

Phenome-Wide Association Study (PheWAS). Genomic methodology that investigates the association between genetic information for a trait (in the form of single SNPs or the aggregation of SNPs in PGSs) and a wide range of phenotypes or traits, allowing for a comprehensive exploration of the pleiotropic effects of genetic variants across phenotypes.

Pleiotropy. Phenomenon where a single genetic variant influences multiple (seemingly related or unrelated) phenotypic traits. Pleiotropy plays a significant role in complex traits and diseases, where multiple genetic and environmental factors contribute to their development.

Population stratification. The presence of the systematic allele frequency differences observed in a population as a consequence of ancestry, which can act as a confounding factor and result in false-positive associations.

Single-nucleotide polymorphisms (**SNPs**). Single base-pair changes in the DNA sequence, which occur with high frequency in the human genome.

SNP-based heritability (h^2_{SNP}). Estimate of the additive genetic variance in a phenotype that can be explained by common SNPs.

Target sample (in the context of PGS studies)**.** Group or population of individuals for whom the PGS is being calculated.

1 INTRODUCTION

1. Introduction to SUDs

Substance use disorders (SUDs) are psychiatric conditions characterized by a cluster of cognitive, behavioral, and physiological symptoms, including impaired control, social impairment, risky use, and pharmacological changes indicating that the individual continues using the substance or substances despite its significant substance-related problems. The core feature of SUDs is an underlying alteration in the structure and function of the brain, known as neuroadaptations, that progresses as the individual continues to misuse the substance. As a consequence of these alterations, normal brain function becomes impaired, which ultimately drives the transition from controlled regular use to harmful and compulsive misuse. Moreover, these neuroadaptations can persist long after the individual quits substance use, and, as a result, may experience persistent cravings for the substance, which ultimately may lead to relapse (Volkow et al., 2019).

Initial or experimental substance use is commonly associated with *impulsivity*, where the individual engages in substance use without fully considering the potential consequences. When the experience is positive, this sensation produced by the substance *positively reinforces* the act of substance use, increasing the likelihood of repeated use. As the pattern of substance use continues and the positive reinforcing effects diminish, the individual develops *tolerance*, which leads to the consumption of the substance in greater amounts and/or more frequently. Over time, the absence of the substance will produce *withdrawal* symptoms, characterized by negative emotions such as stress, anxiety, depression and physical sickness. Consequently, the individual will transition from impulsive to *compulsive* use with the aim to alleviate these negative feelings, known as *negative reinforcement* (Koob & Le Moal, 2001).

1.1. History of Addiction and The Diagnostic and Statistical Manual of Mental Disorders

Before scientific psychiatry emerged, individuals with addiction were commonly seen as moral transgressors or sinners, and their behavior was often attributed to the will of gods (Nathan et al., 2016). However, in the late 1700s, physicians like Benjamin Rush in the United States and Thomas Trotter in England began to conceptualize addiction as a medical disease rather than a matter of bad character. Their observations of patients and

their families led them to document the medical and psychiatric consequences of excessive alcohol use, the progressive nature of the disease, its intergenerational impact, the persistence of cravings, and the inability to control alcohol consumption despite repeated attempts (Olsen, 2022). Their contributions shed light on the involvement of biological factors in addiction and marked a significant shift in understanding addiction as a medical condition.

The development of the Diagnostic and Statistical Manual of Mental Disorders (DSM-I) in 1952 was a crucial milestone in establishing a comprehensive psychiatric nomenclature (Nathan et al., 2016) (**Figure 1**). However, the initial edition lacked sufficient detail to enable reliable and replicable diagnoses, as it included only 106 disorders. Both the DSM-I and its successor, the DSM-II, categorized alcoholism and drug addiction as personality disorders under sociopathic personality disturbance (**Figure 1**). Addiction was viewed as a symptom of an underlying personality disorder, characterized by maladjustment, neurotic character traits, and emotional immaturity. Individuals with alcoholism were seen as engaging consistently and purposefully in uncontrolled and compulsive drinking behaviors (Nathan et al., 2016).

In 1980, the third edition of the DSM (DSM-III) represented a significant advancement in diagnostic reliability by providing a comprehensive and detailed summary of signs, symptoms, and operational criteria based on empirical evidence (American Psychiatric Association, 1980). Within this framework, SUDs were detailed in a separate section, with clear distinctions among substance use, substance abuse, and substance dependence. Substance abuse was defined by a pattern of pathological use, impairment in social or occupational functioning, and at least one month of use. Substance dependence, a more severe form of a SUD, required one or more signs of *physiological dependence*, such as tolerance or withdrawal (**Figure 1**). Some substances, like cocaine and hallucinogens, were primarily associated with symptoms of abuse and not considered to cause dependence.

In 1983, articles highlighting the role of genetics in alcoholism emerged, suggesting a heritable physiological predisposition to alcohol response (Mayer, 1983). The fourth edition of the DSM (DSM-IV) in 1994 embraced a biological perspective on addiction, renaming the chapter to "substance-related disorders" and grouping substances into 11

distinct classes (American Psychiatric Association, 1994). The distinction between substance abuse and substance dependence remained, with abuse characterized by maladaptive and repeated patterns of use leading to adverse consequences, while dependence involved cognitive, behavioral, and physiological symptoms indicating the continued use of a substance despite substance-related problems (**Figure 1, Box 1**).

The latest edition of the DSM (DSM-5) in 2013 expanded the substance-related and addictive disorders chapter to cover 10 substance classes, gambling disorders, and substance-induced disorders (American Psychiatric Association, 2013). The DSM-5 emphasized the central role of brain reward mechanisms in addiction onset and maintenance, maintaining the biological/genetic model of addiction. Notably, the DSM-5 eliminated the distinction between substance abuse and substance dependence, replacing it with a dimensional assessment of substance use severity based on the number of observed symptoms (**Figure 1, Box 1**). However, it faced criticism for over-diagnosing and overpromising the development of biomarkers and dimensional diagnosis, potentially influenced by large drug companies (G. Young, 2016).

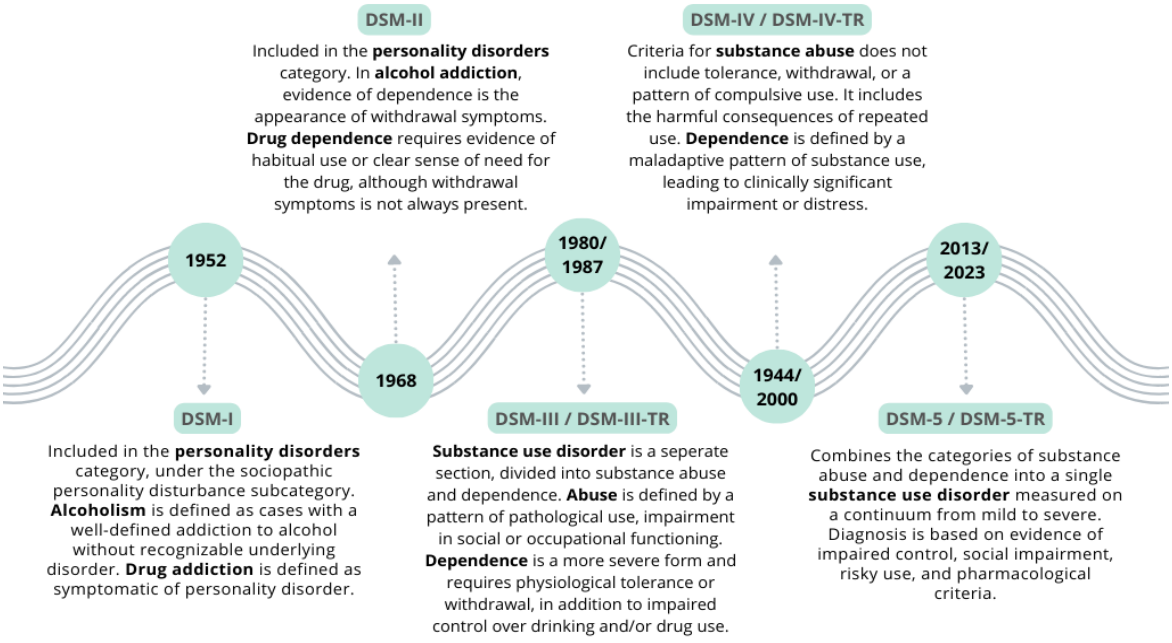


Figure 1. Chronology of SUDs diagnosis in the five editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM)

Both the DSM-IV and DSM-5 hold significance in the context of the current thesis. Box 1 provides a side-by-side comparison of the diagnostic criteria for SUDs used in both editions (**Box 1**).

Box 1. DSM-IV and DSM-5 criteria for SUDs

DSM-IV

Substance Abuse
One (or more) of the following, occurring within a 12-month period:

1. Recurrent substance use resulting in a failure to fulfill major role obligations at work, school, or home
2. Recurrent substance use in situations in which it is physically hazardous
3. Recurrent substance-related legal problems
4. Continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance

Substance Dependence
Three (or more) of the following, occurring any time in a 12-month period:

1. Tolerance, as defined by either of the following:
   a. a need for markedly increased amounts of the substance to achieve intoxication or desired effect, or
   b. markedly diminished effect with continued use of the same amount of the substance
2. Withdrawal, as manifested by either of the following:
   a. the characteristic withdrawal syndrome for the substance, or
   b. the same (or closely related) substance is taken to relieve or avoid withdrawal symptoms
3. The substance is often taken in larger amounts or over a longer period than intended
4. There is a persistent desire or unsuccessful efforts to cut down or control substance use
5. A great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover from its effects
6. Important social, occupational, or recreational activities are given up or reduced because of substance use
7. Continued substance use despite knowledge of having a persistent physical or psychological problem that is likely to have been caused by or exacerbated by the substance

DSM-5

Two (or more) of the following, occurring within a 12-month period:

1. The substance is often taken in larger amounts or over a longer period than intended
2. There is a persistent desire or unsuccessful efforts to cut down or control substance use
3. A great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover from its effects
4. Craving for the drug
5. Recurrent substance use resulting in a failure to fulfill major role obligations at work, school, or home
6. Continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance
7. Important activities are given up or reduced because of drug use
8. Recurrent substance use in situations in which it is physically hazardous
9. Continued substance use despite knowledge of having a persistent physical or psychological problem that is likely to have been caused by or exacerbated by the substance
10. Tolerance, as defined by either of the following:
    a. a need for markedly increased amounts of the substance to achieve intoxication or desired effect, or
    b. markedly diminished effect with continued use of the same amount of the substance
11. Withdrawal, as manifested by either of the following:
    a. the characteristic withdrawal syndrome for the substance, or
    b. the same (or closely related) substance is taken to relieve or avoid withdrawal symptoms

Mild: Presence of 2-3 symptoms
Moderate: Presence of 4-5 symptoms
Severe: Presence of 6 or more symptoms

1.2. Current Prevalence of SUDs

The World Drug Report (WDR), an annual publication by the United Nations Office on Drugs and Crime (UNODC), informed in 2021 that approximately 296 million individuals aged 15-64 worldwide, which accounts for 5.8% of the global population, had engaged in illicit substance use within the past 12 months. This represents a notable 23% increase over the previous decade, partly attributed to population growth (UNODC, 2023). Cannabis is the most used substance in the vast majority of countries, with a prevalence of 4.3% of the global adult population, followed by opioids, with a prevalence of 1.18% (UNODC, 2023) (**Figure 2A**).

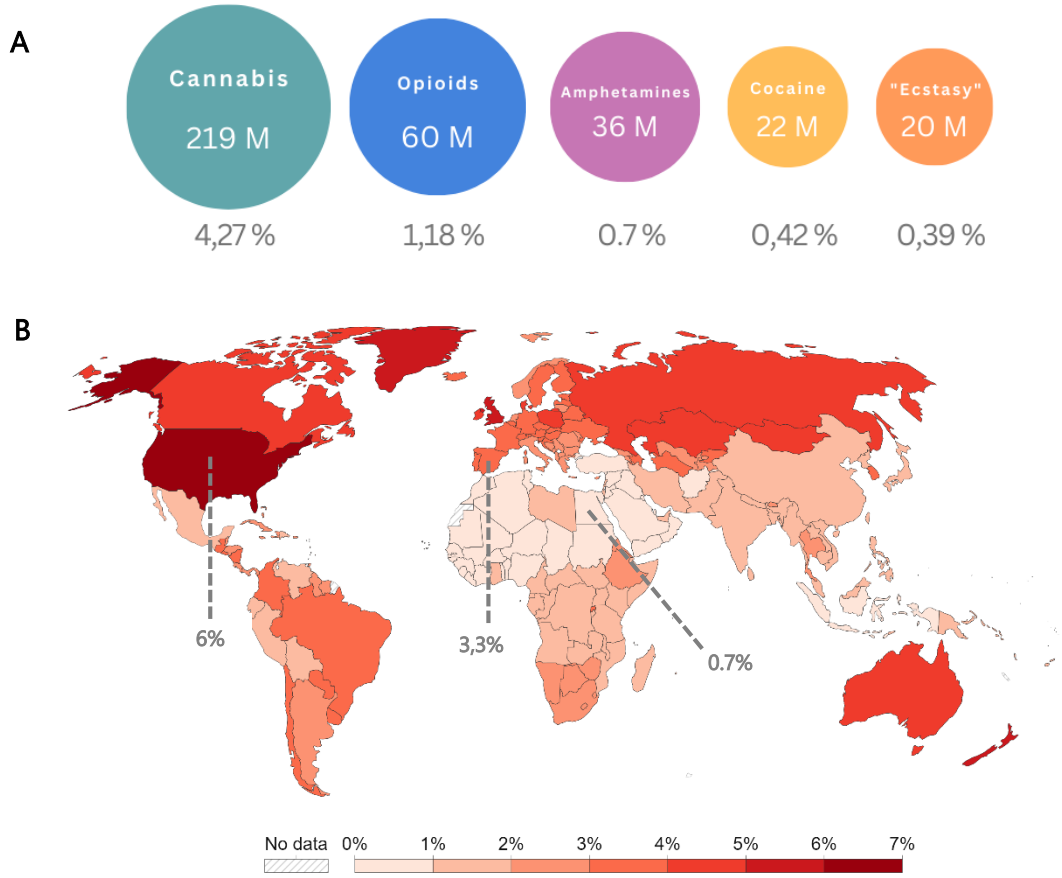


Figure 2. Annual prevalence of substance use and SUDs. (**A**) Global estimates of number of substance users in millions (number inside the circle) and prevalence of substance use (percentage below de circle) of selected substance groups (2021). (**B**) Share of the population with alcohol or substance use disorders (2019). *Sources: World drug report 2023, IHME, global burden of disease 2019 and our world in data.*

Furthermore, an estimated 13.6% of substance users (around 39.5 million individuals) suffer from an illicit SUD, which corresponds to a prevalence of 0.76% among the global

population aged 15-64 (Degenhardt, Bharat, Glantz, Sampson, Scott, et al., 2019; UNODC, 2023). Another well-documented source, the Global Burden of disease (GBD), which also accounts for the prevalence of alcohol use disorder, reported in 2019 that over 2% of the global population were dependent on alcohol or an illicit substance (Hannah Ritchie et al., 2022).

Nevertheless, substantial differences in SUDs prevalence are observed across different regions. Higher-income countries, such as the United States of America (USA) and certain Eastern European nations, exhibit a higher overall SUDs prevalence, with approximately 4-6% of the population dependent on alcohol or illicit substances. In contrast, lower-middle-income countries, including several African nations, show lower SUDs prevalence, of around 0.7% (Hannah Ritchie et al., 2022) (Figure 2B). Additionally, regional differences extend to the preferred type of substance. For instance, in the USA, SUDs prevalence is dominated by illicit substances, whereas in regions such as Russia and Eastern Europe, alcoholism is more prevalent (Degenhardt, Bharat, Glantz, Sampson, Scott, et al., 2019).

Historically, both substance use and SUDs have exhibited higher prevalence rates among men compared to women (Rehm & Shield, 2019; UNODC, 2023). This difference has mainly been attributed to gender-related cultural and environmental factors rather than sex-related biological factors. Fundamentally, men have had more access to substances than women (McHugh et al., 2018). However, it is important to note that women engage in the use of certain substance types nearly as frequently as men, such as non-medical pharmaceutical opioids, sedatives, and tranquilizers (Figure 3). Moreover, women tend to increase the amount of consumption more rapidly than men and, therefore, typically show an accelerated progression from the initiation of substance use to the development of a SUD (Fonseca et al., 2021). Nevertheless, recent epidemiologic studies suggest that the differences in prevalence rates between genders might be getting narrower (Castelpietra et al., 2022). This may be attributed to the fact that women are exposed to negative life events, such as trauma and intimate partner violence, which is directly related with the rapid development of SUDs and contributes to narrowing the gap in substance use between genders (Fonseca et al., 2021). Despite that, women remain

underrepresented in SUDs treatment, likely due to social stigma or fear of legal sanctions (UNODC, 2023).

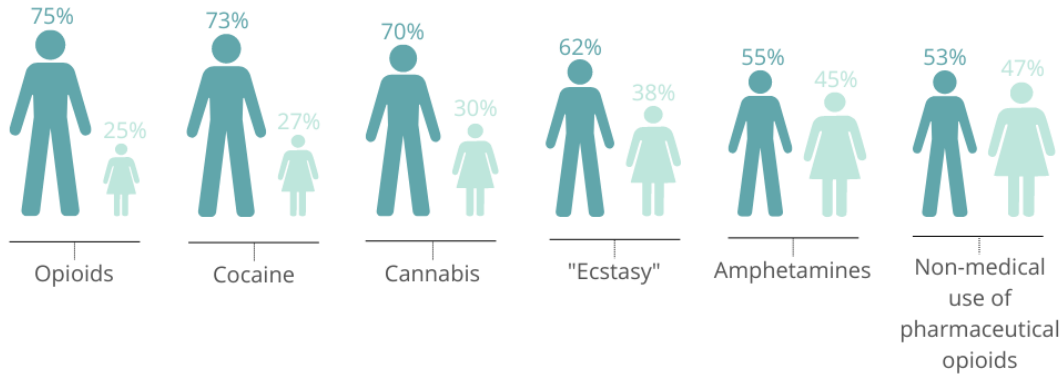


Figure 3. Prevalence of substance use by sex. About 70% of cannabis users globally are men, although in some regions the gender divide is reducing, such as North America were women account for 42% of cannabis users. The proportion of female users is higher in the case of amphetamine-type stimulants (45% of users are women) and non-medical use of pharmaceuticals (between 45% and 49% of users are women), whereas the highest share of men is found in users of opioids (75%) and cocaine (73%). *Adapted from The World drug report 2023 (UNODC, 2023).*

1.3. Disease Burden, Mortality and Social Consequences of SUDs

In 2019, 3.2 million people died due to all SUD-related causes, with approximately 10% (300,000 deaths) directly attributed to substance or alcohol overdose (Abbafati et al., 2020). To date, opioids represent the illicit substance with the biggest impact on the global burden of disease, contributing to the highest number of both direct (fatal overdoses) and indirect deaths (70% and 77% respectively) (Abbafati et al., 2020; UNODC, 2023) (**Figure 3**). A significant number of these deaths, over half, occur among individuals under the age of 50 (Abbafati et al., 2020). While deaths among young people are more commonly associated to overdose, older people are more likely to die from somatic causes, mainly liver diseases, as a consequence to long-term substance use (UNODC, 2023). Importantly, liver diseases attributed to hepatitis C account for more than half of the total number of deaths attributed to the use of substances (Zhang et al., 2022) (**Figure 4**). This can be partly attributed to the high prevalence of substance injection, which experienced an 18% increase in 2021 compared to the previous year. Estimates indicate that 50% of individuals who inject substances are living with hepatitis C, while 12% are living with HIV (UNODC, 2023).

Overall, substance use and associated disease burden have been estimated to increase over the past few decades (UNODC, 2023). Specifically, the burden of disease associated with alcohol and illicit SUDs as measured by Disability-Adjusted Life Years (DALYs), reached as high as 6% in USA and up to 2% in the European Region in 2019 (Vos et al., 2020). DALYs represent the combined years lost due to premature death and years lived with disability, serving as a comprehensive indicator of the disease's impact on population health.

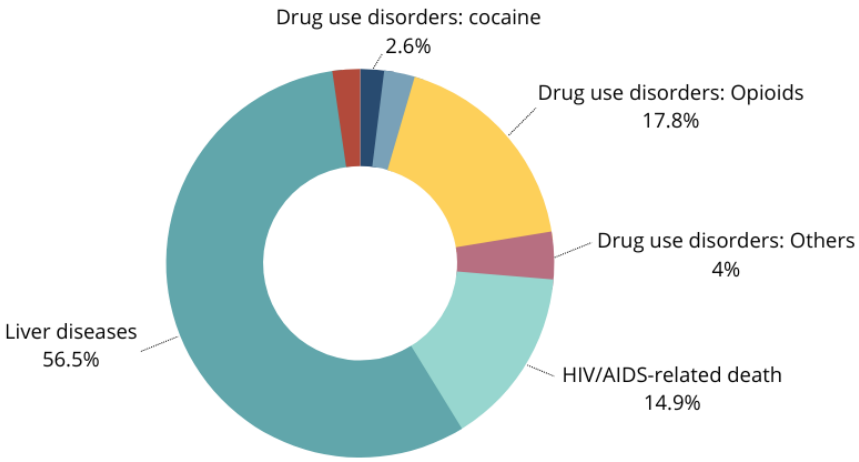


Figure 4. Causes of death related to substance use globally in 2019. Liver diseases attributed to hepatitis C are a major cause of substance-related deaths, accounting for more than half of the total number of deaths attributed to the use of substances. Moreover, two thirds of direct substance-related deaths are due to opioids. Source: Global Burden of Disease Study 2019. *Adapted from The World Drug Report 2023 (UNODC, 2023).*

Furthermore, SUDs have an immense impact on an individual's life, affecting the family circle, the interpersonal relationships and the job environment (Sheidow et al., 2012). Additionally, its repercussions extend beyond the individual, impacting the society as a whole through violence, crime, incarceration, poverty and homelessness (Sheidow et al., 2012).

In a longitudinal study of individuals entering opioid treatment programs in the USA, it was found that less than 8% of the participants had earned a bachelor's or advanced degree. This is in contrast to the estimated 32.2% of the US population with such degrees (Ellis et al., 2020). Another twin study revealed that individuals who initiated alcohol use before the age of 18 and/or had a lifetime diagnosis of alcohol use disorder completed fewer years of education compared to their cotwins without a SUD (Grant et al., 2012). Moreover, young adults not attending higher education, as well as unemployed adults

with no higher education, are more likely to abuse alcohol and/or illicit substances (Martins et al., 2015; Melchior et al., 2015). SUDs also increase the likelihood of unemployment and reduce the prospects of finding and maintaining a job (Henkel, 2011). The impact of SUDs also extends to the criminal justice system, as evidenced by a study showing that individuals with SUD by age 16 face a higher risk of incarceration for substance-related offenses in early adulthood and exhibit greater involvement in the criminal justice system (Slade et al., 2008).

1.4. Complexity and Dynamics of SUDs Trajectory

Substance use and SUDs exists on a continuum of severity (**Figure 5**). At its initial stages, where substance use is occasional, the urge of consuming the substance can be regulated. As the disease advances, there is a progressive loss of control over substance use. Individuals experience an increasing difficult time resisting the urge to consume the substance, transitioning to a state of compulsive use. A variety of factors, which will be further discussed, are known to influence each stage of substance use vulnerability, from substance use initiation to how rapidly it transitions to a severe SUD.

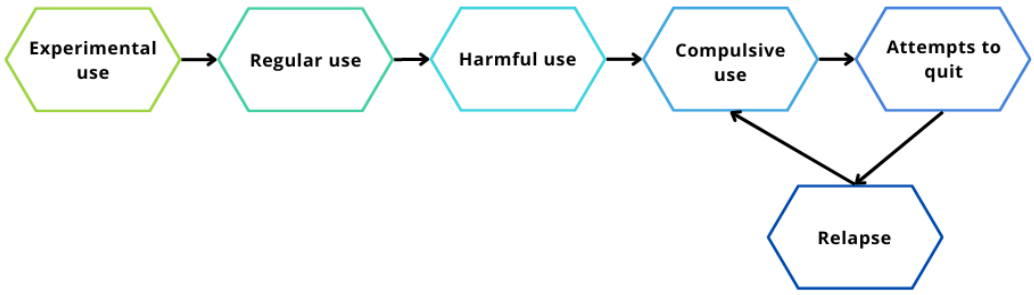


Figure 5. The multiple stages of substance use vulnerability. The phases of substance use can go from social use (experimental use) to the development of addiction (compulsive use) to vulnerability to relapse (failing to quit). *Adapted from Sanchez-Roige & Palmer (2020).*

SUDs are more likely to develop during young adulthood, between the ages of 18 and 29 (Kessler et al., 2007). The rate of transition from substance use to SUD varies by the type of substance, highly influenced by its pharmacological properties, availability and legality. However, studies have shown that early adolescent onset of substance use, before the ages of 13-16, is a robust predictor of a future SUD. Alcohol, cannabis, and illicit substance use before these ages have been associated with higher rates of alcohol dependence and SUDs in the late 20s (Ganguli et al., 2002; S. Y. Kim et al., 2023; McGue

et al., 2001; Rial Boubeta et al., 2020; Rioux et al., 2018). Earlier exposure, especially to cannabis, increases the odds of developing a SUD for heavier substances of abuse such as cocaine or opioids, compared to individuals with later onset of cannabis use, who present higher rates of cannabis or alcohol use disorder (Butelman et al., 2021).

Additionally, the disorder is characterized by considerable chronicity, with persistence rates estimates reaching as high as 45.6% (Degenhardt, Bharat, Glantz, Sampson, Al-Hamzawi, et al., 2019). Chronic SUDs are generally associated with lower quality of life, both mentally and physically (Armoon et al., 2022; Na et al., 2022). The individual's mental health status, as well as polysubstance use, are key factors in the persistence of the disorder (Degenhardt, Bharat, Glantz, Sampson, Al-Hamzawi, et al., 2019; Farmer et al., 2015).

A substantial proportion of individuals with a SUD achieve remission, as indicated by a meta-analysis from 2016 which reported remission rates ranging from 35% to 54.4%. (Fleury et al., 2016). Specially, individuals seeking outpatient treatment, those who use drug-free recovery housing exhibit higher rates for substance abstinence, as well as other positive outcomes such as increased employment and decreased involvement in the criminal justice system (Mericle et al., 2022). General social support is also crucial for the recovery process and has been positively associated with abstinence-specific self-efficacy in in-treatment SUDs patients (Stevens et al., 2015).

However, most individuals with a SUD alternate between periods of remission and relapse. Vulnerability to relapse is especially high during the first 12 months after achieving remission (McLellan et al., 2000). Relapse rates vary widely across studies, likely due to various external factors that can influence the remission/relapse cycle. For example, people in recovery from a SUD are at a higher risk of relapse if they are unemployed (Henkel, 2011) or if they come from families with a high density of SUDs, which refers to the proportion of first-degree relatives with a SUD (Farmer et al., 2022). Additionally, age can be a factor in relapse, with younger individuals who have abstained from substance use (ages 18-24) being more likely to relapse at 3- and 10 year follow-up than older abstainers (aged 55 and over) (Dawson et al., 2007; Mertens et al., 2012).

1.5. Current Treatment for SUDs

Despite the availability of effective interventions for SUDs, less than 20% of people with the disorder are in treatment (Harris et al., 2019; UNODC, 2023). In addition, access is highly unequal, especially for women who are highly underrepresented in substance treatment programs. This is particularly apparent for women who use amphetamine-type stimulants. Nearly half the users are female, yet only one in four individuals in treatment is a woman (UNODC, 2023). In the majority of European countries, opioids are the most prevalent primary substance of people in substance treatment, accounting for 38% of treatment for SUDs in 2021 (UNODC, 2023).

The "International Standards for the Treatment of Drug Use Disorders" document outlines various pharmacological interventions used in the treatment of SUDs (Busse et al., 2015). These interventions include opioid agonist therapy, which utilizes medications like methadone, buprenorphine, or morphine to alleviate withdrawal symptoms and cravings in patients with opioid use disorders. Naltrexone is another medication that blocks opioid effects and helps prevent relapse. Acamprosate and disulfiram are used to treat alcohol use disorders by reducing cravings and inducing unpleasant side effects when alcohol is consumed. Nicotine replacement therapy involves the use of nicotine patches or gum to reduce cravings and withdrawal symptoms in nicotine use disorders, while varenicline works by blocking nicotine effects on the brain. These pharmacological interventions are often combined with psychosocial interventions provided by psychologists, such as cognitive-behavioral therapy, motivational interviewing, and contingency management. Psychosocial interventions aim to develop coping skills, enhance self-esteem, and strengthen social support networks for patients with a SUD.

1.6. Neurobiology of SUDs

From a psychopathological framework, SUDs are chronically relapsing disorders that cycle through three stages: (i) binge/intoxication, (ii) withdrawal/negative effect, and (iii) preoccupation/anticipation (**Figure 6**). These stages involve neuroplastic changes in neurobiological mechanisms, such as brain reward, stress and executive function, differentially involved in the transition from recreational to compulsive substance use. In addition, each stage is specifically associated with *allostatic* changes in three key

neurocircuits: (i) basal ganglia, (ii) extended amygdala, and (iii) prefrontal cortex (Koob & Volkow, 2016) (**Figure 6**).

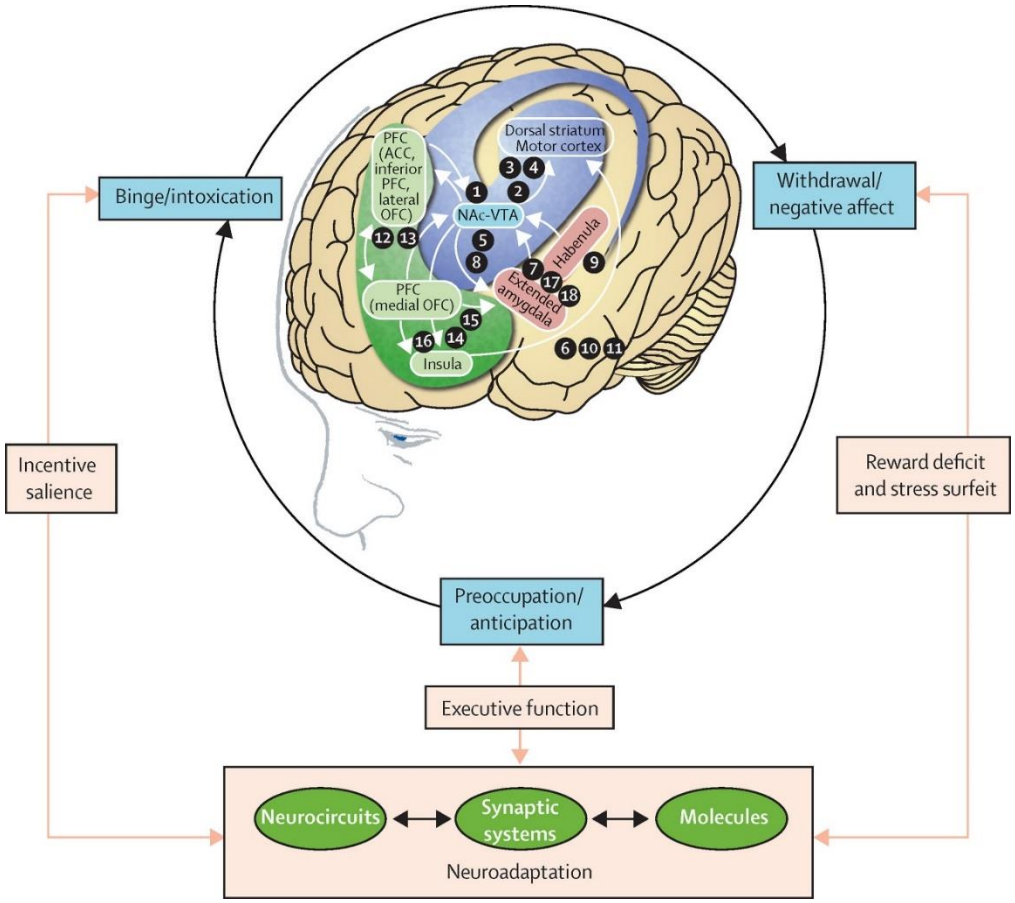


Figure 6. Neural circuitry associated with the three stages of the addiction cycle. The overall neurocircuitry domains correspond to three functional domains: binge/intoxication (reward and incentive salience: basal ganglia [blue]), withdrawal/negative affect (negative emotional states and stress: extended amygdala and habenula [red]), and preoccupation/anticipation (craving, impulsivity, and executive function: PFC, insula, and allocortex [green]). Arrows depict major circuit connections between domains, and numbers refer to neurochemical and neurocircuit-specific pathways known to support brain changes that contribute to the allostatic state of addiction. PFC=prefrontal cortex. ACC=anterior cingulate cortex. OFC=orbitofrontal cortex. NAc-VTA=nucleus accumbens-ventral tegmental area. *Adapted from Koob & Volkow (2016).*

In the binge/intoxication stage, an individual consumes a substance or drug of abuse and experiences its rewarding or pleasurable effects. This stage involves three main contributors: (i) The acute positive hedonic value of drugs, which referrers to the fast and intense pleasant sensations produced by drugs; (ii) Sensitization of incentive salience, where neutral stimulus acquires incentive value through association with a drug; (iii) Inherent poor cognitive insight, including cognitive deficiencies such as impairment of objectivity and resistance to corrective feedback (George & Koob, 2017) (**Figure 6**).

Drugs activate the brain's reward system, increasing dopamine release through dopamine D1 receptors in the ventral striatum, creating a sense of "high" (Volkow et al., 2007). Other relevant neurotransmitters and neuromodulators, such as opioid peptides, serotonin, glutamate, γ-aminobutyric acid (GABA), acetylcholine and endocannabinoid systems, play a role in this rewarding and reinforcing effects of drugs. Through these neurotransmitters, drugs disrupt inhibitory control, decision making and normal functioning of reward, motivation, stress and memory circuits (Koob & Volkow, 2016). Additionally, dopamine neurons exhibit phasic responding to drug reward, which means that after repeated exposure, the same neurons stop responding to a predictable reward, leading to sensitization of incentive salience and triggering the desire for the drug (craving) and compulsive use in response to stressful environments (George & Koob, 2017) (**Figure 6**).

In the withdrawal/negative affect stage the individual experiences a negative emotional state in the absence of the drug. During this stage, chronic irritability, emotional pain, stress and loss of motivation for natural rewards, drive drug consumption. This process involves within-system neuroadaptations, where the primary target for the drug adapts to neutralize the effect of the drug. Decreases in dopaminergic, serotonergic and GABAergic transmission and increases in μ opioid receptor and NMDA glutamatergic transmission, are studied neuroadaptations during this stage (Melis et al., 2002). Moreover, between-system neuroadaptations, which refers to systems other than those involved in the positive rewarding effects of drugs, are recruited or dysregulated by chronic drug use to oppose the rewarding effects of the drug (Koob & Le Moal, 2008). These include, the dysregulation of the hypothalamic-pituitary-adrenal axis, the brain stress system mediated by corticotropin-releasing factor and the dynorphin-κ opioid receptor system (Whitfield et al., 2015). All these processes together are implicated in the long-term biochemical changes that contribute to the clinical manifestation of the withdrawal syndrome and tolerance to the drug (Whitfield et al., 2015) (**Figure 6**).

Lastly, in the preoccupation/anticipation stage the individual seeks the drug again after a period of abstinence, commonly called craving. Two opposite systems are implicated: (i) In the go system (cue-induced craving) there is a reactivation of the dopamine release during acute craving episodes (Niendam et al., 2012). (ii) In the stop

system (inhibitory function), increases in GABAnergic activity and decreases in dopamine D2 receptor availability cause chronic executive dysfunction, including impairments in decision making, self-regulation, inhibitory control, attention and working-memory (Volkow et al., 2010) (**Figure 6**).

KEY POINTS SECTION 1

- SUDs are neuropsychiatric disorders characterized by a recurring desire to continue taking a substance or drug regardless of its destructive consequences.
- Among the various clinical tools design to diagnose SUDs, the DSM is widely used. The latest edition (DSM-5) follows a 11-item criteria list to diagnose SUDs, ranging from mild (2-3 items) to moderate (4-5 items) to severe (6 or more items) SUDs.
- Over 2% of the population worldwide are dependent on alcohol or an illicit substance, although it varies across regions. Cannabis is the most used drug worldwide whereas opioids are the main drug impacting the global burden of disease.
- SUDs have detrimental effects on individuals' lives, families, and societies, contributing to unemployment, crime, and poverty. However, treatment for SUDs remains limited, especially for women.
- Substance use and SUDs exists on a continuum of severity: drug experimentation → regular use → harmful use → compulsive use → quit attempts → relapse.
- Neurobiological adaptations occur in the brain during substance use that drive the transition from initial to compulsive use. Three stages of addiction are identified: binge/intoxication, withdrawal/negative affect, and preoccupation/anticipation, involving neuroadaptations in different brain circuits.

2. SUDs and Psychiatric Comorbidity

Many individuals who develop a SUD also experience comorbid psychiatric disorders, making it a rule rather than an exception. According to recent data (Carliner et al., 2017), approximately 53% of individuals with a primary diagnosis of a SUD have at least another co-occurring diagnosis for a mental, behavioral, or emotional disorder. Moreover, this estimate is even higher (58%) when looking at the young adult population, between 18 and 25 years old.

Numerous systematic reviews, as well as population-based studies, have assessed the comorbidity rates between SUDs and relevant psychiatric disorders. Overall, post-traumatic stress disorder (PTSD) and schizophrenia have the highest comorbidity rates with SUDs, with estimates around 46% and 40%, respectively, followed by bipolar disorder with 35%, major depressive disorder, anxiety disorder and attention-deficit/hyperactivity disorder (ADHD) with 25% each, and obsessive compulsive disorder (OCD) with 11% (Hunt et al., 2016, 2018, 2020; Lochner et al., 2014; Pietrzak et al., 2011; Toftdahl et al., 2016) (**Figure 7**). SUDs also present high comorbidity rates with personality disorders, specially borderline and antisocial personality disorder, with estimates up to 46% (Toftdahl et al., 2016) (**Figure 7**).

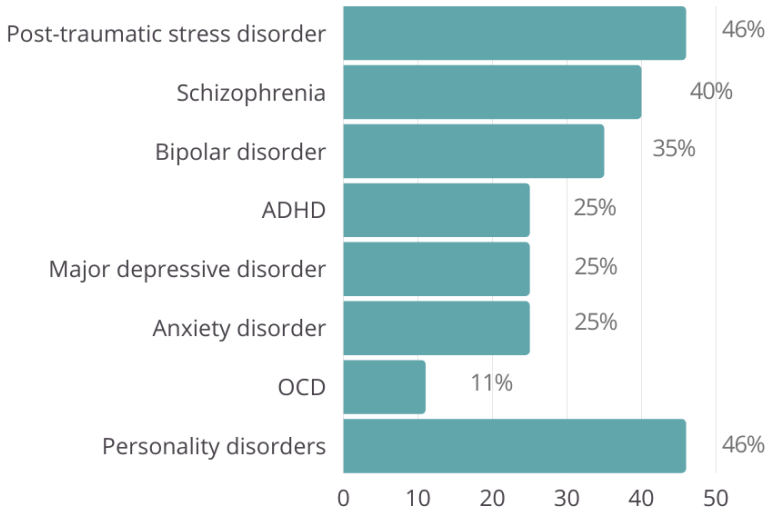


Figure 7. Prevalence rates of co-occurring SUDs and major mental health disorders. ADHD= Attention-deficit/hyperactivity disorder. OCD= Obsessive compulsive disorder.

Overall, having lifetime diagnoses of mental health disorders predisposes to an increased risk of initiating substance use and developing a SUD. Around 30-50% of individuals with a mental disorder also experience a SUD (Carliner et al., 2017). Particularly, psychotic disorders have the strongest associations with early onset of SUDs (Hunt et al., 2018; Lev-Ran et al., 2013). In addition, the co-occurrence of SUDs and mental disorders is linked to a more severe and debilitating course of illness, as well as unfavorable social and clinical outcomes (Andersson et al., 2019; Harris et al., 2019; Krawczyk et al., 2017; Levola et al., 2022). For instance, personality disorders such as antisocial, borderline, and schizotypal personality disorders, as well as comorbid major depressive, anxiety and sleep-related disorders, have been linked to SUDs persistence in epidemiologic and longitudinal studies (Crum et al., 2004; Fenton et al., 2012; D. Hasin et al., 2011; Tuithof et al., 2013). This co-occurrence also represents a challenge for the treatment and remission of SUDs. It has been extensively reported that individuals with a SUD and psychiatric comorbidity experience poorer treatment adherence and completion (Andersson, Lauvsnes, et al., 2021; Charney et al., 2005; Dodge et al., 2005; Hesse, 2009; Krawczyk et al., 2017; Lipsky et al., 2010; Ostacher, 2007) and higher risk for relapse (Andersson et al., 2019; Boschloo et al., 2012; Crum et al., 2004; Lipsky et al., 2010). Personality profiles can also impact treatment outcomes. Traits such as neuroticism and impulsivity have been associated with higher risk of relapse and increased symptom severity in individuals undergoing treatment (Bucher et al., 2019; Turner et al., 2021). Conversely, extraversion, agreeableness and openness are personality traits that have been associated with favorable treatment outcomes (Zilberman et al., 2018).

On the other hand, developing a comorbid SUD in individuals with a mental health disorder diagnosis is also associated with increased symptom severity and worse disease trajectory. The presence of a comorbid SUD has been associated with an earlier onset of schizophrenia, as well as increased hospitalization and mortality. Mortality rates are 12.7% higher for individuals with comorbid schizophrenia and a SUD than for those with schizophrenia alone (Lähteenvuo et al., 2021). Moreover, the presence of a comorbid SUD and mood disorders, such as bipolar and major depressive disorder, increases the risk of depressive episodes and mortality, specially due to suicide (Blanco et al., 2012; Levola et al., 2022), as well as legal and academic difficulties during youth (Goldstein & Bukstein,

2010). A 42-year follow up study on SUDs patients admitted into a detoxification unit, found that comorbid primary psychoses or mood disorders predicted higher risk of premature mortality and deaths by overdose (Fridell et al., 2019).

Moreover, individuals with comorbid PTSD and SUDs present an overall lower physical and mental health status (Mills et al., 2006), probably as the result of a higher prevalence of depression, anxiety, suicidality, increased unemployment, and social impairment (Flanagan et al., 2016; Pietrzak et al., 2009, 2011). Although PTSD is deeply studied in adults and veterans, these impairments can also be present in youth. For instance, a study in adolescence with psychotic symptoms showed that co-occurring SUDs and PTSD was predictive of increased psychotic symptoms compared to individuals with only SUDs (Basedow et al., 2023).

SUDs diagnosis is an important risk factor for suicidal thoughts and behaviors (Yuodelis-Flores & Ries, 2015). Furthermore, the impact on suicidality is even greater when individuals have multiple SUDs diagnoses compared to having just one. A recent study discovered that individuals with five SUDs diagnoses, including alcohol, cannabis, cocaine, tobacco, and opioids, exhibited a 6.77-fold increase in suicidal ideation and a 3.61-fold increase in suicide attempts when compared to those with a single SUD diagnosis (Polimanti et al., 2021).

2.1. Causal Hypotheses of the Comorbidity between SUDs and other Psychiatric Conditions

The relationship between psychiatric comorbidity and SUDs is complex, and it is still unclear whether one precedes the other or whether they are mutually induced (**Figure 8**). However, it is likely that there is no single explanation for this phenomenon. When examining the co-occurrence of externalizing or internalizing disorders with SUDs, it becomes clear that there are various risk processes differentially relevant across subgroups of individuals. Externalizing disorders are associated with a higher risk of SUDs through undercontrol/disinhibition behaviors (Zucker et al., 2011). This framework proposes that one of the core risk pathways that converge in both externalizing disorders and SUDs involves a vulnerability to disinhibitory processes, which is expressed at the behavioral level by high undercontrol (Zucker et al., 2011). Studies show that individuals

exhibiting higher levels of externalizing behaviors, such as impulsivity, sensation seeking, and risk taking, as well as externalizing psychopathology diagnoses in youth, such as conduct disorder, oppositional defiant disorder and ADHD, are at higher risk of early onset of substance use and SUDs in early/mid adolescence. Similarly, youth with increased neurobehavioral disinhibition, described by indicators of executive cognitive function, emotional regulation and behavioral control, also exhibit higher risk for SUDs (Tarter et al., 2003).

Conversely, internalizing disorders can increase the risk of SUDs through self-medication of negative affect mechanisms, whereby individuals use substances in an attempt to reduce or relieve symptoms associated with internalizing psychopathology (Hussong et al., 2011; O'Neil et al., 2011). Studies show that high levels of internalizing behaviours in childhood, such as hopelessness, depressive and anxiety disorders have been associated with increased risk of early onset of substance use and SUDs (Bushnell et al., 2019; Malmberg et al., 2010; Virtanen et al., 2021).

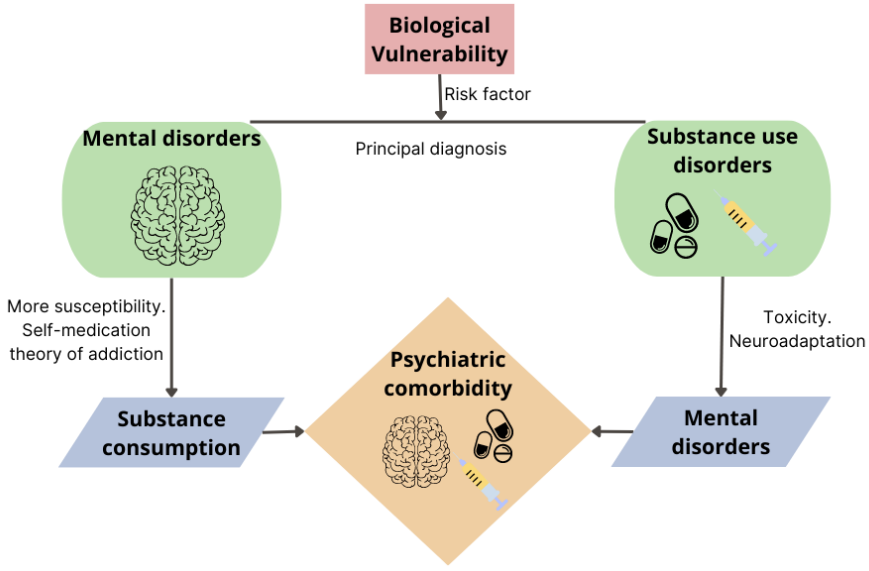


Figure 8. Hypothesis about the causal relationships between SUDs and psychiatric comorbidity. *Adapted from Ornell et al. (2021).*

Both the undercontrol/disinhibition and the self-medication hypotheses suggest that psychiatric disorders precede SUDs. However, the use of psychoactive substances can also contribute to the development of mental illness by inducing changes in brain areas that are disrupted in other mental disorders, predisposing the individual to develop secondary psychopathologies (Balhara et al., 2017) (**Figure 8**). Additionally, there is a

growing body of evidence that supports common neurobiological, genetic and environmental influences underlying the high comorbidity observed between SUDs and mental health related conditions (Ornell et al., 2021; Uher & Zwicker, 2017) (**Figure 8**).

2.2. Comorbidity Between SUDs and ADHD

Given the focus of this thesis, the relationship between SUDs and ADHD needs special attention. ADHD is a common neurodevelopmental disorder that severely impairs the daily functioning of the patients due to age-inappropriate levels of impulsivity and hyperactivity, and/or difficulties in focusing attention (Faraone et al., 2015). The estimated prevalence of ADHD during childhood ranges from 4% to 7%, and there is growing evidence indicating its persistence into adulthood in approximately 15% to 65% of individuals (Faraone et al., 2015).

The prevalence of ADHD in individuals with a SUD varies across studies and populations, ranging from 14% to 56% (Icick et al., 2020; Oliva et al., 2021; van de Glind et al., 2014; Van Emmerik-van Oortmerssen et al., 2014). A recent meta-analysis compiling 31 studies estimated the prevalence of ADHD among SUDs populations at 21% (Rohner et al., 2023). Cannabis, alcohol or tobacco are the most common substances of abuse among adolescents with ADHD (Gujska et al., 2023; S. Young et al., 2023). Research indicates that individuals with co-occurring ADHD and SUDs show more severe symptoms of SUDs, including earlier onset of substance use, faster transition to a SUD and higher frequency of polysubstance dependence (Charach et al., 2011; Icick et al., 2020; Ilbegi et al., 2018; S. S. Lee et al., 2011; Molina et al., 2018). A study using Danish large national registers data reported that the mean age of a SUD diagnosis in individuals with ADHD was 25 years old (Steinhausen & Bisgaard, 2014). A four-year follow-up study found that adolescents diagnosed with ADHD during childhood were two times more likely to develop a SUD and more than eight times more likely to develop nicotine dependence, compared to healthy controls (Groenman et al., 2013). Moreover, ADHD diagnosis after the age of 13, parental history of mental disorders, including a SUD, and low parental socio-economic status are risk factors associated with increased risk for SUDs in ADHD individuals (Wimberley et al., 2020).

In addition, treatment seeking patients with comorbid SUDs and ADHD experience greater difficulty remaining abstinent (Kaye et al., 2019), with severity of SUDs being a predictive factor for treatment outcome (Akalın & Bilici, 2022). Moreover, these individuals exhibit greater psychiatric comorbidity, especially major depressive and personality disorders (Regnart et al., 2017; Van Emmerik-van Oortmerssen et al., 2014), increased risk of suicide attempts, and overall reduced quality of life (Icick et al., 2020; Katzman et al., 2017). Individuals with ADHD also present high rates of conduct disorder and oppositional defiant disorder, which are suggested to mediate some of the disruptive behaviors that predispose ADHD individuals into developing a SUD, as well as with increased symptom severity (Regnart et al., 2017; Torok et al., 2012).

Large nationwide and observational data shows that methylphenidate given to young ADHD patients might be contributing to a decreased risk of developing SUDs (Groenman et al., 2019; Quinn et al., 2017; Steinhausen & Bisgaard, 2014). The onset of medication can also be an interfering factor in the occurrence of SUDs in these individuals, since longer and earlier age at onset of medication treatment has been associated with lower rates of SUDs (Steinhausen & Bisgaard, 2014).

KEY POINTS SECTION 2

- It is estimated that 53% of individuals with a primary diagnosis of a SUD have at least another co-occurring diagnosis for a psychiatric or behavioral disorder, the most prevalent being post-traumatic stress disorder, schizophrenia and personality disorders.
- Comorbid mental health disorders in SUDs cases are associated with a more severe and disabling course of illness, poorer social and clinical outcomes, and treatment complications.
- The relationship between SUDs and comorbid disorders is complex, involving undercontrol disinhibition and self-medication processes. The use of psychoactive substances can also induce the development of secondary mental disorders.
- Common risk factors (genetic and environmental) can contribute to the co-occurrence of SUDs and psychiatric disorders.
- ADHD frequently co-occurs with SUDs, with an estimated prevalence of 21%, and its associated with earlier onset of substance use. Cannabis, alcohol or tobacco are the most common substances of abuse among adolescents with ADHD.

3. Etiology of SUDs

The etiology of SUDs is complex and multifactorial, involving an interplay between the type of substance or substances consumed, the individual's genetic background, and environmental factors (C. P. O'Brien, 2011). Many factors are known to be involved in the development of SUDs, mainly genetic predisposition, exposure to adverse childhood experiences, the stage of life during which substance use began, personality traits, and comorbid mental health conditions (Prom-Wormley et al., 2017). These individual aspects are further influenced by broader social-related factors such as the level of familial and community support, socioeconomic status and the legal status and availability of the substances. The intricate interplay between these factors explains the substantial heterogeneity observed in individuals with SUDs, evidenced by the diversity of clinical profiles, symptom trajectories, comorbidity patterns and chronicity of the disorder (Beseler et al., 2006; Prom-Wormley et al., 2017). Additionally, the distinct pharmacological characteristics of different substances also significantly contribute to the risk of addiction, impacting the speed at which substance use can escalate to a SUD (**Figure 9**).

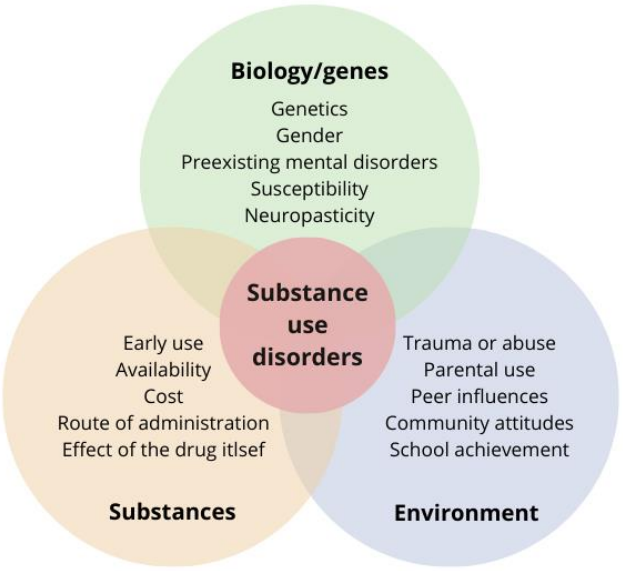


Figure 9. Factors that increase the risk for substance use and SUDs.

To investigate the etiology of complex disorders like SUDs, family, adoption, and twin studies are well-suited as they allow for the control of shared and individual genetic and environmental effects (Agrawal & Lynskey, 2008) (**Box 2**).

Box 2. Population-based study designs to explore SUDs etiology

Family studies: Case-control family studies provide initial clues to potential heritable influences on addictive disorders by examining the risk of SUDs in the first-degree relatives of individuals either with or without a SUD.

Adoption studies: Adoption studies are based on a comparison of the concordance or correlation between offspring disease status and the characteristics of both the biological and adoptive parents: similarity between offspring and biological parents is suggestive of genetic influences, while similarity between offspring and adoptive parents is suggestive of environmental influences.

Twin studies: Twin studies use monozygotic and dizygotic twin pair variances (that share 100% and about 50% of their genetic material, respectively) to estimate the proportion of the total phenotypic variance of a trait due to additive genetic (additive genetic effects of alleles at every locus), shared environment (common in both twins), and unique environmental influences (not shared by members of the twin pair).

Family studies have demonstrated that SUDs tend to cluster within families (Kendler et al., 1997). In fact, individuals with a first-degree relative with a history of a SUD have up to an eight-fold higher risk for developing the disorder (Merikangas et al., 1998). Furthermore, there is evidence of specificity of familial aggregation of the predominant substance of abuse, suggesting that, in some cases, there may be risk factors that are specific to particular substances (Bierut et al., 1998; Merikangas et al., 1998). Nevertheless, family studies alone cannot distinguish weather the causes of familial similarity are genetic or environmental.

Adoption studies have been challenging to conduct due to difficulties in accessing adoption records. However, the few studies that have been conducted suggest that the rates of alcohol problems and other psychiatric disorders are higher for adopted children of biological parents with alcohol problems (Cadoret et al., 1995; Goodwin et al., 1973). These findings provide support for a direct genetic effect between the biological parent's substance use and the offspring's risk for developing a SUD. Furthermore, Cadoret et al., (1995) highlighted the significant role played by specific environmental factors within

adoptive families, such as parental divorce and parental psychiatric disorder, in the progression of SUDs.

Literature on twin studies provides valuable insights into specific sources of variation in the etiology of SUDs. These studies consistently report that substance initiation is influenced by a combination of genetic factors, as well as shared and unique environmental factors (Huizink et al., 2010). Meanwhile, the progression from regular substance use to a SUD is predominantly driven by addictive genetic factors (Agrawal & Lynskey, 2008; Fowler et al., 2007; Kendler et al., 2003). Additionally, these studies reveal the presence of both shared and substance-specific genetic factors in the etiology of SUDs. On the other hand, shared environmental effects appear to impact on the risk for substance use and SUDs in a nonspecific-substance manner, with significant contributions observed during adolescence but not in adulthood (Hopfer et al., 2003; Kendler et al., 2003).

Twin studies have also allowed to estimate the proportion of the total phenotypic variance that can be attributed to genetic factors, referred to as *heritability (h^2).* SUDs heritability has been estimated to be of 40-60%, although it can vary among specific substances (Deak & Johnson, 2021; Lopez-Leon et al., 2021). The heritability for opioid use disorder has been estimated to be of 50%, and similar heritability has been estimated for alcohol use disorder (50-64%) and for cannabis use disorder (51-59%) (Berrettini, 2017; Verhulst et al., 2015). For cocaine use disorder, heritability estimates range from 40-80%, although there is little evidence of cocaine-specific genetic influences (Kendler et al., 2007). Tobacco use disorder heritability shows substantial variability, ranging from 30-70%, which could be explained, at least in part, due to different scales and questionnaires used to assess tobacco use disorder, namely the DSM or the Fagerström Test for Nicotine Dependence (FTND) (Deak & Johnson, 2021) (**Figure 10**).

Most studies examining SUDs have focused on a single substance of abuse. These studies suggest that part of the genetic factors attributed to the heritability of SUDs are substance-specific, particularly for tobacco (Kendler et al., 2007). When it comes to illicit substances, the influence of substance-specific genetic factors appears to be modest overall (Kendler et al., 2003; Tsuang et al., 2001), with the exception of opioids, where it

has been estimated that 38% of the variation in opioid use disorder can be attributed to specific genetic factors (Nielsen & Kreek, 2012; Tsuang et al., 1998). However, many individuals use more than one substance (Morley et al., 2015) and there is compiling evidence of the common genetic architecture and genetic liability underlying all SUDs (Hatoum et al., 2022; Palmer et al., 2015).

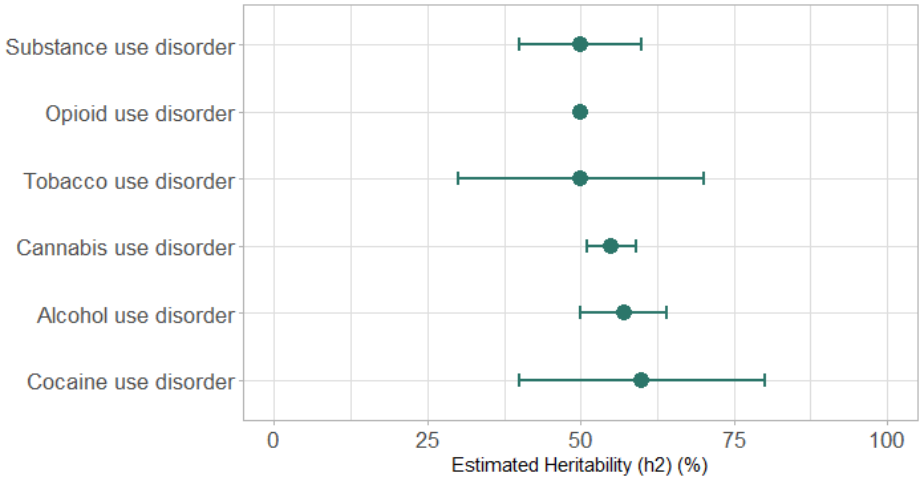


Figure 10. Ranges of estimated heritability of SUDs across substances.

3.1. Genetic Factors of SUDs

In the scope of this thesis, the goal of genetic studies is to identify genetic risk factors that may have a role in the development or progression of complex disorders, such as SUDs. There are many different methodologies, study designs and analytical tools for identifying genetic risk factors. This thesis will focus on genome-wide association studies (GWASs). The current literature on GWASs, has shown that most traits are influenced by thousands of *causal genetic variants* that individually confer very little risk, are often associated with multiple traits at once and are correlated with causal and non-causal variants that are physically close in the DNA sequence (Uffelmann et al., 2021). Given these limitations, it can be challenging to draw clear conclusions from GWASs alone, and post-GWASs functional studies have had an essential role in understanding the genetic architecture of complex traits and predicting disease risk (Gelernter & Polimanti, 2021). All these aspects will be discussed further in this section, along with the most relevant findings for SUDs. While this section focuses on exploring findings related to common genetic variation from GWASs, it's essential to acknowledge that there are other types of genetic variations that also significantly contribute to the genetic etiology of SUDs,

including copy number variants (CNV) and rare variants (Cabana-Domínguez et al., 2016; D. Li et al., 2015; Rajagopal et al., 2023; Sulovari et al., 2018).

3.1.1. Genome-Wide Association Studies

GWASs aim to identify genetic variations that are associated with a particular trait or disease by testing for differences in the allele frequency of genetic variants between individuals who differ phenotypically (Uffelmann et al., 2021). This approach can be implemented for both categorical (e.g., case/control disease status) and continuous (e.g., height, educational attainment) phenotypes. The most frequently studied genetic variation in GWASs are single-nucleotide polymorphisms (SNPs), which are the most abundant form of genetic variation in the human genome (Ku et al., 2010). In addition, SNPs classify as a type of common genetic variation, which means that they are present in a large proportion of the human population, with a Minor Allele Frequency (MAF) greater than 1%. This form of common genetic variation generally has small effects on the phenotype or complex trait, whether other forms of genetic variation, such as rare variants (less frequent in the population, MAF less than 1%), tend to have larger effects on the phenotype. Other types of genetic variation that can be studied in GWASs include insertions, deletions and structural variants (Ku et al., 2010).

The success of GWASs relies on the ability to capture the whole genome common genetic variation by only targeting a subset of tag SNPs, via *linkage disequilibrium* (LD) (Bush & Moore, 2012). LD is the non-random co-occurrence of alleles at different *loci*, meaning that certain alleles are inherited together more often than expected by chance, and therefore form a *haplotype* (Slatkin, 2008). Generally, physically close variants are at higher LD than distant variants, and a high LD implies that the information in one SNP is strongly predictive of the other SNP. Therefore, a GWAS containing information about the association of a subset of tag SNPs on a given phenotype can detect the association of a SNP that is in LD with a *causal variant* for that phenotype (Bush & Moore, 2012). In addition, it is possible to aggregate the contribution of each genome-wide locus estimated in a GWAS to determine the proportion of variance in *liability* explained by these loci together, thus quantifying the effects of all SNPs. This is referred to as *SNP-*

based heritability (h^2_{SNP}). **Box 3** presents an overview of the general workflow for GWAS, which will be elaborated below.

Commercial genotyping arrays vary in their genome coverage and disease specificity. For example, Illumina offers microarrays with coverage ranging from 640 K to 2.4 million markers (Verlouw et al., 2021). Following genotyping, data must undergo rigorous quality control at both the individual and SNP level to ensure reliable results from GWAS. Briefly, quality control involves removing rare or monomorphic variants, those not in *Hardy-Weinberg equilibrium (HWE)*, and SNPs missing from a proportion of individuals (Reed et al., 2015). Moreover, individuals with missing genotypes, duplicates, related individuals, sex discrepancy and outlying *heterozygosit*y rate are also removed (Reed et al., 2015). Furthermore, ancestry and *population stratification* must be carefully accounted for through *Principal Component Analysis (PCA),* which clusters individuals by genotype similarity and allows to detect and exclude ancestry outliers (Hellwege et al., 2017; Uffelmann et al., 2021). Next, imputation increases genomic coverage by inferring variants that have not been directly genotyped (Bush & Moore, 2012), using reference panels with known genotypes and LD patterns, such as the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016) and the 1000 Genomes (1KG) Project (Auton et al., 2015) (**Box 3**).

The core step of the GWAS is the single-locus statistic test, examining each SNP independently for association with the phenotype (Bush & Moore, 2012). GWASs results are reported in *summary statics*, which includes effect sizes (odds ratio (OR) or beta (β)), standard errors and p-values, among other parameters, and visualized using Manhattan plots. Genome-wide significance (GWS) is determined using a stringent multiple-testing threshold of 1 million independent tests, resulting in a *Bonferroni corrected* threshold of p-value$<5x10^{-08}$ (Uffelmann et al., 2021). Downstream analyses are necessary to identifying causal variants, their functional significance, and any potentially meaningful biological pathways (**Box 3**). To obtain larger sample sizes and detect loci influencing complex traits, GWASs meta-analyses (GWASMAs) are performed by combining data from multiple sites (Zeggini & Ioannidis, 2009) (**Box 3**). **Box 4** details some of the major biobanks and collaborative projects that have provided insights into the genetics of SUDs.

Box 3. Overview of the general workflow of a GWAS

Data collection
Data can be collected from study cohorts or available genetic and phenotypic information can be used from biobanks or repositories. *Confounders* need to be carefully considered and recruitment strategies must not introduce biases such as *collider bias*.

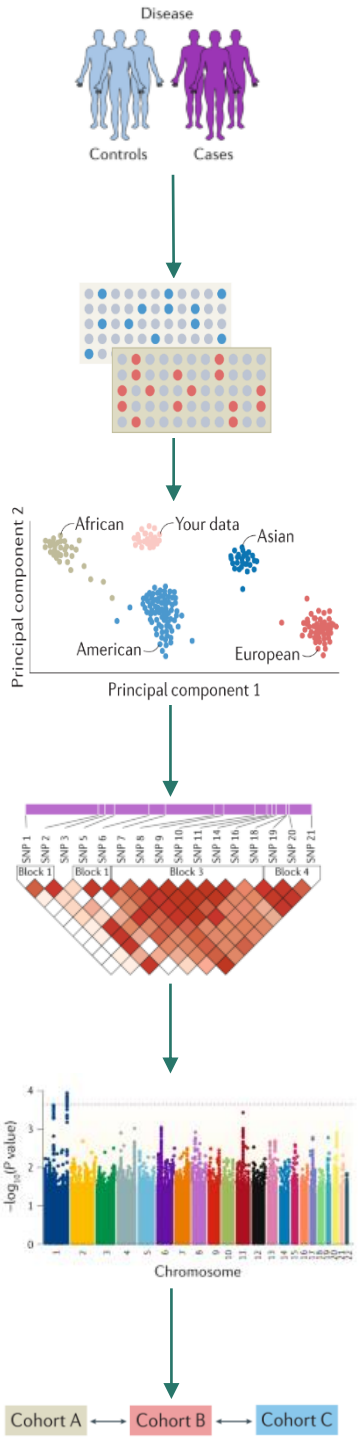
Genotyping
Genotypic data can be generated using microarrays to capture common variants, with coverage ranging from 640K to 2.4 million markers.

Quality control
Quality control steps include deletion of bad quality SNPs and samples, detection of population stratification in the sample and calculation of principle components, through principal component analysis (PCA).

Imputation
Imputation involves the statistical inference of genotypes that have not been assayed directly using information from matched reference populations. Genomic coverage increases up to 8-10 million variants.

Association testing
Genetic association tests are run for each genetic variant, using an appropriate model (e.g., linear or logistic regression). Confounders are corrected for, including principal components, sex and age. Output is visualized in Manhattan plots and summary statistics are generated.

Meta-analysis
Results from multiple cohorts are combined using standardized statistical pipelines.



Adapted from Uffelmann et al. (2021)

Box 4. Relevant biobanks and collaborative projects for substance use and SUDs

- **23andme** (https://www.23andme.com/) is a biotechnology company that offers direct to consumer genetic testing. Customers can participate in research programs by completing an online surveys regarding health-related outcomes. This data is used to investigate the genetics of common diseases and traits.
- The GWAS and Sequencing Consortium of Alcohol and Nicotine use (**GSCAN**) (https://genome.psych.umn.edu/index.php/GSCAN) is an international meta-analysis consortium with a focus on understanding the etiology of alcohol and nicotine use and addiction, that aggregates genetic association findings across studies with millions of individuals.
- The Million Veteran Program (**MVP**) (Gaziano et al., 2016) is a national research program looking at how genes, lifestyle, military experiences, and exposures affect health and wellness in Veterans.
- The Psychiatric Genomics Consortium (**PGC**) (https://pgc.unc.edu/) is a large collaborative effort including 800+ investigators, 36 countries, and >400K subjects aimed toward elucidating the genetic contributions across psychiatric disorders. The PGC consists of 14 working groups, including the Substance Use Disorders (PGC-SUD) Working Group.
- **UK Biobank** (Bycroft et al., 2018) is a large-scale biomedical database and research resource, containing in-depth genetic and health information through self-reported questionnaires from half a million UK participants.

3.1.1.1. Genome-Wide Association Studies of SUDs

Large scale GWASs have mainly focused on investigating individual substances to uncover the polygenic architecture of SUDs, including alcohol, tobacco, cannabis, opioid or cocaine use disorder. In addition, GWASs have utilized various intermediate phenotypes related to SUDs, such as quantitative measures like number of cigarettes smoked per day or drinks consumed per week, as well as indicators of substance use initiation, such as smoking or cannabis use initiation. These intermediate phenotypes allow the study of SUDs-related behaviors to uncover the substantial genetic complexity involved in the development of SUDs (Deak & Johnson, 2021). Nevertheless, given the high prevalence of polysubstance use (Morley et al., 2015), GWASs conducted on individual substances are likely to include subjects with multiple SUDs diagnoses. This

circumstance can potentially diminish the power to identify genetic risk factors specific to each substance. For an overview of the largest GWAS conducted to date on SUDs and SUD-related phenotypes, see **Table 1**.

Alcohol. Alcohol-drinking behaviors are among the few complex mental health phenotypes that have common risk alleles with relatively large effect sizes. The influence of genes encoding alcohol-metabolism enzymes, such as Alcohol Dehydrogenase 1B (*ADH1B*) and Aldehyde Dehydrogenase 2 (*ALDH2*), has been consistently associated with alcohol use disorder and alcohol-related phenotypes. In addition, recent GWAS with larger sample sizes have successfully identified numerous other risk loci for alcohol behaviors. A GWAS of self-reported alcohol consumption identified two GWS loci in the β-Klotho (*KLB*) gene, previously associated with alcohol consumption and with alcohol preference in mice studies, and a few novel GWS loci, one in the Dopamine Receptor D2 gene (*DRD2*), which plays an important role in addiction, and another in the Glucokinase Regulator gene (*GCKR*) (Clarke et al., 2017) (**Table 1**). A following GWAS, which investigated separately alcohol consumption and problematic drinking, replicated the well-stablished *ADH1B*, *ADH1C*, *KLB* and *GCKR* gene findings as well as finding several novel associations, including the Junctional Cadherin 5 Associated gene (*JCAD*) and the Solute Carrier Family 39 Member 13 gene (*SLC39A13*) (Sanchez-Roige et al., 2019) (**Table 1**). Similarly, a GWAS for alcohol consumption and alcohol use disorder was able to replicate the previous findings in a multi-ancestry population (Kranzler et al., 2019). From the 18 GWS loci, only five loci were associated with both traits (alcohol consumption and alcohol use disorder), including loci in *ADH1B* gene as the lead finding, *ADH1C* gene and the Solute Carrier Family 39 Member 8 gene (*SLC39A8*), among others. *KLB* was only found to be associated with alcohol consumption, while *DRD2* only showed an association with alcohol use disorder (**Table 1**). Another GWAS assessed the maximum habitual alcohol intake and, again, identified the *ADH1B* gene as the lead finding for all ancestries, while also replicating relevant alcohol use loci for heavy alcohol use (Deak, Levey, et al., 2022) (**Table 1**). The alcohol-related trait "drinks per week" has also been used in numerous studies as an intermediate phenotype for alcohol use disorder. The largest GWAS for drinks per week found 81 risk loci associated with this phenotype (M. Liu et al., 2019) (**Table 1**). The most relevant GWAS for alcohol use disorder to date is a

meta-analysis combining 435,563 individuals assessed for alcohol use disorder, alcohol dependence, and problematic drinking, which allowed the study of the trait denominated problematic alcohol use. This study identified 29 independent risk variants, 19 of them novel, elucidating relevant findings, such as the Phosphodiesterase 4B gene (*PDE4B*), the Cell Adhesion Molecule 2 gene (*CADM2*), and a novel rare variant in *ADH1B* (Zhou, Sealock, et al., 2020) (**Table 1**). This study estimated a h^2_{SNP} of 6.8%, which is in line with the other alcohol use disorder GWASs (Kranzler et al., 2019; Zhou, Sealock, et al., 2020) (**Table 1**).

Tobacco. Tobacco smoking-related phenotypes are the best studied among SUDs, due to its high prevalence and well-powered studies. The strongest finding in tobacco use and nicotine dependence is within the gene cluster encoding the Neuronal Acetylcholine Receptor (*CHRNA3– CHRNA5–CHRNB4*). Hancock et al., replicated this finding in two different GWASs of nicotine dependence (Hancock et al., 2015, 2018) (**Table 1**). In the 2018 GWAS of for nicotine dependence, where African-American and European populations were included, Hancock et al, (2018) found a novel variant in the DNA Methyltransferase gene (*DNMT3B*). Another GWAS for nicotine dependence found five GWS loci, providing further evidence of *CHRNA5* and *CHRNA4* risk genes, and highlighting other findings, such as the Teneurin Transmembrane Protein 2 (*TENM2*) and Dopamine β-Hydroxylase (*DBH*) (Quach et al., 2020). This study estimated an h^2_{SNP} for nicotine dependence of 8.6% (Quach et al., 2020) (**Table 1**). Many GWASs for nicotine have focused on behaviors related to tobacco smoking and smoking trajectories. The largest to date included up to 1.2 million individuals of European Ancestry assessed for four smoking related phenotypes: smoking initiation, smoking cessation, age of smoking initiation and cigarettes per day. This incredibly well-powered study found 406 risk loci associated across all stages of tobacco use, including loci in all nicotinic receptor genes (except *CHRNA7*), the Phosphatase 1 Regulatory Subunit 1B gene (*PPP1R1B*) with smoking initiation, and the *DBH* gene with cigarettes per day and smoking cessation (M. Liu et al., 2019) (**Table 1**). Furthermore, a GWAS for longitudinal smoking phenotypes, including "mostly current smoking", "mixed smoking-nonsmoking" and "mostly never smoking", identified 18 loci, some of them in novel genes, such as the Neuronal Growth Regulator 1 gene (*NEGR1*) and Cyclin and CBS Domain Divalent Metal Cation Transport

Mediator 2 gene (*CNNM2*), associated with smoking trajectories (Xu et al., 2020) (**Table 1**). The latest and largest GWAS meta-analysis on tobacco use disorder (preprint) comprising 898,680 multi-ancestry individuals, identified 88 risk loci and estimated a h^2_{SNP} of 7%. Findings included the Nicotinic Acetylcholine 466 receptor genes (*CHRNA5-A3-B4*, *CHRNB2*, *CHRNA2*, *CHRNA4*), as well as *CYP2A6* previously linked to heavy smoking, and *PDE4B* also associated to other addiction phenotypes (Toikumo et al., 2023) (**Table 1**).

Cannabis. Cannabis is the most commonly used illegal substance throughout most of the world and is becoming increasingly socially accepted, which can affect the incidence of cannabis use disorder in the population. Despite that, current GWASs for cannabis use disorder have limited power to detect robust and replicable risk loci, compared to alcohol or nicotine, due to the small sample sizes available. The first relevant GWAS for cannabis dependence found three GWS risk loci in a multi-ancestry population, with the most notable finding being in the CUB and Sushi Multiple Domains 1 gene (*CSMD1*), which has also been linked to schizophrenia (Sherva et al., 2016) (**Table 1**). A following GWAS only in European-ancestry individuals found a novel GWS loci enriched for H3K4me1 and H3K427ac histone modifications, but it could not be replicated in an independent cohort (Agrawal et al., 2018) (**Table 1**). In an effort to achieve larger sample sizes, two GWASs were performed in cannabis-use related phenotypes, including lifetime cannabis use and age at first cannabis use. The GWAS on lifetime cannabis use found eight GWS, with the top SNP in *CADM2*, a gene previously associated with alcohol consumption, and risk-taking behavior (Pasman et al., 2018) (**Table 1**). Meanwhile, a GWAS for age at first cannabis use found five GWS all in the Calcium-transporting ATPase gene (*ATP2C2*). This finding could not be replicated in an independent cohort (Minică et al., 2018) (**Table 1**). Two large GWASs directly assessing cannabis use disorder have been performed so far. The first one found a relevant GWS association in the Cholinergic Receptor Nicotinic α2 Subunit (*CHRNA2*), specifically with an under-expression of this gene in the cerebellum of cannabis use disorder individuals (Demontis, Rajagopal, et al., 2019) (**Table 1**). The latest GWAS on cannabis use disorder (N = 20,916 cases and 363,116 controls) found two GWS loci, the first locus located in the novel Forkhead Box Protein P2 gene (*FOXP2*) and the second one near *CHRNA2* – supporting the evidence from the previous GWASs

- and Epoxide Hydrolase 2 gene (*EPHX2*). This study estimated an h^2_{SNP} for cannabis use disorder of 6.7% (E. C. Johnson, Demontis, et al., 2020) (**Table 1**).

Opioids. The interest in performing large GWASs for opioid use disorder has increased over the past few years due to the increasing cases of opioid overdose in the USA, which has been denominated the "opioid epidemic". The first GWASs of opioid use disorder yielded GWS loci in the Repulsive Guidance Molecule BMP Coreceptor A gene (*RGMA*) and in the Cornichon Family AMPA Receptor Auxiliary Protein 3 gene (*CNIH3*) (Cheng et al., 2018; Nelson et al., 2016) (**Table 1**). Furthermore, the first large-scale opioid use disorder GWAS observed a GWS variant in the μ-opioid receptor *OPRM1* gene. This gene is considered the main biological target of opioids. This study also estimated a h^2_{SNP} of 11% (Zhou, Rentsch, et al., 2020) (**Table 1**). Another GWAS compared opioid dependent individuals, opioid-exposed controls and opioid-unexposed controls (Polimanti et al., 2020). There were some genetic differences, including the association of the Solute Carrier Family 30 Member 9 gene (*SLC30A9)* and the BEN Domain Containing 4 gene (*BEND4)* with opioid use (exposed vs. unexposed controls GWAS), but not with the opioid dependence (opioid dependence vs. exposed or unexposed controls GWAS) (**Table 1**). These analyses highlighted the difference between dependence and exposure and the importance of considering the definition of controls (exposed versus unexposed) in SUD studies. Two large opioid use disorder GWASs have been conducted in the past year (Deak, Zhou, et al., 2022; Kember et al., 2022). Deak et al., found three GWS variants in a European population of 639,063 individuals, one in the Paired Basic Amino Acid Cleaving Enzyme gene (*FURIN*) and the other two in *OPRM1* (Deak, Zhou, et al., 2022). Moreover, Kember et al., identified ten independent GWS loci in a multi-ancestry GWAS of 425,944 individuals of a less stringent opioid use disorder diagnosis, and one additional loci when considering only a stringent opioid use disorder definition. Findings replicated previous associations with *OPRM1* and *FURIN*, as well as identified novel loci in the Rab9 Effector Protein With Kelch Motifs gene (*RABEPK*) and the Neural Cell Adhesion Molecule 1 gene (*NCAM1*), among others (Kember et al., 2022). Moreover, the estimated h^2_{SNP} in both GWAS was of 12-13% (Deak, Zhou, et al., 2022; Kember et al., 2022) (**Table 1**).

Cocaine. Cocaine use disorder is notably the SUD with the lower sample sizes in GWASs. The first GWAS on cocaine dependence identified a GWS variant in the Family

with Sequence Similarity 53, Member B gene (*FAM53B*) (Gelernter et al., 2014) (**Table 1**). Another GWAS performed a few years later did not identify GWS risk loci (Cabana-Domínguez et al., 2019). However, it identified the gene *HIST1H2BD* in a gene-based analysis, previously associated to schizophrenia (**Table 1**). The largest GWAS to date (N = 9,965) used cluster analyses to identify cocaine use disorder subtypes with reduced phenotypic heterogeneity (Sun et al., 2020). Five clusters were identified, which were then used as traits for the GWAS. From the 13 GWS loci identified, three loci were replicated in a replication sample, located in the Trafficking Kinesin-Binding Protein 2 gene (*TRAK2*), the Latrophilin 2 gene (*LPHN2*) and the Transmembrane Protein 51 gene (*TMEM51*). This study, though, could not replicate the previous *FAM53B* finding (Sun et al., 2020) (**Table 1**).

SUD. Recent efforts have been made to explore the common underlying genetic architecture for all SUDs, that is a general addiction genetic factor that conveys vulnerability to multiple SUDs. Hatoum et al., (2022) combined the latest GWASs summary statistics for problematic alcohol use, problematic tobacco use, cannabis use disorder and opioid use disorder into a multi-ancestry sample of 1,118,180 individuals and defined "The Addiction-Risk-Factor" (addiction-rf) as the unidimensional shared genetic liability between all SUDs, and independent of substance use. Next, a multivariate GWAS meta-analysis was performed to disaggregate general (addiction-rf) and substance-specific (alcohol, nicotine, cannabis and opioids) loci from published summary statistics of these same SUDs (problematic alcohol use, problematic tobacco use, cannabis use disorder and opioid use disorder) in a mixed-ancestry sample of 1,118,180 individuals (Hatoum et al., 2023). The addiction-rf was associated with 17 GWS risk loci, with the top finding in *DRD2*, and other findings highlighting *FTO*, *PDE4B*, *GTF3C2*, *ZNF512*, *ADH1C* and *SIX3* (**Table 1**). In addition, substance-specific associations highlighted six GWS loci for problematic alcohol use with the top signal in *ADH1B*, 12 loci for problematic tobacco use with the top finding being *CHRNA5*, five loci for cannabis use disorder with the top signal in *FAM19A5*, and one loci for opioid use disorder in *OPRM1*.

Table 1. Largest genome-wide association studies available for main phenotypes related to SUD

Study (author, year)	SUD outcome	Description	Population	Sample size (case/control)	GWS loci	Relevant genes	h^2_{SNP} (%)
Alcohol							
Gelernter, 2014	alcohol dependence	DSM-IV alcohol dependence diagnosis derived from the semi-strucutred assessment for drug dependence and alcoholism (SSADDA)	African-Americans and European-Americans	16087	2	*ADH1B, ADH1C,*	-
Adkins, 2015	alcohol consumption	Mean of drinks per week repeated assessments collected across adolescence and the transition to adulthood (maximum age range: 12–21 years)	NS	2126	4 (FDR p<0,1)	*GABA transporter SLC6A1*	-
Schumann, 2016	alcohol consumption	Drinks per week variable was dichotomized into heavy drinking (> 14-21) or light drinking (<14-21)	European	105898	5 (p<1x10e-6)	*KLB*	-
Clarke, 2017	Alcohol consumption	Self-reported current drinking status and drinks per week	European	112,117	14	*ADH1B/ADH1C/ADH5, KLB, CADM2, DRD2, PDE4B*	13 ± 1
Walters, 2018	Alcohol dependence	DSM-IV alcohol dependence diagnosis (DSM-IV and DSM-IIIR)	African-Americans and European-Americans	14,904 / 37,994	1	*ADH1B*	9 ± 1.9
Sanchez-Roige, 2019	Alcohol use disorder	AUDIT, total score, AUDIT-C score (consumption) and AUDIT-P score (problematic alcohol use)	European	141.932	10	*ADH1B, ADH1C, KLB, GCKR, JCAD, SLC39A13*	1.2 ± 0.48
Kranzler, 2019	Alcohol consumption	Alcohol consumption measured with the AUDIT-C scores, comprising the firts 3 items of the AUDIT	Multi-ancestry	272.842	13	*GCKR, KLB, ADH1B, ADH1C, SLC39A8*	6.8 ± 0.5
Kranzler, 2019	Alcohol use disorder	Alcohol use disorder diagnosis based on ICD-9/10 codes for abuse and dependence	Multi.ancestry	55,584 / 218,807	10	*ADH1B, ADH1C, ADH4, DRD2, FTO, GCKR, SLC39A8*	5.6 ± 0.4
Liu, 2019	Drinks per week	Average number of drinks a participant reported drinking each week	European	941.280	81	*ADH1B, DRD, GRIK2, CRHR, UCN*	4.2 ± 0.2
Zhou, 2020	Problematic alcohol use	Meta-analysis of cohorts assessed using different phenotypic definitions: alcohol dependence and alcohol use disorder diagnoses with ICD codes and DSM-IV, and AUDIT- P.	European	435.563	29	*GCKR, SIX3, KLB, ADH1B, ADH1C, SLC39A8, DRD2, FTO, CADM2, THSD7B, SLC39A13*	6.8 ± 0.4
Deak, 2022	Maximum alcohol intake	Maximum alcohol intake was estimated from MVP lifestyle survey quesiton: in a typical month, what is/was the largest number of drinks of alcohol you may have had in one day?	African-Americans and European-Americans	247.755	15	*ADH1B, SLC39A8, CRHR1, AGBL2, GCKR, DRD2*	6.5 ± 0.41
Tobacco smoking							
Hancock, 2015	Nicotine dependence	Nicotine dependence defined by the FTND, cateorized into mild, moderate and severe.	European	24.543	2	*CHRNA5-CHRNA3-CHRNB4, CHRNA4*	Not reported
Hancock, 2018	Nicotine dependence	Nicotine dependence defined by the FTND, cateorized into mild, moderate and severe.	African-Americans and European-Americans	38.602	2	*CHRNA5-CHRNA3-CHRNB4, DNMT3B*	Not reported
Liu, 2019	Age of Initiation of Regular Smoking	Age at which an individual started smoking cigarettes regularly	European	341.427	10	*GRK4*	4.7 ± 0.3
Liu, 2019	Cigarretes per day	Average number of cigarettes smoked per day	European	337.334	40	*CYP2A6, DRD2, DBH, CHRNA5, GRK4*	8 ± 0.8

Table 1. Continuation

Study (author, year)	SUD outcome	Description	Population	Sample size (case/control)	GWS loci	Relevant genes	h^2_{SNP} (%)
Liu, 2019	Smoking cessation	Binary phenotype with current smokers coded as cases and former smokers coded as controls	European	139,453 / 407,766	16	*CYP2A6, DRD2, DBH, CHRNA5*	4.6 ± 0.2
Liu, 2019	Smoking initiation	Binary phenotype with ever regular smokers coded as cases and never regular smokers coded as controls	European	557,337 / 674,754	259	*PPP1R1B, GRIN2A, HOMER2, GRM1*	7.8 ± 0.2
Quach, 2020	Nicotine dependence	Nicotine dependence defined by the FTND, cateorized into mild, moderate and severe.	European	58.000	5	*MAGI2/GNAI1, TENM2, DBH, CHRNA5, CHRNA4*	8.6 ± 1.2
Xu, 2020	Smoking trajectory	Smoking trajectory phenotype accounts for variation in smoking status over time: mostly current smoking, mixed smoking and nonsmoking, mostly never smoking	Multi-ancestry	286.118	18	*CHRNA2, DBH, DRD2, RIMS1*	018 ± 1 (current vs. never) 5.8 ± 0.5 (current vs. mixed)
Sylvanus, 2023	Tobacco use disorder	Tobacco use disorder assessed with ICD codes	Multi-ancestry	898.680	88	*CHRNA5-A3-B4, CHRNB2, CHRNA2, CHRNA4, CYP2A6, PDE4B, GPX, GMPPB, NT5C2*	5 ± 2
Cannabis							
Sherva, 2016	Cannabis dependence	Criterion count for DSM-IV cannabis dependence	African-Americans and European-Americans	14.754	3	*SLC35G1, CSMD1, RP11*	Not reported
Agrawal, 2018	Cannabis dependence	Cases were defnied as meeting criteria for DSM-IV cannabis dependence, and controls were exposed individuals without a lifetime history of cannabis dependence	European	2,080 / 6,436	1	*H3K4me1, H3K427ac*	Not reported
Pasman, 2018	Lifetime cannabis use	Self-report data on whether the participant had ever used cannabis during their lifetime	European	184.765	8	*CADM2, ZNF704, SDK1, NCAM1, RABEP2/ATP2A1, SMG6*	11 ± 1
Minica, 2018	Age at first cannabis use	Age at first cannabis use was assessed from questionnaires or clinical interviews	European	24.953	5	*ATP2C2*	Not significant
Demontis, 2019	Cannabis use disorder	Cannabis use disorder diagnosis based on ICD10 codes	European	2,387 / 48,985	1	*CHRNA2*	9 ± 3
Johnson, 2020	Cannabis use disorder	Cannabis use disorder diagnosis based on DSM- IV or DSM-III- R criteria or ICD-10 codes	African-Americans and European-Americans	20,916 / 363,116	2	*FOXP2, CHRNA2, EPHX2*	6.7 ± 0.6
Opioids							
Nelson, 2016	Opioid dependence	Participants were divided into opioid dependent daily injectors as cases, and opioid users with impeded progression as controls (including non-dependent opioid misuers and opioid dependent without a history of daily injections)	European	1,167/161	1	*CNIH3*	Not reported

Table 1. Continuation

Study (author, year)	SUD outcome	Description	Population	Sample size (case/control)	GWS loci	Relevant genes	h^2_{SNP} (%)
Cheng, 2018	Opioid dependence	Opioid dependence was assessed with DSM-IV and the phenotype was defnied as the sum of DSM-IV criterion counts for all Opioid dependence-exposed individuals	European	3.058	1	*RGMA*	Not reported
Zhou, 2020	Opioid use disorder	Opioid use disorder diagnosis and opioid exposure based on ICD-9 and ICD-10 codes	African-Americans and European-Americans	15,756 / 99,039	1	*OPRM1*	11.3 ± 1.8
Polimanti, 2020	Opioid dependence	Opioid dependence diagnosis based on DSM-IV criteria. Two control groups were defined: individuals without opioid dependence diagnosis who were exposed to opioids at least once, and unexposed controls	African-Americans and European-Americans	4,503 / 4,173 / 17,700	0/1	*C18orf32*	28 ± 1 0 (OD unexposed)
Polimanti, 2020	Opioid use	Cases were defied as individuals without opioid dependence diagnosis who were exposed to opioids at least once, and controls were unexposed individuals	African-Americans and European-Americans	4,173 / 32,500	1	*SLC30A9, BEND4*	Not significant
Sanchez-Roige, 2021	Problematic opioid use	Problematic opioid use was defined as using prescription opioids 'not as prescribed'	European	27,805 / 104,246	2	*KDM4A, LRRIQ3*	4 ± 1
Deak, 2022	Opioid use disorder	Opioid use disorder diagnosis based on ICD codes or DSM-IV depending on the cohort	African-Americans and European-Americans	20,686 / 618,377	3	*FURIN , OPRM1*	12.8 ± 1.1
Kember, 2022	Opioid use disorder	Opioid use disorder diagnosis based on ICD codes (less stringent definition: only 1 ICD code required)	Multi-ancestry	31,473 / 394,471	14	*CDKAL1, BTNL2, OPRM1, RABEPK, FBXW4, NCAM1, FURIN, KCNN1, RNF114*	12 ± 1
Cocaine							
Gelernter, 2014	Cocaine dependence	Phenotyope was defnied as symptom count for cocaine dependence diagnosis based on DSM-IV criteria	African-Americans and European-Americans	5.697	1	*FAM53B, NCOR2, CDK1*	Not reported
Cabana-Dominguez, 2019	Cocaine dependence	Cocaine dependence diagnosis bases on DSM-IV criteria	European	2,085 / 4,293	0	*HIST1H2BD*	30 ± 6
Sun, 2020	Cocaine use disorder	Cocaine use disorder subtypes were derived by cluster analysis using 25 questions from the SSADDA: 5 cocaine use disorder subtypes were evaluated	African-Americans and European-Americans	9.965	13	*LPHN2, FN1, TENM3, TRAK2, LINC01411*	Not reported
SUD							
Hatoum, 2023	The addiction risk factor	Multivariable meta-analysis combining problematic alcohol use, problematic tobacco use, cannabis use disorder and opioid use disorder	African-Americans and European-Americans	1.118.180	17	*DRD2, FTO, PDE4B, GTF3C2, ZNF512, ADH1C, SIX3*	Not reported

DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; SSADDA: Semi-structured Assessment for Drug Dependence and Alcoholism; AUDIT: Alcohol Use Disorders Identification Test; MVP: Million Veteran Program; FTND: Fagerström Test for Nicotine Dependence.

3.1.2. Post-GWAS Analysis

The results from GWASs can be used as the bases for a wide range of analytical approaches, starting by the discovery of genes and biological pathways implicated in the pathology of complex traits. Within this thesis, findings obtained from GWASs were utilized to gain insights into the underlying genetic architecture and the complex interplay between correlated traits, employing three main post-GWASs methodologies: genome-wide polygenic score (PGS), genetic correlation and Mendelian randomization (MR) analyses (**Figure 11**).

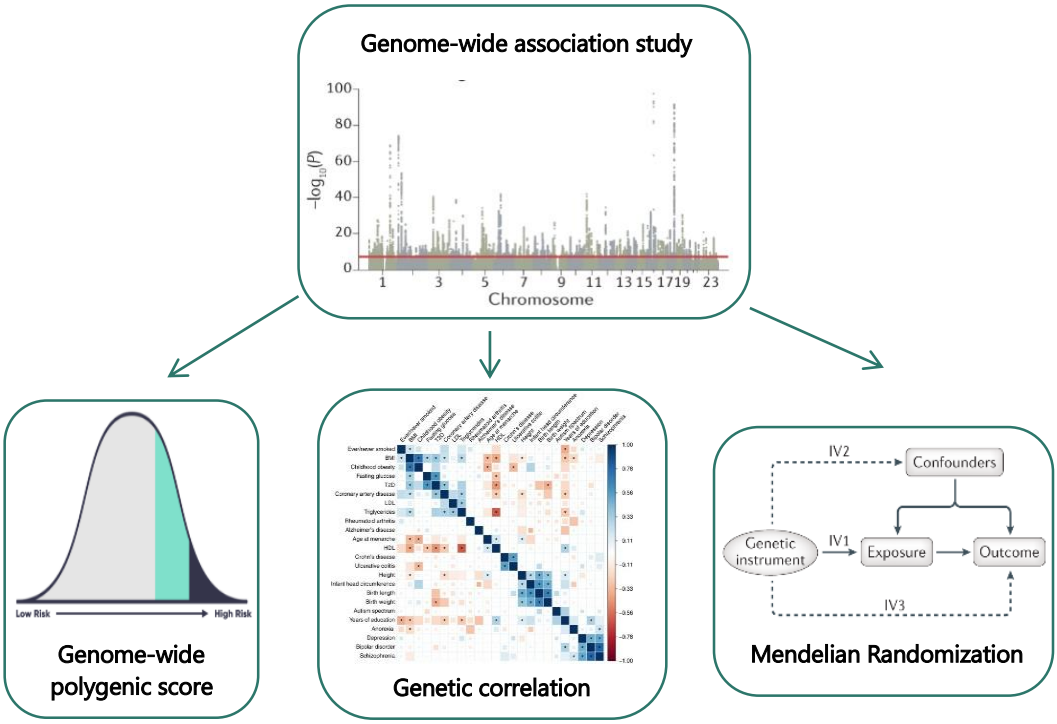


Figure 11. Overview of post-GWAS analyses employed in this thesis. GWASs data sets can be used as the basis for multiple analytical approaches, such as genome-wide polygenic score (PGS), genetic correlation and Mendelian randomization (MR) analyses, to understand the mechanisms underlying the associations observed between complex traits. *Adapted from Gelernter et al. (2021).*

3.1.2.1. Genetic Correlation

Information from two GWASs summary statistics can be used to estimate the correlation in allele effects between two traits across the genome (van Rheenen et al., 2019). This parameter is defined as the genetic correlation (rg) between two traits. In

other words, genetic correlation quantifies the *pleiotropy* between two traits, understanding pleiotropy as the ability of a genetic locus to affect more than one trait (**Figure 12**). Pleiotropy between two traits can reflect different methods of action, mainly horizontal pleiotropy, where the genetic variant contributes directly to the risk of both traits, and vertical pleiotropy where there is a causal relationship between the two traits and the direct effect of the genetic variant on the first trait generates a response on the second trait. Genetic correlation can capture these two forms of pleiotropy, so the underlying architecture of correlations at the individual genetic variant can vary (**Figure 12**).

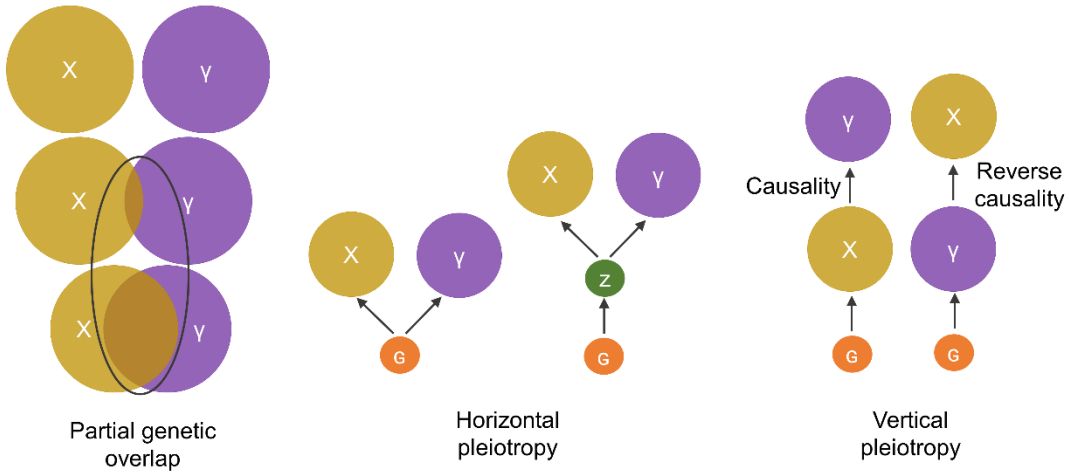


Figure 12. Different mechanisms of pleiotropy between two diseases. X and Y correspond to two different traits or diseases; G=genetic variant; Z=intermediate phenotype. *Adapted from van Rheenen et al. (2019).*

Genetic correlation of SUD and related traits: Estimating the genetic correlation across multiple disorders and/or traits has become a standard practice after any SUD-related primary GWAS. Although SNP-based heritability is still low based on current data, these analyses have helped to elucidate the consistent genetic overlap between SUDs and a wide range of psychiatric disorders and behavioral traits.

As expected, SUDs are generally correlated with each other, for instance, alcohol use disorder shows strong genetic correlation with smoking initiation, nicotine dependence and lifetime cannabis use, and vice versa (Deak, Zhou, et al., 2022; E. C. Johnson, Demontis, et al., 2020; Quach et al., 2020; Zhou, Sealock, et al., 2020). In addition, genetic correlation analysis from all the SUD-specific GWASs previously mentioned (**Table 1**) consistently show positive genetic correlations with psychopathology, with the strongest

overlap observed for major depressive disorder, anxiety, ADHD, bipolar disorder, PTSD and schizophrenia, as well as with behavioral traits such as neuroticism and risk taking. Conversely, SUDs show negative genetic correlations with cognitive traits, the most relevant being years of education and intelligence (Abdellaoui et al., 2021; Cabana-Domínguez et al., 2019; Deak, Zhou, et al., 2022; E. C. Johnson, Demontis, et al., 2020; Quach et al., 2020; Zhou, Sealock, et al., 2020).

Intriguingly, this pattern of correlations is not always consistent when looking into quantity/frequency substance use phenotypes. For instance, alcohol consumption frequency showed negative correlation with major depressive disorder and positive correlation with years of education (Kranzler et al., 2019). This pattern was not seen when alcohol consumption quantity (maximum habitual alcohol intake) was assessed, which was more genetically similar with alcohol use disorder and psychopathology (Deak, Levey, et al., 2022). Even though alcohol consumption is a necessary component for alcohol use disorder, seen by the presence of substantial genetic correlation between them (rg =0.52) and the high overlap of associated risk genes, these findings suggest that the genetic architecture they share with other psychiatric and behavioral traits differs completely between them. It is likely that GWASs for alcohol consumption frequency are capturing socially accepted drinking behaviors, which may be influenced by sociodemographic status.

Moreover, smoking-related phenotypes, which measure quantity and frequency of use, showed modest genetic correlations with psychopathology (M. Liu et al., 2019). Smoking initiation was found to be only modestly genetically correlated with nicotine dependence (rg = 0.40), while cigarettes per day showed a highly positive genetic correlation with nicotine dependence (rg = 0.95). This suggests that, similarly to alcohol drinking-related phenotypes, traits assessing quantity of use, rather than frequency, are genetically more similar to dependence phenotypes (Quach et al., 2020). Furthermore, lifetime cannabis use only showed modest genetic correlation with cannabis use disorder (rg = 0.50) and weaker genetic relationships with educational attainment and body mass index (E. C. Johnson, Demontis, et al., 2020; Pasman et al., 2018).

The latest GWAS on the addiction-rf replicated the genetic correlation between SUD and all stage-based facets of addiction: risk-taking (binge/intoxication), executive function (preoccupation/anticipation), and neuroticism (negative affect) (Hatoum et al., 2022). It was also strongly genetically correlated to suicide attempt, externalizing behaviors, self-medication, unemployment, maternal smoking around birth and age first had sexual intercourse, among others (Hatoum et al., 2023) (**Figure 13**).

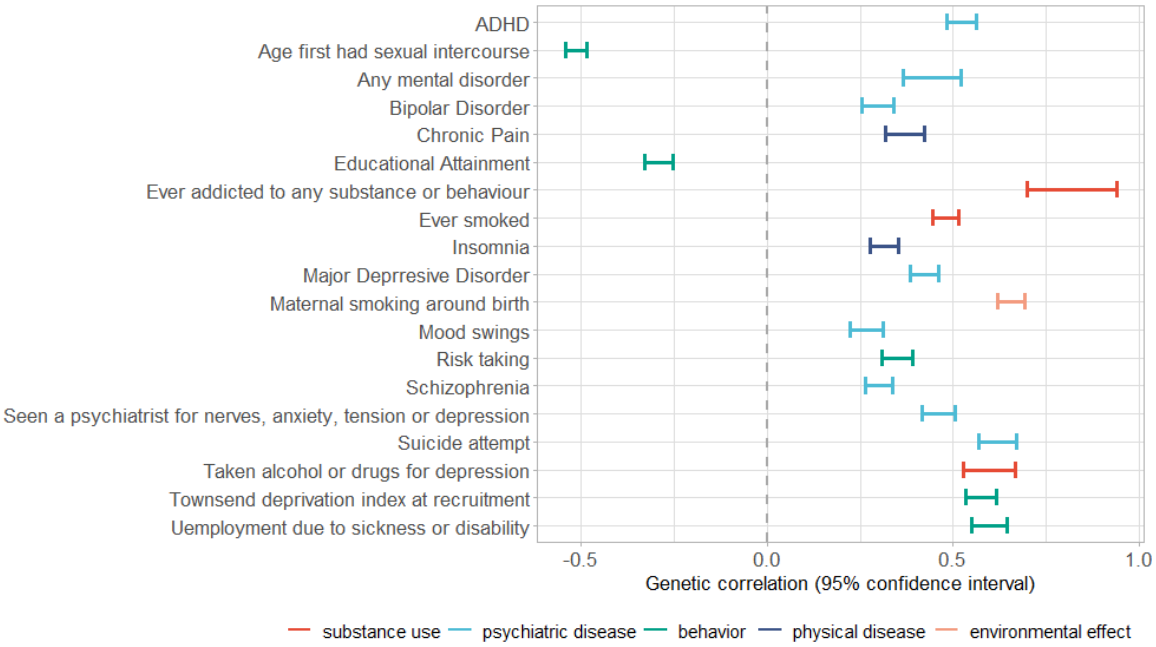


Figure 13. Genetic correlation of the addiction-rf and relevant phenotypes. List of selected relevant phenotypes showing a high genetic correlation with the addicition-rf. The 95% confidence interval of the genetic correlation estimates were obtained from Hatoum et al. (2023).

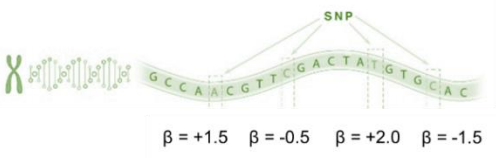
3.1.2.2. Genome-Wide Polygenic Scores

PGSs - also referred to as polygenic risk scores (PRSs) in the context of disease - are the quantitative measure of the total genetic burden of a trait and/or disease over multiple susceptibility variants (Chatterjee et al., 2016). These scores are calculated for individuals in a *target sample* by summing up the number of risk alleles they carry weighted by their effect size estimated in a large-scale GWAS from an independent sample (*discovery sample*) (**Box 5**). The computation of PGSs was initially introduced in the context of psychiatric diseases. It was based on the premise that, due to their high polygenicity, insufficient sample sizes in early GWASs produced few robust associations, but the aggregation of many loci below the GWS threshold could significantly predict

disease risk in new studies (Wray et al., 2007, 2014). In order to do so, several methods can be used to calculate PGSs, depending on how the SNPs included in the analysis are selected. One approach is to select SNPs based on pre-defined p-value thresholds of association in the GWAS summary statistics (e.g., p-value <.00001, .0001, .001, .01, .1). Alternatively, genome-wide SNPs can be included, which requires accounting for the LD structure between variants and to re-estimate SNPs weights.

Box 5. Overview of the steps for calculating PGSs

Step 1: GWAS summary statistics are obtained from a discovery sample, which detail the effect size of each SNP on the phenotype of interest.



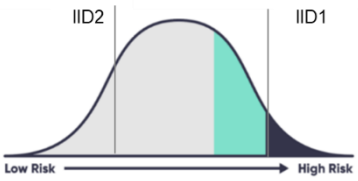
Step 2: Genotype data for a target sample are referenced against GWAS summary statistics.

IID	SNP1	SNP2	SNP3	SNP4
IID1	AT	CG	TT	AA
IID2	TT	CC	GT	AC

Step 3: PGSs can be calculated for each individual in the target sample by summing up the number risk alleles weighted by the effect size from the discovery summary statistics.

IID	SNP1	SNP2	SNP3	SNP4	Total score
IID1	+ 1.5	- 0.5	+ 4.0	- 0.0	**= 5.0**
IID2	+ 0.0	- 1.0	+ 2.0	- 1.5	**= -0.5**

Step 4: Association analysis is performed to assess the effect of the PGSs on the outcome measure.



From a clinical perspective, PGSs can be used to identify individuals at high risk of disease for prevention and early intervention. In current genetic research, and in this thesis, PGSs are a tool to quantify the genetic liability for a trait explained by GWASs associated SNPs and to assess the association with that trait in an independent sample. PGS studies also allow cross-phenotype analysis, were the phenotype of the discovery sample, on which the GWAS is conducted, differs from the phenotype of the target sample used to construct the PGSs, allowing the exploration of the genetic overlap among related traits and diseases (Wray et al., 2014).

Furthermore, PGSs can be utilized in *phenome-wide association studies* (PheWASs) to conduct a comprehensive examination of the genetic liability for a single or multiple trait(s), represented by PGSs, with a wide range of phenotypes or traits.

PGS studies of SUDs and related traits. SUDs PGSs show limited predictive power on an individual basis, meaning they are not (yet) good at predicting disease risk in an individual. However, they can be used to understand the genetic overlap between different SUDs and a wide range psychiatric and behavioral traits (Lewis & Vassos, 2020).

A PheWAS of PGSs for alcohol and opioid use disorder, smoking initiation and lifetime cannabis use was conducted in a deeply phenotyped sample, to examine the patterns of pleiotropy of these four PGSs with multiple phenotypic domains (Kember et al., 2023). This study replicated known and identified novel phenotypic associations between SUDs and major depressive disorder, poor school performance, PTSD, lifetime trauma assessment and family history of SUDs (Kember et al., 2023).

Moreover, the addiction-rf PGS was found to be associated with various medical conditions, including psychiatric illnesses, self-harming behaviors, and somatic diseases, such as chronic pain, as well as polysubstance use disorder and individual SUDs (Hatoum et al., 2023). A following PheWAS revealed correlations between the addiction-rf and maternal tobacco smoking during pregnancy and ADHD, consistent with existing evidence suggesting that the impacts of the prenatal environment may be influenced by inherited risk genes. Moreover, it was associated with family history of serious mental illness, and indices of socioeconomic status and disability, further supporting the association between environmental risk factors and common genetic effects on SUDs liability (Hatoum et al., 2023). In substance-naïve children, the addiction-rf PGS was positively associated with parental substance use problems, externalizing behavior, fun-seeking behavior, family history of SUDs and hospitalization, childhood externalizing behaviors, and socioeconomic disadvantage (Hatoum et al., 2023).

On the other hand, studies have also examined the association between PGSs for psychiatric disorders and substance use or dependence. A study found significant associations between PGSs for individual psychiatric diagnoses and substance-specific use or dependence, including major depressive disorder PGS with cannabis use and

cocaine dependence, and schizophrenia PGS with cannabis use cannabis dependence and cocaine dependence (Carey et al., 2016). A recent study revealed that, across six different psychiatric disorders previously associated with SUD, schizophrenia PGS explained the highest variance in SUD genetic susceptibility (Gurriarán et al., 2019), which aligns with the reports from epidemiological and genetic correlation studies (Lähteenvuo et al., 2021).

Within this thesis, the association between SUDs and ADHD holds particular interest. Genetic correlations and PGS analyses support a shared genetic background between ADHD and SUDs, reporting evidence of positive genetic correlations between ADHD and different SUDs (Hatoum et al., 2023; Pasman et al., 2018; Walters et al., 2018) In addition, Wimberley et al., (2020) found that high ADHD PGS increased the risk of any SUD in an ADHD cohort, although ADHD still explained a minor proportion of the variance in SUDs. In another study, ADHD PGS was significantly associated with alcohol frequency intake, alcohol dependence and ever smoking (Du Rietz et al., 2018). Moreover, ADHD PGS also showed suggestive associations with cannabis use and nicotine dependence (Carey et al., 2016).

3.1.2.3. Mendelian Randomization

MR is a technique used to investigate causal relationships between exposures and outcomes by leveraging genetic variants as instrumental variables (Sanderson et al., 2022). This approach uses genetic instruments associated with the exposure to estimate causal effects not affected by the presence of unobserved confounding (Uffelmann et al., 2021). The rationale of MR is similar to that of randomized control trials, where the random assignment of a treatment between two groups allows the evaluation of its effect on an outcome avoiding the effect of potential confounding factors. In MR, genetic variants serve as a naturally occurring form of randomization (Sanderson et al., 2022) (**Box 6**).

There are three assumptions each genetic variant needs to meet to be considered a valid instrument for causal interference and effect estimation: First, it should be robustly associated with the exposure of interest. Secondly, it should not directly affect the outcome except through its influence on the exposure. Thirdly, it should not be

associated with the outcome due to confounding factors (Burgess et al., 2020) (**Figure 14**). When multiple genetic variants that fulfil these assumptions can be identified, the statistical power to estimate true causal effects improves by increasing the proportion of the exposure variance explained by the instruments (Burgess et al., 2013).

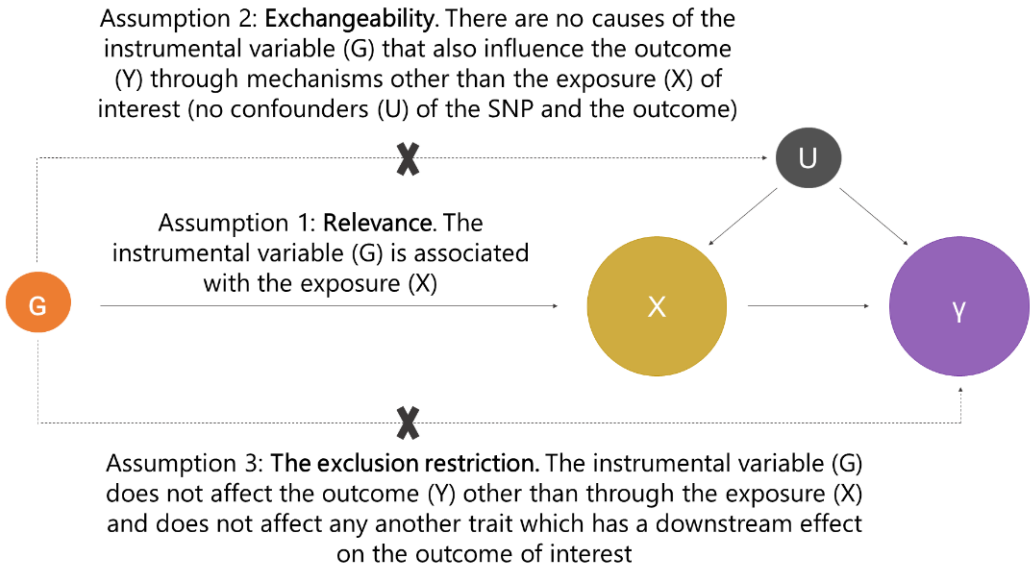


Figure 14. Mendelian randomization model and three key assumptions of a Mendelian randomization analysis. G: genetic variant; X: exposure; Y: outcome; U: confounder.

MR studies can be conducted using either individual-level data or, more commonly, summary data obtained from GWASs. While individual-level data includes genetic and phenotype measures for each participant, summary data estimates the association between the SNP(s) and both the exposure and the outcome traits separately (Sanderson et al., 2022). When using summary data, the association between the SNP(s) and the exposure can be estimated in an independent sample, distinct from the one used to estimate the effect of the SNP(s) on the outcome, which is referred to as "two-sample MR" (Burgess et al., 2013).

In any MR analysis, a range of sensitivity analyses need to be employed to test for violations of any the three assumptions, which can invalidate MR results (Burgess et al., 2020). The first assumption can be tested through the association between the SNPs and the exposure. If the genetic instruments are not strongly associated with the exposure, then weak instrument bias can be introduced into the MR estimation (Davies et al., 2015).

Although the second and third assumptions cannot be proven to be true, there are a variety of methods that focus on detecting and accounting for horizontal pleiotropy - a violation of the third assumption – which occurs when genetic instruments have a direct effect on the outcome or on another phenotype with a main effect on the outcome (Katikireddi et al., 2018). An important limitation of MR studies is the presence of unmeasured confounding of the genetic variants and the outcome, which violates the second assumption, as current MR sensitivity analysis do not account or correct completely for this potential bias. Therefore, it is valuable to interpret MR findings within a triangulation of evidence framework, considering results from complementary approaches that rely on different assumptions (Munafò et al., 2021).

Box 6. Principles of the Mendelian randomization approach

The MR approach utilizes Mendel's laws of genetic inheritance as its foundation, specifically the law of segregation and the law of independent assortment. According to the law of segregation, offspring randomly inherit one allele from each parent at every point in the autosomal genome. The law of independent assortment states that these alleles are passed down to offspring independently of each other, except in regions of the genome that are genetically linked.

The concept of MR was initially described in the context of family-based studies, drawing an analogy to randomized controlled trials by comparing the random allocation of genetic variants from parents to their children. However, due to limited availability of family-based data, population-based MR studies have been introduced as an alternative. The rationale behind population studies is that genetic variants can identify groups within the population that differ on average in terms of an exposure. In these studies, group distribution based on genetic variation is assumed to be unrelated to confounding factors such as behavioral, social, and physiological exposures occurring after conception. Therefore, genetic associations between traits are expected to be free from confounding, and any differences in outcomes observed between groups defined by genetic variation can be attributed to the exposure, assuming no selection bias arising from that genetic variation.

Adapted from Sanderson et al. (2022).

Mendelian Randomization analyses of SUDs. In recent years, MR has been increasingly applied to understand the causal relationship between SUDs and related traits. Since the power to detect robust MR results rely on the number of genetic instruments available, some of the most relevant findings have been made for smoking and alcohol-related traits, which present the largest GWASs sample sizes.

For instance, there are some evidence that the genetic liability for neuroticism and extraversion are causally linked to smoking heaviness (cigarettes per day) and smoking initiation, respectively (Sallis et al., 2019). Moreover, a few studies report that the genetic liability for major depressive and bipolar disorder have a bidirectional relationship with lifetime smoking and smoking initiation, meaning that there is evidence of causal effects in both directions (Vermeulen et al., 2021; Wootton et al., 2020). Having a mood disorder diagnosis can lead to smoking-behaviors, and smoking can also increase the risk of mood disorders. On the other hand, for schizophrenia, this relationship has only shown strong evidence in the causal direction of smoking (initiation and lifetime) on the risk for schizophrenia (Wootton et al., 2020), but not in the other direction (Gage, Jones, Taylor, et al., 2017). However, evidence of bidirectional causal effects was observed between schizophrenia and cannabis use, adding to the observational evidence linking these two conditions (Gage, Jones, Burgess, et al., 2017; Vaucher et al., 2018). Moreover, major depressive disorder has shown to have causal effect on the risk for alcohol use disorder but not on alcohol frequency or quantity (Polimanti et al., 2019), which is consistent with the genetic correlation findings discussed earlier.

In the context of ADHD, current MR studies have shown evidence of the causal effect of ADHD on smoking initiation and severity (cigarettes per day, and impaired smoking cessation), cannabis use initiation and alcohol use disorder (Fluharty et al., 2018; Soler Artigas et al., 2020; Treur et al., 2021). Moreover, smoking has also shown to increase the risk for ADHD, which is in line with previous literature indicating that smoking can have detrimental, long-term effects on attention (Treur et al., 2021). However, MR analyses present several limitations related to the strong assumptions they rely. Therefore, the interpretation of MR findings in the context of SUD and other complex traits should be considered carefully.

3.2. Environmental Factors and Gene-Environment Interactions in the Etiology of SUDs

As evidenced, common genetic variation has an important role in the etiology of SUDs, but, to date, its contribution to the total variance in SUDs remains relatively small, accounting for less than or slightly over 10% of the heritability (Deak & Johnson, 2021). Twin and family studies provide evidence supporting the involvement of environmental factors in the etiology of SUDs. These studies show that, during adolescence, the initiation of substance use is predominantly influenced by shared environmental factors, while genetic factors and unique environmental factors contribute to a lesser extent. However, at later age, and with more severe measures of substance use, the individual predisposition is mostly explained by genetic factors and unique environmental factors (Kendler et al., 2003; Prom-Wormley et al., 2017).

Numerous environmental factors play a critical role in the onset of SUDs. A recent study reported that the extent of illicit substance use in an individual's birth cohort was associated with significantly increased risk of substance use initiation, even after controlling for individual-level substance use history (Degenhardt, Bharat, Glantz, Sampson, Al-Hamzawi, et al., 2019). Moreover, youth experiencing psychosocial problems, including problems within the family, peer relationships, delinquency and school-related problems exhibit earlier age at onset of substance use (Poudel & Gautam, 2017; Van Den Bree & Pickworth, 2005). Similarly, in a meta-analytical review, Elliott et al. reported that family history of alcohol use disorder does not impact on the amount of drinking overall but is associated with higher rates of problematic alcohol use (Elliott et al., 2012). Having a history of maltreatment during childhood or other forms of early life stress are also risk factors for an earlier onset of alcohol drinking and substance use initiation in adolescence with progression toward heavy substance use and increased risk for SUDs (Capusan et al., 2021; Cicchetti & Handley, 2019; Dube et al., 2006; Kirsch & Lippard, 2022). Moreover, having a family history of a SUD or having experienced childhood maltreatment are risk factors associated with the persistence of the disorder (Chassin et al., 2004; Elliott et al., 2014, 2016).

In addition, current research provides evidence that SUDs etiology, as many psychiatric disorders, results from the interplay between genetic and environmental factors, rather than from their independent main effects (Dick & Kendler, 2012; Pasman et al., 2019). These include gene-environment interaction (GxE) and gene-environment correlation (rGE). Gene-environment interaction refers to the moderation of the genetic predisposition as a consequence of the environmental exposure (Agrawal et al., 2012). Under this premise, genetic factors underlie biological mechanisms that make a person more or less vulnerable to environmental circumstances (Belsky & Pluess, 2009). For example, genetic influences on adolescent substance use are enhanced in environments with lower parental monitoring and easy availability of alcohol (Dick & Kendler, 2012). Alternatively, gene-environment correlation refers to the genetic predispositions that influence the likelihood of being exposed to a certain environment (Jaffee & Price, 2007). For example, genetic factors linked to socialization in childhood have been found to influence the risk for substance abuse in late adolescence (Hicks et al., 2013). However, extensive data on environmental factors are often lacking in large genetic studies. The inclusion of environmental factors in genetic studies will not only give insight into the underlying biological mechanisms of SUDs, but will also characterize subgroups (based on these environmental factors) at high risk for addictive behaviors.

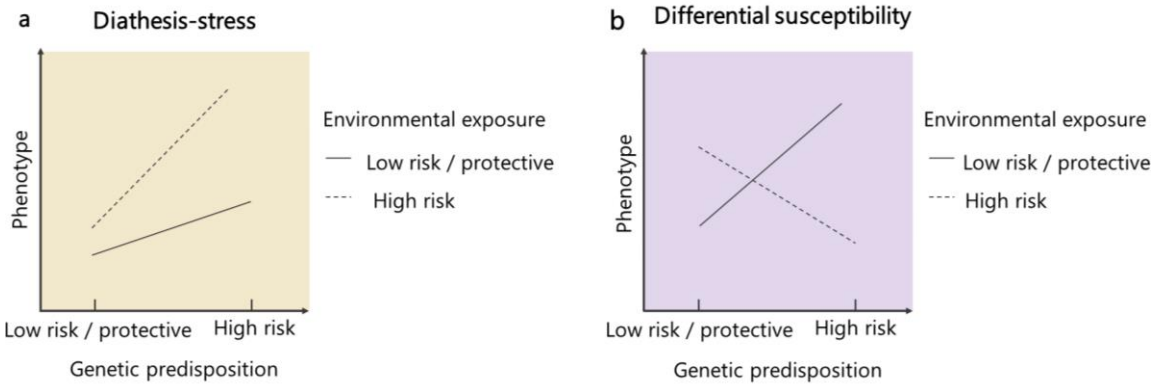


Figure 15. Two models of Gene-Environment interactions. **a.** Under the diathesis-stress framework, genetic factors and environmental factors reinforce each other. **b.** Under the differential susceptibility framework, the effect of a genetic factors is reversed as a function of an environmental factor (or, vice versa). *Adapted from Pasman et al. (2019).*

In the present thesis, we focused on GxE. Generally, the interpretation of GxE can be made according to two models. The first model can be explained in the diathesis-stress framework, where adverse environmental circumstances enhance the chance that genetic vulnerability comes to expression (example in **Figure 15a**). In this model, the greater the genetic vulnerability, the less environmental exposure is required to trigger the disorder. The second model reflects the differential susceptibility framework, posing that genetic predisposition might enhance the effect of adverse, but also positive, environmental factors (example in **Figure 15b**). In other words, that individuals may vary in their sensitivity or susceptibility to environmental influences depending on their genetic predisposition.

3.2.1. Gene-Environment Interaction Studies of SUDs

Most current studies assessing GxE in the context of SUDs have mainly focused on tobacco, alcohol and cannabis, as these substances are the most commonly used, and more data on environmental exposures are available (Pasman et al., 2019). One approach to study GxE involves investigating specific candidate genes. These kind of studies have mainly focused on interactions between genes, such as the Serotonin Transporter gene (*SLC6A4*), dopaminergic genes and alcohol-metabolizing genes with different stages of substance use, moderated by factors such as stressful life events, childhood adversity, and educational attainment (Milaniak et al., 2015). For example, one study found that an interaction between history of childhood maltreatment and the 5-HTTLPR polymorphism in the *SLC6A4* gene, which is associated with reduced transcription and functional capacity of the serotonin transporter, increased risk for early alcohol use (Kaufman et al., 2007). Similarly, the 5-HTTLPR polymorphism showed a significant interaction with stressful life events, impacting on increased frequency and heavy drinking, as well as increased substance use in college students (Covault et al., 2007). Furthermore, a study examining nicotine dependence in adults reported a significant interaction between peer smoking (which represents the socially reinforcing context of friends' encouragement and approval that contributes to smoking behavior) and a variant of the *CHRNA5* gene (rs16969968), so that peer smoking had a lower effect on the risk for nicotine dependence among those with the high risk genotype (E. O. Johnson et al., 2010).

Moreover, GxE studies can utilize polygenic measures such as PGSs. One study found that the impact of the alcohol problems PGS on alcohol problems in an independent sample was greater under conditions of low parental monitoring or high peer substance use, compared to situations of high parental monitoring or low peer substance use (Salvatore et al., 2014). Similarly, another study found that lower levels of family support interacted with the alcohol consumption PGS, leading to an increased risk of high alcohol use, while the interaction of high friend support with the PGS attenuated the association with alcohol use (Su et al., 2021). A recent study showed that high socio-economic status interacts with the alcohol use PGS by increasing the risk for higher alcohol use. However, this interaction was not significant for the PGSs for smoking heaviness, smoking initiation or cannabis initiation (Pasman et al., 2020).

A study in African-Americans showed that the association between PGS and smoking was stronger among individuals who experienced an increased number of traumatic events in their lifetimes, and the association was diminished among individuals who lived in neighborhoods with greater social cohesion (Meyers et al., 2013). In addition, in a mixed-ancestry population, trauma exposure moderated the effects of the cannabis use PGS on lifetime cannabis use, so that the PGS only influenced cannabis use among those exposed to trauma (Meyers et al., 2019). High parental monitoring and low substance use among friends, combined with the PGSs for smoking and cannabis use were able to predicted lower smoking and cannabis use, respectively (Musci et al., 2015).

Furthermore, environmental factors can interact with the PGSs for psychiatric diseases, moderating SUD outcomes. For instance, a study conducted in US Army soldiers revealed that exposure to trauma interacted with the bipolar disorder PGS, increasing the risk of alcohol misuse. Specifically, the PGS was positively associated with alcohol misuse in soldiers exposed to trauma but negatively associated in trauma-unexposed soldiers (Polimanti et al., 2018).

KEY POINTS SECTION 3

- Family, adoption, and twin studies provide valuable insights into the genetic and environmental influences on SUDs, showing that substance use initiation is influenced by a combination of genetic, as well as shared and unique environmental factors, while progression to a SUD is predominantly driven by genetic factors, particularly in adulthood.
- The heritability of SUDs is estimated to be range from 40-60%, although there is variation across substances, which can be attributed to both substance-specific genetic factors and a shared genetic liability for addiction.
- GWASs aim to identify genetic risk factors for complex disorders like SUDs by testing for differences in the allele frequency of genome-wide common genetic variants, mainly SNPs, between individuals who differ phenotypically.
- GWASs have identified various genetic risk loci associated with substance-specific SUDs including, the alcohol- metabolism genes *ADH1B* and *ALDH2* for alcohol use disorder, the *CHRNA5–CHRNA3–CHRNB4* nicotinic receptor gene cluster for tobacco use disorder and the *OPRM1* gene for opioid use disorder. In addition, the addiction-rf combining all SUDs subtypes was associated with 17 GWS risk loci.
- Post-GWAS methodologies, including PGS, genetic correlation and MR analyses, are key to gain insights into the genetic architecture and interplay between correlated traits.
- SUDs are generally correlated with each other and show positive genetic correlations with psychopathology, while displaying negative correlations with cognitive traits.
- PGS analyses have provided valuable insights into the shared genetic background between SUDs and related traits, including ADHD, but they have limited predictive power on an individual basis.
- MR analyses have been used to understand the causal relationship between SUDs and psychiatric disorders, shedding light on bidirectional causal effects in some cases.
- GxE studies on SUDs have focused mainly on tobacco, alcohol, and cannabis, reveling that environmental factors, like parental monitoring, peer substance use and trauma exposure, can modify the impact of genetic risk on SUDs outcomes.

2

HYPOTHESIS AND OBJECTIVES

1. Study 1: Genetic Overlap and Causality between Substance Use Disorder and Attention-Deficit and Hyperactivity Disorder

Given the following evidence:

- Individuals diagnosed with ADHD have a higher likelihood of heavy substance use compared to those without ADHD, predominantly cigarette smoking, cannabis use and alcohol use.
- Co-occurring ADHD and SUDs leads to more severe SUDs symptoms more frequent polysubstance dependence, greater difficulty remaining abstinent and increased risk for other mental health problems.
- One of the core risk pathways that converge in both ADHD and SUDs involves increased undercontrol/disinhibition behaviors, described by indicators of impaired executive cognitive function, emotional regulation and behavioral control.
- The use of drugs of abuse can also increase ADHD symptoms by inducing changes in brain areas involved in attention and impulse control, such as the prefrontal cortex.
- SUDs and ADHD are complex and multifactorial, with both genetic and environmental influences and heritability ranging from 30% to 70% for individual substances and around 76% for ADHD.
- ADHD and SUDs share a genetic background, supported by positive genetic correlations and associations between ADHD PGS and smoking, alcohol, or cannabis dependence.
- Evidence indicates that the genetic liability to ADHD has a causal role on an increased risk to smoking, cannabis use, and possibly alcohol dependence, but more research is needed.

1.1. Hypotheses

We propose the following hypotheses:

1. There is a high positive genetic correlation among substance-specific SUDs phenotypes and between ADHD and SUDs.

2. The genetic liability to SUDs shares a common background in the general population and individuals diagnosed with ADHD.
3. There is a causal effect of the genetic liability to ADHD on a higher risk of smoking-related phenotypes, cannabis use, alcohol dependence, cocaine dependence and ever addicted to illicit drugs.
4. For some specific substances, there might be a bidirectional causal effect, meaning that there is also a casual effect of the genetic liability to SUDs on an increased risk of ADHD.

1.2. Objectives

To explore these hypotheses, the following objectives were established:

1. To explore whether the genetic liability to substance-specific SUDs, in the form of PGSs, is associated with its respective phenotype in a clinical ADHD cohort.

    1.1. To construct PGSs for smoking initiation, alcohol dependence, lifetime cannabis use, cocaine dependence and ever addicted to illicit drugs in a clinical sample of 989 individuals with ADHD, using pre-existing GWAS summary statistics.

    1.2. To test whether the genetic background for these SUD phenotypes is shared between the general population and individuals with ADHD.

2. To explore pair-wise genetic correlations between four smoking-related phenotypes, alcohol dependence, lifetime cannabis use cocaine dependence and ADHD using pre-existing GWAS summary statistics.

3. To explore the bidirectional causal relationship between ADHD and substance-specific SUD phenotypes using genetic variants identified in GWAS as instrumental variables.

    3.1. To infer the average causal effect of the exposure on the outcome across genetic variants associated with the exposure.

3.2. To explore whether the genetic variants used to assess causality have horizontal pleiotropic effects on both the exposure and the outcome traits.

2. Study 2: Disentangling Heterogeneity in Substance Use Disorders: Insights from Genome-Wide Polygenic Scores

Given the following evidence:

- SUDs are highly heterogeneous across a wide range of phenotypic outcomes, such as type of substance(s), age at onset of SUDs, individual personality profiles, presence of comorbid conditions, and disease trajectory.
- Most available inpatient and outpatient treatments for SUDs are not well suited to accommodate the observed clinical heterogeneity, resulting in high rates of early treatment termination and relapse.
- Approximately 53% of individuals with a primary diagnosis of a SUD have at least another co-occurring diagnosis for a mental, behavioral, or emotional disorder.
- The presence of psychiatric comorbidity in SUDs has been associated with adverse disease trajectory, increased rates of suicide and worse physical and mental health.
- SUDs present substantial genetic overlap with psychiatric disorders and behavioral traits, with the strongest genetic correlations observed for ADHD, PTSD, anxiety, schizophrenia, depression, bipolar disorder and risk-taking behaviors.
- PheWAS analyses using SUDs PGSs have identified associations across major psychiatric disorders in deeply phenotyped samples.
- The etiology of SUDs is highly influenced by environmental factors, and GxE contribute to the individual differences in disease trajectory of SUDs.

2.1. Hypothesis

We propose the following hypotheses:

1. There is a shared genetic liability among different psychiatric disorders and behavioral traits and SUD-related phenotypes, including sociodemographic and health outcomes, comorbidity and personality traits and SUD variables.
2. The genetic liability for a wide range of mental health-related traits influences the heterogeneity observed in individuals with SUDs.
3. The exposure to adverse environmental factors, particularly emotional, physical, and/or sexual abuse, has an impact on the heterogeneity among individuals with SUDs, which is, in part, mediated through its interaction with genetic factors.

2.2. Objectives

To explore these hypotheses, the following objectives were established:

1. To understand the heterogeneity of SUDs by analyzing how the genetic liability for different mental health-related traits is associated with various SUD-related phenotypes, including sociodemographic and health outcomes, comorbidity and personality traits and SUD variables.

    1.1. To systematically investigate how PGSs for specific mental health-related traits are associated with different SUD-related phenotypes in a clinical sample of 1,427 individuals with SUDs.

2. To explore the impact of GxE between the genetic liability to mental health-related traits and lifetime emotional, physical, and/or sexual abuse on the SUD-related phenotypes under study.

3 RESULTS

STUDY 1

Genetic Overlap and Causality between Substance Use Disorder and Attention-Deficit and Hyperactivity Disorder

Laura Vilar-Ribó[1,2], Cristina Sánchez-Mora[1,2,3,4], Paula Rovira[1,2], Vanesa Richarte[2,3,5], Montserrat Corrales[2,3,5], Christian Fadeuilhe[2,3,5], Lorena Arribas[1,2], Miquel Casas[2,3,5], Josep Antoni Ramos-Quiroga[1,2,3,5], Marta Ribasés[1,2,3,4 *], María Soler Artigas[1,2,3,4 *]

[1] Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.
[2] Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain.
[3] Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain.
[4] Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain.
[5] Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain.

*Corresponding authors

Published in: Am J Med Genet B Neuropsychiatr Genet. 2021 Apr;186(3):140-150.
doi:10.1002/ajmg.b.32827.

Received: 3 August 2020 | Revised: 13 October 2020 | Accepted: 8 November 2020

DOI: 10.1002/ajmg.b.32827

ORIGINAL ARTICLE

American Journal of Medical Genetics Part B Neuropsychiatric Genetics | WILEY

Genetic overlap and causality between substance use disorder and attention-deficit and hyperactivity disorder

Laura Vilar-Ribó[1,2] | Cristina Sánchez-Mora[1,2,3,4] | Paula Rovira[1,2] | Vanesa Richarte[2,3,5] | Montserrat Corrales[2,3,5] | Christian Fadeuilhe[2,3,5] | Lorena Arribas[1,2] | Miquel Casas[2,3,5] | Josep Antoni Ramos-Quiroga[1,2,3,5] | Marta Ribasés[1,2,3,4] | María Soler Artigas[1,2,3,4]

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain

[2]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

[3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

[4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain

[5]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

Correspondence
Marta Ribasés and María Soler Artigas, Psychiatric Genetics Unit, Vall d'Hebron Research Institute, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain.
Email: marta.ribases@vhir.org (M. R.); maria.soler@vhir.org (M. S. A.)

Funding information
Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Grant/Award Number: 2017SGR1461; European Union H2020 Programme, Grant/Award Numbers: 667302, 728018; Fundació la Marató de TV3, Grant/Award Number: 092330/31; Instituto de Salud Carlos III, Grant/Award Numbers: P19/01224, PI15/01789, PI16/01505, PI17/00289, PI18/01788, PI19/00721; Pla estratègic de recerca i innovació en salut (PERIS), Grant/Award Number: SLT006/17/287

Abstract

Substance use disorder (SUD) often co-occur at high prevalence with other psychiatric conditions. Among them, attention-deficit and hyperactivity disorder (ADHD) is present in almost one out of every four subjects with SUD and is associated with higher severity, more frequent polysubstance dependence and increased risk for other mental health problems in SUD patients. Despite studies suggesting a genetic basis in the co-occurrence of these two conditions, the genetic factors involved in the joint development of both disorders and the mechanisms mediating these causal relationships are still unknown. In this study, we tested whether the genetic liability to five SUD-related phenotypes share a common background in the general population and clinically diagnosed ADHD individuals from an in-house sample of 989 subjects and further explored the genetic overlap and the causal relationship between ADHD and SUD using pre-existing GWAS datasets. Our results confirm a common genetic background between ADHD and SUD and support the current literature on the causal effect of the liability to ADHD on the risk for SUD. We added novel findings on the effect of the liability of lifetime cannabis use on ADHD and found evidence of shared genetic background underlying SUD in general population and in ADHD, at least for lifetime cannabis use, alcohol dependence and smoking initiation. These findings are in agreement with the high comorbidity observed between ADHD and SUD and highlight the need to control for substance use in ADHD and to screen for ADHD comorbidity in all SUD patients to provide optimal clinical interventions.

KEYWORDS
attention-deficit and hyperactivity disorder, substance use disorder, polygenic risk score, Mendelian randomization, genetic correlation

1 | INTRODUCCTION

Substance use disorder (SUD) is a psychiatric condition characterized by a hazardous use of a legal or illegal drug or medication and an inability to reduce the frequency of consumption (American Psychiatric Association, 2013). According to the World Drug Report 2019, about 270 million people (or about 5.5% of global population aged 15–64) had used psychoactive drugs during 2019 and about 35 million people are estimated to be affected by drug use disorders (United Nations Office on Drugs and Crime [UNODC], 2019). Abuse

of tobacco, alcohol, and illicit drugs have a huge impact on crime, work productivity and the healthcare system (National Institute on Drug Abuse [NIDA], 2020a). In addition, global deaths directly caused by the use of drugs increase every year and alcohol is the leading risk factor for premature mortality and disability among those aged 15 to 49 years old (UNODC, 2019).

SUD often co-occur at high prevalence with other psychiatric conditions and mental disorders, which reduces life expectancy and increases disease burden and societal impact. Among them, attention-deficit and hyperactivity disorder (ADHD) is present in almost one out of every four patients with SUD and is associated with an increased risk of substance use, abuse and dependence (van de Glind et al., 2014; van Emmerik-van Oortmerssen et al., 2012). Children diagnosed with ADHD are at higher risk of developing SUD and nicotine dependence in adolescence (Charach et al., 2011; Ilbegi et al., 2018) and a four year follow up study found that adolescents with a childhood ADHD diagnosis were 1.8 and 8.6 times more likely to develop psychoactive SUD and nicotine dependence, respectively, compared to healthy controls (Groenman et al., 2013). In addition, ADHD comorbidity is associated with higher severity, more frequent polysubstance dependence and increased risk for other mental health problems in SUD patients (Icick et al., 2020).

SUD and ADHD are complex and multifactorial, with both genetic and environmental influences and heritability ranging from 50% to 70% for individual substances (Wetherill et al., 2019; Wong & Schumann, 2008) and around 76% for ADHD (Wendt et al., 2020). Several studies also support a shared genetic background between them, reporting evidence of positive genetic correlations between ADHD and different SUD (Pasman et al., 2018; Soler Artigas et al., 2019; Walters et al., 2018), and associations between ADHD Polygenic Risk Score (PRS) and SUD (Goldman, 2017; Wimberley et al., 2020), smoking (Du Rietz et al., 2018) and alcohol or cannabis dependence (Du Rietz et al., 2018; Wimberley et al., 2020). Inconsistent results, however, were observed, with some studies reporting a lack of association between ADHD-PRS and smoking, cannabis dependence or SUD (Carey et al., 2016; Gurriarán et al., 2019; Rabinowitz et al., 2018).

Additionally, there is evidence that liability to ADHD is on the causal pathway to smoking-related phenotypes and SUD (Elkins et al., 2018; Fluharty et al., 2018). In this context, Soler Artigas et al found a causal effect of liability to ADHD on cannabis use (Soler Artigas et al., 2019). Likewise, Treur et al reported that liability to ADHD also increases risk to smoking, cannabis use and, tentatively, alcohol dependence, and that liability to smoking initiation also increases ADHD risk (Treur et al., 2019).

Despite the high incidence of SUD in subjects with ADHD, the associated social and clinical difficulties, and previous studies suggesting a genetic basis in the co-occurrence of these two conditions, the causal relationship that may exist between them is still unclear. The aim of the present study was to use an in-house sample of 989 subjects with ADHD and pre-existing GWAS datasets on ADHD and SUD to (i) test whether the genetic liability to SUD shares a common background in the general population and clinically diagnosed ADHD individuals through PRS analyses and (ii) further explore the genetic overlap and the causal relationship between ADHD and SUD.

2 | MATERIALS AND METHODS

2.1 | Samples

2.1.1 | In-house ADHD cohort

The in-house cohort consisted on 989 subjects with ADHD. The clinical assessment of ADHD was conducted by a psychiatrist based on the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II) and the Conner's Adult ADHD Diagnosis Interview for DSM-IV (CAADID Parts I and II). Information about lifetime substance use was collected by a psychiatrist with the Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I) and additional information on smoking was available by an open interview from the majority of participants. All individuals were unrelated, of European descendent and had been recruited at the Program of Adult ADHD of Vall d'Hebron University Hospital of Barcelona (Spain). The study was approved by the Clinical Research Ethics Committee (CREC) of the Hospital Universitari Vall d'Hebron, all methods were performed in accordance to the relevant guidelines and regulations and written informed consent was obtained from all subjects before inclusion into the study.

Genomic DNA was isolated from peripheral blood leukocytes by the salting-out procedure. Subjects were genotyped in four different waves using the HumanOmni1-Quad (N = 558), HumanOmni 2.5 (N = 218), the PsychChip (N = 60) and the GSA (N = 153) Illumina arrays. Pre-imputation quality control and principal components analysis were implemented with the Ricopili pipeline (https://sites.google.com/a/broadinstitute.org/ricopili/), and ancestry outliers or relatives were excluded from the analyses. Genotype imputation was performed using the European population haplotypes of the 1000 Genomes Project Phase I as the reference panel for waves 1 and 2 and the 1000 Genomes Project Phase III for waves 3 and 4 (The 1000 Genomes Project Consortium, 2015). Individuals with > 2% best-guess genotype missingness were removed, as well as SNPs with low call rate (< 0.95), with minor allele frequency (MAF) < 0.01, INFO score below 0.8 or failing the Hardy–Weinberg equilibrium test ($P < 1e-06$). Post-imputation best-guess genotype data from a total of 4,105,370 markers were available in all four datasets.

2.2 | Pre-existing ADHD and substance use GWAS-MA datasets

2.2.1 | ADHD

Psychiatric Genomics Consortium (PGC) and iPSYCH data from the largest GWAS-MA on ADHD performed to date (N = 55,374) was

considered (Demontis et al., 2019). Because of sample overlap between the ADHD and lifetime cannabis use datasets, we removed the effect of two studies (Barcelona (572 cases and 425 controls) and Yale-Penn (182 cases and 1315 controls)) from the ADHD GWAS-MA using an inverse variance weighted difference for the beta and standard error estimates. This provided a restricted PGC + iPSYCH ADHD sample of 18,345 cases and 32,454 controls that was used in all analyses where both traits were considered (Table 1).

2.2.2 | Substance use

Data on four smoking-related phenotypes were obtained from a meta-analysis of over 30 GWAS datasets performed in participants of European ancestry (Liu et al., 2019). Smoking-related phenotypes included smoking initiation, indicating whether an individual had ever smoked regularly (*N* = 632,802; 46% ever smokers), age of initiation of regular smoking (*N* = 262,990), cigarettes per day, as a measure of smoking heaviness (*N* = 263,954) and smoking cessation, contrasting current versus former smokers (*N* = 312,821; 39% current smokers) (Table 1).

Results from the GWAS-MA on alcohol dependence (AD) considering the subset of genetically unrelated European individuals by the PGC were used (8,485 cases and 20,272 controls). AD was defined as meeting criteria for a DSM-IV (or DSM-IIIR in one instance) diagnosis of AD (Table 1). More details are available in Walters et al (Walters et al., 2018).

GWAS-MA on lifetime cannabis use included data from the International Cannabis Consortium (ICC) and UK Biobank in a total of 162,082 European ancestry individuals (Pasman et al., 2018). Because of sample overlap with the in-house ADHD cohort and to avoid biases, we removed the effect of the Barcelona study (981 individuals) from the ICC sample using an inverse variance weighted difference for the beta and standard error estimates. This provided a sample size of 161,101 individuals that was used in all analyses where lifetime cannabis use and ADHD were considered (Table 1).

For the GWAS-MA on cocaine dependence we considered 2,085 cases that met DSM-IV criteria for cocaine dependence and 4,293 controls of European-ancestry (Table 1) (Cabana-Domínguez et al., 2019).

UK Biobank data for ever addicted to illicit drugs was available for 6,944 participants (518 cases and 6,426 controls) (Table 1). More details about UK Biobank data can be found at http://www.nealelab.is/uk-biobank/. Additional quality control filters of INFO score < 0.8 and MAF < 0.01 were undertaken in this UK Biobank dataset prior to all analyses.

The studies included in the GWAS-MA of the smoking-related phenotypes have all been imputed to Haplotype Reference Consortium, 1000 Genomes or a combination including more specific reference panels (Liu et al., 2019). GWAS-MA for alcohol dependence, lifetime cannabis use (ICC sample) and cocaine dependence used the 1000 Genomes Phase III reference panel to perform the imputation (Cabana-Domínguez et al., 2019; Pasman et al., 2018; Walters et al., 2018) and UK Biobank data for ever addicted to illicit drugs and lifetime cannabis use were imputed using the haplotype reference consortium (HRC) reference panel (http://www.nealelab.is/uk-biobank/).

2.3 | Polygenic risk score analyses

Five SUD-related phenotypes were collected in an in-house cohort of ADHD individuals where data on SUD were available: smoking initiation, alcohol dependence, lifetime cannabis use, cocaine dependence and ever addicted to illicit drugs. Controls for each cohort were individuals that did not consume that substance, but could be consumers for any of the others substances.

PRS for smoking initiation, alcohol dependence, lifetime cannabis use, cocaine dependence and ever addicted to illicit drugs were constructed using pre-existing GWAS-MA summary statistics and tested for association with the five SUD-related phenotypes in the in-house ADHD cohort using the PRSice software v.1.25. Independent variants

TABLE 1 Summary of the pre-existing GWAS-MA datasets on five different SUD-related phenotypes (ever smoking, alcohol dependence, lifetime cannabis use, cocaine dependence and ever addicted to illicit drugs) and ADHD

Phenotype	*N* cases	*N* controls	*N* total	*N* effective[a]	*N* SNPs	*N* GWS SNPs[b]	Reference
Smoking initiation	329,057	303,745	632,802	631,790	11,802,365	127	Liu et al., 2019
Age of smoking initiation	–	–	262,990	–	11,983,807	8	Liu et al., 2019
Cigarettes per day	–	–	263,954	–	12,003,614	54	Liu et al., 2019
Smoking cessation	122,000	190,821	312,821	297,680	12,197,134	14	Liu et al., 2019
Alcohol dependence	8,485	20,272	28,757	23,926	9,142,831	3	Walters et al., 2018
Lifetime cannabis use	43,380	118,702	162,082	127,079	9,076,507	8	Pasman et al., 2018
Cocaine dependence	2,085	4,293	6,378	5,614	8,591,128	0	Cabana-Domínguez et al., 2019
Ever addicted to illicit drugs	518	6,426	6,944	1,917	13,791,467	1	http://www.nealelab.is/uk-biobank/
ADHD	20,183	35,191	55,374	51,306	8,094,094	12	Demontis et al., 2019

[a]N effective sample sizes were calculated following the equation: Neff = 4/(1/Ncases+1/Ncontrols).
[b]Number of genome-wide significant SNPs.

from each discovery GWAS-MA were selected by clumping (p_1 = 1, p_2 = 1, r^2 = 0.2, kb = 250), and PRS were constructed as the sum of the risk alleles across SNPs of each individual, weighted by their discovery effect sizes and divided by the total number of alleles included. P-value thresholds from 0 to 0.5 with increments of 0.00005 were considered and logistic regression was used to test the association between each PRS and the SUD phenotype in the ADHD cohort adjusting for age, sex, genotyping wave and the first five principal components. The best-fit PRS was determined by the *p*-value threshold with the most predictive PRS on the phenotype and 10,000 permutations were computed at the *p*-value threshold that explained the most variance (best-fit) to correct for multiple testing. As well as Nagelkerke's R^2, we reported R^2 on the liability scale and adjusted for ascertainment in case-control studies (Lee's R^2) (Lee et al., 2012). Population prevalence used for each substance were the following: smoking initiation = 28.7% (World Health Organization [WHO], 2018); alcohol dependence = 5.6% (National Institute on Drug Abuse [NIDA], 2020b); lifetime cannabis use = 2.5% (World Health Organization [WHO], 2020); cocaine dependence = 0.4% (Grant et al., 2016); ever addicted to illicit drugs = 3.9% (Grant et al., 2016).

2.4 | Genetic correlation

Cross-trait LD score regression was used to estimate the genetic correlation for all possible pairs of the following traits: ADHD, smoking initiation, age of smoking initiation, cigarettes per day, smoking cessation, alcohol dependence, lifetime cannabis use and cocaine dependence using previously published GWAS-MA summary statistics (Demontis et al., 2019; Liu et al., 2019; Walters et al., 2018; Pasman et al., 2018; Cabana-Domínguez et al., 2019; http://www.nealelab.is/uk-biobank/). N effective sample sizes were calculated following the equation: Neff = 4/(1/Ncases+1/Ncontrols), and only traits with Neff > 5,000 were considered. For this analysis we used HapMap3 SNPs and LD scores computed from the 1000 Genomes Project reference panel (Bulik-Sullivan et al., 2015).

2.5 | Mendelian randomization

Causality between ADHD and SUD-related phenotypes were assessed using previously published GWAS-MA summary statistics and bidirectional two-sample Mendelian Randomization (MR) (Burgess et al., 2017). Age of smoking initiation, cigarettes per day and smoking cessation could not be used as exposures, since these phenotypes can only be assessed in smokers and results stratified by smoking initiation were not available in the outcome GWAS-MA (ADHD).

Strict clumping was undertaken in the exposure population considering r^2 = 0.05, kb = 500 and p_2 = 0.5, using Plink 1.9 software (Purcell et al., 2007). Then, we selected SNPs at a threshold of *p*-value < 5e-08 to be used as instruments, and identified the same SNPs in the outcome population. When there was a limited number of SNPs (<5), surpassing the genome-wide significant threshold of *p*-value <5e-08 in the exposure population the threshold was lowered to *p*-value <5e-06.

We used the multiplicative random effects inversed-variance weighted (IVW) as the main method to obtain the average effect across genetic variants and additional MR methods were implemented as sensitivity analyses for IVW significant findings: MR-Egger regression (Bowden et al., 2015), weighted median regression (Bowden, Davey Smith, et al., 2016a), MR-PRESSO (Verbanck et al., 2018), and Steiger filtering (Hemani et al., 2018). When I^2 < 0.6, Egger results were considered not valid due to violation of the NOME (NO Measurement Error) assumption (Bowden, Del Greco, et al., 2016b). Additionally, we ran heterogeneity tests with Cochran Q statistics and repeated analyses removing one genetic variant at a time (leave-one-out analyses) for the main analysis when we obtained significant results. All analyses were run with the "TwoSampleMR" R package and, additionally, with the "MRPRESSO" R package for the MR-PRESSO sensitivity analysis. For further information on the MR methods used, please see Supplementary methods.

3 | RESULTS

The ADHD cohort included 657 males (66.4%) and 332 females (33.6%), with a mean age of 32.96 years ± 10.7. Fifty-four percent of ADHD patients (N = 539) had used at least one of the substances considered in the study and 6.5% (N = 64) were consumers for all substances (Figure 1).

To assess whether SUD-related phenotypes in the general population and in clinically diagnosed ADHD individuals share a common genetic load, we constructed PRSs for smoking initiation, alcohol or cocaine dependence, lifetime cannabis use and ever addicted to illicit drugs in the general population using data from pre-existing GWAS datasets (Table 1) and tested their association with these SUD-related phenotypes in our in-house ADHD sample. We found evidence of association for lifetime cannabis use, alcohol dependence and smoking initiation (*p*-value= 9e-03, 5e-03 and 1.50e-03, respectively). The best-fit PRS was set at a *p*-value threshold (P_T) of 0.26 for alcohol dependence and explained 1.0% of the variance, at P_T = 0.07 for lifetime cannabis use, explaining 0.6% of the variance and at P_T = 9.50e-04 for smoking initiation and explained 1.4% of the variance in the phenotype (Figure 2, Supplementary Table 1). No significant results were found for cocaine dependence or ever addicted to illicit drugs (Supplementary Table 1, Supplementary Figure 1).

When we estimated pairwise genetic correlations (rg) for ADHD and SUD-related phenotypes we found evidence of genetic correlations between all traits (Figure 3). The strongest positive correlations were found between alcohol and cocaine dependence (rg = 0.76, se = 0.15, *p*-value = 5.62e-07) and between smoking initiation and alcohol dependence (rg = 0.69, se = 0.09, *p*-value = 3.05e-14). Pairs of traits with weaker genetic correlations were ADHD and lifetime cannabis use (rg = 0.15, se = 0.04, *p*-value = 4e-04) and alcohol dependence and lifetime cannabis use (rg = 0.16, se = 0.08, *p*-value = 0.05).

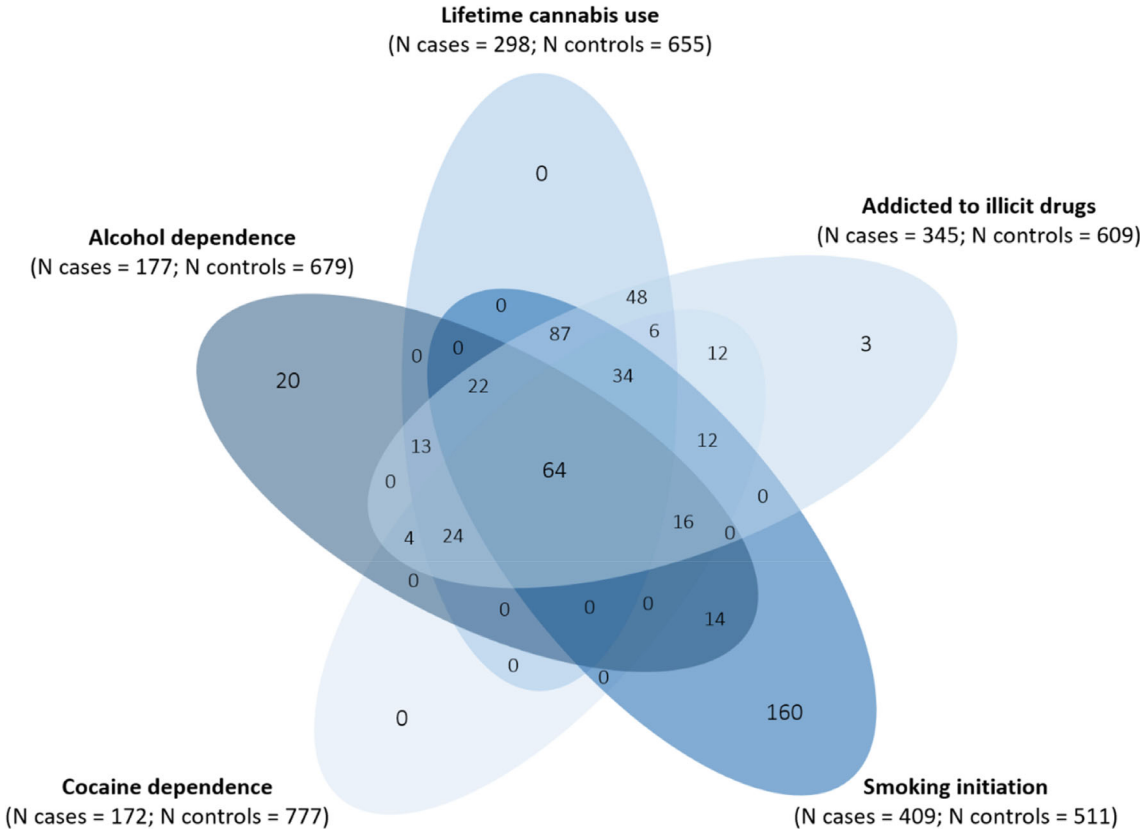


FIGURE 1 Venn diagram showing the sample overlap between the five SUD-related phenotypes from the in-house cohort of 989 subjects with ADHD: Smoking initiation, alcohol dependence, lifetime cannabis use, cocaine dependece and ever addicted to illicit drugs [Color figure can be viewed at wileyonlinelibrary.com]

Similarly, strong negative genetic correlations were observed between age of smoking initiation and alcohol dependence (rg = −0.68, se = 0.10, *p*-value = 1.52e-11) and ADHD (rg = −0.62, se = 0.04, *p*-value = 1.69e-49) (Supplementary Table 2).

Lastly, our results for the bidirectional two-sample Mendelian Randomization analyses support a causal effect of the genetic liability to ADHD on smoking initiation (IVW *p*-value = 2.24e-21 and OR = 1.20), age of smoking initiation (IVW *p*-value = 4.27e-04 and OR = 0.94) and cigarettes per day (IVW *p*-value = 8.23e-04 and OR = 1.08), but not on smoking cessation (IVW *p*-value = 0.09 and OR = 1.05) (Table 2). There was no evidence that these results were driven by a single SNP and all SNPs were valid instruments for the analysis according to Steiger filtering. However, we found evidence of heterogeneity for smoking initiation and age of smoking initiation (Cochran Q Statistics *p*-value = 0.02 and 4.18e-04, respectively) as well as evidence of horizontal pleiotropy for age of smoking initiation (MR-PRESSO Global Test *p*-value = 1e-03). MR-PRESSO outlier-corrected results for age of smoking initiation remained significant and reported a similar effect size (*p*-value = 5.47e-05 and OR = 0.93) (Table 2 and Supplementary Table 3).

Significant results were found when smoking initiation was considered as exposure and ADHD as outcome (IVW *p*-value = 1.36e-35 and OR = 2.58), but evidence of heterogeneity and horizontal pleiotropy were detected in the sensitivity analyses (Cochran Q Statistics *p*-value = 3.21e-08 and MR-PRESSO Global Test *p*-value = 1e-04). MR-PRESSO outlier-corrected results remained significant, but the effect size was reduced (*p*-value = 3.66e-17 and OR = 2.46). There was no evidence that these results were driven by a single SNP and with all SNPs being valid instruments for the analysis according to the Steiger filtering.

A causal effect was also found for the liability to ADHD on lifetime cannabis use (IVW *p*-value = 2.01e-03 and OR = 1.15) and vice-versa (IVW *p*-value = 8.00e-04 and OR = 1.46), with no evidence of horizontal pleiotropy or heterogeneity in any direction. The leave-one-out analysis gave no evidence that these findings were driven by any single variant and the Steiger filtering showed that all variants were more predictive of the exposures than of the outcomes (Table 2 and Supplementary Table 3).

Finally, no causal effects of the liability to ADHD on alcohol dependence, cocaine dependence or ever addicted to illicit drugs and vice-versa were found (Table 2).

4 | DISCUSSION

For the first time, we provide evidence of a common genetic background between lifetime cannabis use, alcohol dependence and smoking initiation in the general population and in subjects with ADHD and support previous findings of a causal relationship for the liability to

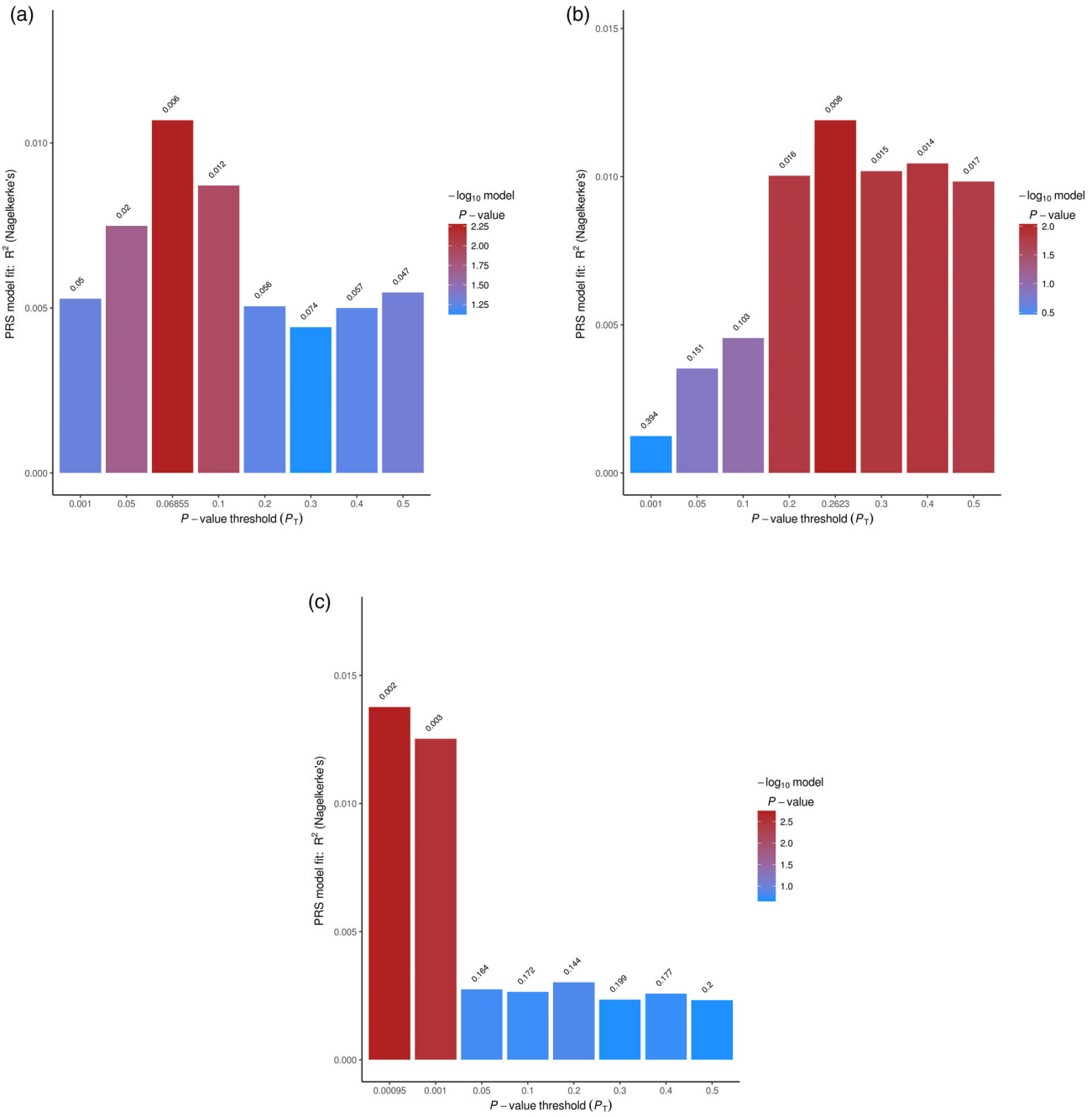


FIGURE 2 Bar plots showing results from the PRS analysis based on (a) lifetime cannabis use, (b) alcohol dependence and (c) smoking initiation at broad *p*-value thresholds (p_T = 0.001, p_T = 0.05, p_T = 0.1, p_T = 0.2, p_T = 0.3, p_T = 0.4, p_T = 0.5) and at the best-fit PRS [Color figure can be viewed at wileyonlinelibrary.com]

ADHD on the risk for lifetime cannabis use, smoking initiation, age of smoking initiation and cigarettes per day. These results are in agreement with epidemiological evidence showing increased risk and higher severity of substance use, abuse and dependence in ADHD subjects (Groenman et al., 2013; Groenman et al., 2017; van de Glind et al., 2014; van Emmerik-van Oortmerssen et al., 2012; Wilens & Morrison, 2012) and support previous reports showing that the liability to ADHD is on the causal pathway to smoking and SUD (Artigas Soler et al., 2019; Treur et al., 2019; Jang et al., 2020).

In addition to the finding supporting that ADHD risk impacts on lifetime cannabis use, we found evidence of reverse causation, with a causal effect of genetic liability to lifetime cannabis use on ADHD. Although these results do not follow the appropriate temporal sequence by which the exposure (lifetime cannabis use) precedes the outcome (ADHD), they may be due to shared risk factors or alternatively, reflect the relationship between parental exposure to cannabis and psychiatric outcomes in offspring described in both human and animal models (El Marroun et al., 2019; Langley et al., 2012).

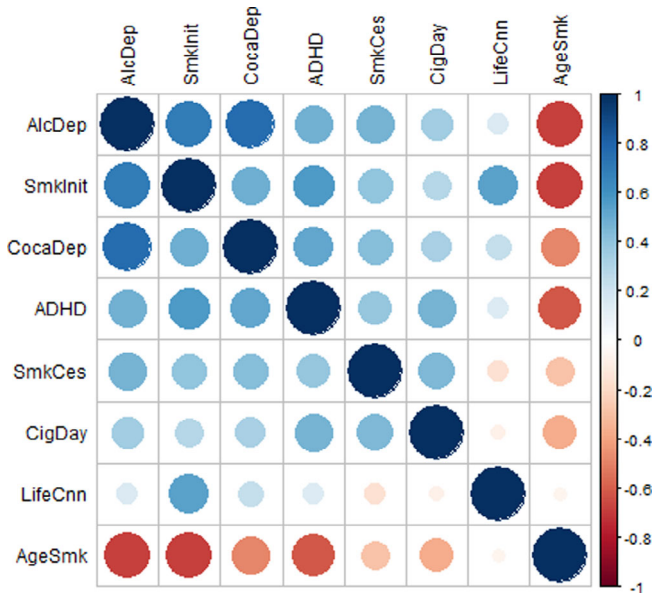


FIGURE 3 Genetic correlation resultsfor all possible pairs of traits. The size and the colour of the circles correspond with the rg value. AlcDep, alcohol dependence; SmkInit, smoking initiation; CocaDep, cocaine dependence; ADHD, attention-deficit and hyperactivity disorder; SmkCes, smoking cessation; CigDay, cigarettes per day, LifeCnn, lifetime cannabis use, AgeSmk, age of smoking initiation [Color figure can be viewed at wileyonlinelibrary.com]

Pre-gestational exposure to cannabis or maternal cannabis use during pregnancy are associated with long-term adverse neurocognitive and behavioural outcomes in the offspring (Levin et al., 2019). Thus, cannabis exposure, particularly during sensitive and critical windows of development, may impact on ADHD through epigenetic modifications and subsequent gene expression alterations (Murphy et al., 2018; Smith et al., 2020). Additionally, male paternal cannabis use or smoking were also associated with offspring behavioural problems, which suggests that, in addition to causal intrauterine effects, these associations may be influenced by a shared genetic background between parental substance use and ADHD symptoms in the offspring (El Marroun et al., 2019; Langley et al., 2012). Our results may therefore reflect dynastic effects, where the exposure trait in a previous generation influences the outcome trait of the current generation (Hartwig et al., 2018), hypothesis that needs to be confirmed by alternative study designs such as sibling pairs.

Additionally, our results may also reflect an effect of lifetime cannabis use on the risk of the persistent form of ADHD across the lifespan. This is supported by the association between hazardous use of cannabis and ADHD symptoms in adulthood (Fergusson & Boden, 2008; Kolla et al., 2016) and by the effect of cannabis use on the impairment of planning, inhibition or decision-making, also associated with ADHD (Crean et al., 2011). Given that there is strong evidence for a common genetic background underlying ADHD in children and adults (Rovira et al., 2020), environmental risk factors, such as cannabis use, may play a role in the different ADHD trajectories across the lifespan. Since we considered ADHD in both childhood and adulthood as outcome, further studies on ADHD in children and adults separately as well as additional research on the effect of cannabis use on late-onset ADHD, where ADHD symptoms arise in the late adolescence or adulthood, (Cooper et al., 2018) are required to clarify the temporal relationship between cannabis use and ADHD.

MR results should be interpreted taking into consideration the evidence of horizontal pleiotropy for some smoking-related phenotypes. These findings, however, add additional evidence to the growing literature showing that individuals with ADHD are at higher risk of initiating smoking at younger ages and of being heavier smokers, as well as the existence of a causal effect of smoking initiation on ADHD (Jang et al., 2020; Treur et al., 2019).

We also found evidence of genetic correlations between all traits considered in the study, namely ADHD, smoking initiation, age of smoking initiation, cigarettes per day, smoking cessation, alcohol or cocaine dependence and lifetime cannabis use. These results are in agreement with extensive literature showing poorer smoking outcomes and higher rates of polysubstance use and comorbidity between ADHD and SUD (Capusan et al., 2019; Hayley et al., 2017; Martínez-Luna et al., 2019; van Emmerik-van Oortmerssen et al., 2014) and add additional evidence for a common genetic background among these phenotypes (Du Rietz et al., 2018; Wimberley et al., 2019, Wilens, 2007; Vink et al., 2014; Chang et al., 2019).

The present study, however, should be considered in the context of some limitations:

First, the limited sample size of the in-house ADHD cohort and the fact that most individuals were polysubstance users may have mitigated specific results in the PRS analyses for individual substances. We did not control our analysis for comorbid conditions, either, meaning that the presence of other psychiatric disorders sharing genetic background with SUD could also bias our results. For instance, externalizing disorders that have been associated with smoking, alcohol and cannabis dependence (Grant et al., 2015; Rabinowitz et al., 2018), may have impacted on the observed associations. Despite that, recent findings show that ADHD-PRS was also associated with SUD after controlling for conduct disorder (Wimberley et al., 2019).

Second, even though we conducted a different set of analyses for each SUD-related phenotype, no multiple testing correction was applied, with the exception of a permutation test performed to correct for the different thresholds in individual PRS analyses. At risk of presenting false positive results, applying more conservative multiple testing corrections may be too strict given the high genetic correlation between the different SUD-related phenotypes considered.

Lastly, we based our analyses on publicly available GWAS-MA on ADHD and SUD-related phenotypes and, although we selected larger and recently published data, the sample sizes differed between phenotypes. This may have reduced power to identify PRS-associations or causal effects through the Mendelian randomization analyses for phenotypes with smaller sample sizes, such as alcohol dependence, cocaine dependence and ever addicted to illicit drugs, and may have impacted on results for directionality between traits, which should be interpreted with caution given the different statistical power for each direction. Also a restrictive threshold was used to select genetic

TABLE 2 Results of the Mendelian Randomization analyses when ADHD was set as the exposure and SUD-related phenotypes as outcomes, and vice-versa

		ADHD → SUD-related phenotype				SUD-related phenotype → ADHD			
Phenotype	**Method**	***N* SNPs**	**OR**	**95% CI**	***p*-value**	***N* SNPs**	**OR**	**95% CI**	***p*-value**
Smoking initiation	Inverse-variance weighted	12	1.20	1.15, 1.234	2.24e-21	118	2.58	2.22, 3.00	1.36e-35
	Weighted median		1.18	1.13, 1.24	2.36e-12		2.56	2.13, 3.07	7.28e-24
	MR-PRESSO outlier-corrected		–	–	–	116	2.46	2.14, 2.38	3.66e-17
Age Smoking initiation	Inverse-variance weighted	12	0.94	0.91, 0.97	4.72e-04	-	-	-	-
	Weighted median		0.93	0.90, 0.96	1.32e-06		-	-	-
	MR-PRESSO outlier-corrected	11	0.93	0.90, 0.95	5.47e-05		-	-	-
Cigarettes per day	Inverse-variance weighted	12	1.08	1.03, 1.13	8.23e-04	-	-	-	-
	Weighted median		1.08	1.02, 1.14	0.01		-	-	-
Smoking cessation	Inverse-variance weighted	12	1.05	0.99, 1.08	0.09	-	-	-	-
Lifetime Cannabis use	Inverse-variance weighted	12	1.15	1.05, 1.25	2.01e-03	8	1.46	1.17, 1.83	8.00e-04
	Weighted median		1.18	1.06, 1.30	1.41e-03		1.46	1.12, 1.90	5.73e-03
Alcohol dependence	Inverse-variance weighted	12	1.15	0.97, 1.35	0.10	21†	0.98	0.93, 1.03	0.28
Cocaine dependence	Inverse-variance weighted	12	1.27	0.91, 1.77	0.15	10†	1.02	0.98, 1.07	0.41
Ever addicted to illicit drugs	Inverse-variance weighted	12	1.00	0.97, 1.03	0.86	7†	1.51	0.83, 2.76	0.17

Note: Inverse variance weighted random effects (IVW) = main MR method. Sensitivity analyses were only conducted for significant IVW results. When I^2 < 0.6, MR-Egger results were considered not valid and are not shown. MR-PRESSO outlier-corrected results only presented for those analysis with *p*-value and global test *p*-value < 0.05. *p*-value threshold of 5e-08 for all analyses, except for alcohol and cocaine dependence and ever addicted to illicit drugs as exposures (N SNPs marked with a †), when *p*-value threshold = 5e-06.

variants for the majority of the Mendelian randomization analyses (*p*-value < 5e-08). Although relaxing the threshold may include a larger the number of variants and, thus, increase power, these variants showing weaker associations could be invalid instruments.

In conclusion, our results confirm a common genetic background between ADHD and SUD and support the current literature on the causal effect of the liability to ADHD on the risk for SUD. For the first time, we add novel findings on the effect of lifetime cannabis use on ADHD and found evidence of shared genetic background underlying SUD in general population and in ADHD, at least for lifetime cannabis use, alcohol dependence and smoking initiation. Although larger studies will be needed to provide more conclusive results, these findings are in agreement with the high comorbidity observed between ADHD and SUD and highlight the need to control for substance use in ADHD and to carefully screen for ADHD in patients seeking treatment for SUD to provide optimal clinical interventions.

ACKNOWLEDGMENTS

Over the course of this investigation, L.V.R. was a recipient of a predoctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285), C.S.M. was a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199), P.R. was a recipient of a predoctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI_B 00899), M.R. was a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023), M.S.A. was a recipient of a contract from the Biomedical Network Research Center on Mental Health (CIBERSAM), Madrid, Spain and a recipient of a Juan de la Cierva Incorporación contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I).

The research leading to these results has received funding from the European Union H2020 Programme (H2020/2014-2020) under grant agreements no. 667302 (CoCA) and no. 728018 (Eat2beNICE), from the Instituto de Salud Carlos III (PI15/01789, PI16/01505, PI17/00289, PI18/01788, PI19/00721, and P19/01224), from the Pla estratègic de recerca i innovació en salut (PERIS); Generalitat de Catalunya (METAL-Cat; SLT006/17/287) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461) and cofinanced by the European Regional Development Fund (ERDF) and by "la Marató de TV3" (092330/31). The work was also supported by the ECNP Network 'ADHD across the Lifespan' (https://www.ecnp.eu/researchinnovation/ECNP-networks/List-ECNP-Networks/). This paper reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained therein.

CONFLICT OF INTEREST

Laura Vilar-Ribó, Dr. Sánchez-Mora, Paula Rovira, Lorena Arribas, Dr. Soler Artigas and Dr. Ribasés reported no biomedical financial interests or potential conflicts of interest. Vanesa Richarte has served on the speakers for Eli Lilly, Rubio and Shire in the last 5 years. She has received travel awards from Eli Lilly and Co. and Shire for participating in psychiatric meetings. The ADHD Program has received unrestricted educational and research support from Eli Lilly and Co., Janssen-Cilag, Shire, Rovi, Psious and Laboratorios Rubió in the past

two years. Montserrat Corrales received travel awards for taking part in psychiatric meetings from Shire. Christian Fadeuilhe received travel awards for taking part in psychiatric meetings from Shire and Lundbeck. Prof. Casas has received travel grants and research support from Eli Lilly and Co., Janssen-Cilag, Shire and Lundbeck and served as consultant for Eli Lilly and Co., Janssen-Cilag, Shire and Lundbeck. Prof. Ramos-Quiroga was on the speakers' bureau and/or acted as consultant for Eli-Lilly, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogui, Lundbeck, Almirall, Braingaze, Sincrolab, Medice and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogui, Bial, Medice and Eli- Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli-Lilly, Lundbeck, Janssen- Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió.

AUTHOR CONTRIBUTIONS

L.V.R., C.S.M., P.R. and L.A. participated in the DNA isolation and preparation of samples. L.V.R., P.R., C.S.M., M.S.A. and M.R. undertook the statistical analyses. V.R., M.C. and C.F. contributed to the clinical assessment and recruitment of patients. M.C. and J.A.R.Q. participated in the study design, clinical assessment and coordination of the clinical research. M.R. and M.S.A. conceived the project, wrote the protocol and coordinated the study design and the statistical analyses. M.S.A. and M.R. supervised the project and the manuscript preparation. All authors contributed to and have approved the final version.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at: https://doi.org/10.1038/s41588-018-0307-5; https://doi.org/10.1038/s41588-018-0269-7; https://doi.org/10.1038/s41593-018-0206-1; https://doi.org/10.1038/s41593-018-0275-1; https://doi.org/10.1016/j.pnpbp.2019.109667 and http://www.nealelab.is/uk-biobank/.

ORCID

Laura Vilar-Ribó https://orcid.org/0000-0001-8758-523X
Cristina Sánchez-Mora https://orcid.org/0000-0003-4211-1107
Marta Ribasés https://orcid.org/0000-0003-1039-1116
María Soler Artigas https://orcid.org/0000-0002-3213-1107

REFERENCES

National Institute on Drug Abuse (NIDA) (2020a). Alcohol Facts and Statistics. Retrieved from https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics.

World Health Organization (WHO) (2020). Alcohol, drugs and addictive behaviours unit. Retrieved from https://www.who.int/teams/mental-health-and-substance-use/alcohol-drugs-and-addictive-behaviours/drugs-(psychoactive)/cannabis.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Philadelphia, PA: American Pshychiatric Association.

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, *44*(2), 512–525. https://doi.org/10.1093/ije/dyv080

Bowden, J., Davey Smith, G., Haycock, P. C., & Burgess, S. (2016a). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology*, *40*(4), 304–314. https://doi.org/10.1002/gepi.21965

Bowden, J., Del Greco, M.,. F., Minelli, C., Davey Smith, G., Sheehan, N. A., & Thompson, J. R. (2016b). Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *International Journal of Epidemiology*, *45*(6), 1961–1974. https://doi.org/10.1093/ije/dyw220

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Burgess, S., Thompson, D. J., Rees, J., Day, F. R., Perry, J. R., & Ong, K. K. (2017). Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. *Genetics*, *207*(2), 481–487. https://doi.org/10.1534/genetics.117.300191

Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N., & Cormand, B. (2019). Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *94*, 109667. https://doi.org/10.1016/j.pnpbp.2019.109667

Capusan, A. J., Bendtsen, P., Marteinsdottir, I., & Larsson, H. (2019). Comorbidity of Adult ADHD and Its Subtypes With Substance Use Disorder in a Large Population-Based Epidemiological Study. *Journal of Attention Disorders*, *23*(12), 1416–1426. https://doi.org/10.1177/1087054715626511

Carey, C. E., Agrawal, A., Bucholz, K. K., Hartz, S. M., Lynskey, M. T., Nelson, E. C., … Bogdan, R. (2016). Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Frontiers in Genetics*, *7*, 149. https://doi.org/10.3389/fgene.2016.00149

Chang, L. H., Whitfield, J. B., Liu, M., Medland, S. E., Hickie, I. B., Martin, N. G., … GSCAN Consortium. (2019). Associations between polygenic risk for tobacco and alcohol use and liability to tobacco and alcohol use, and psychiatric disorders in an independent sample of 13,999 Australian adults. *Drug and Alcohol Dependence*, *205*, 107704. https://doi.org/10.1016/j.drugalcdep.2019.107704

Charach, A., Yeung, E., Climans, T., & Lillie, E. (2011). Childhood attention-deficit/hyperactivity disorder and future substance use disorders: comparative meta-analyses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(1), 9–21. https://doi.org/10.1016/j.jaac.2010.09.019

Cooper, M., Hammerton, G., Collishaw, S., Langley, K., Thapar, A., Dalsgaard, S., … Riglin, L. (2018). Investigating late-onset ADHD: a population cohort investigation. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(10), 1105–1113. https://doi.org/10.1111/jcpp.12911

Crean, R. D., Crane, N. A., & Mason, B. J. (2011). An evidence based review of acute and long-term effects of cannabis use on executive cognitive functions. *Journal of Addiction Medicine*, *5*(1), 1–8. https://doi.org/10.1097/ADM.0b013e31820c23fa

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Du Rietz, E., Coleman, J., Glanville, K., Choi, S. W., O'Reilly, P. F., & Kuntsi, J. (2018). Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. *Biological psychiatry. Cognitive Neuroscience and Neuroimaging*, *3*(7), 635–643. https://doi.org/10.1016/j.bpsc.2017.11.013

El Marroun, H., Bolhuis, K., Franken, I., Jaddoe, V., Hillegers, M. H., Lahey, B. B., & Tiemeier, H. (2019). Preconception and prenatal cannabis use and the risk of behavioural and emotional problems in the offspring; a multi-informant prospective longitudinal study. *International Journal of Epidemiology*, *48*(1), 287–296. https://doi.org/10.1093/ije/dyy186

Elkins, I. J., Saunders, G., Malone, S. M., Keyes, M. A., McGue, M., & Iacono, W. G. (2018). Associations between childhood ADHD, gender, and adolescent alcohol and marijuana involvement: A causally informative design. *Drug and Alcohol Dependence*, *184*, 33–41. https://doi.org/10.1016/j.drugalcdep.2017.11.011

Fergusson, D. M., & Boden, J. M. (2008). Cannabis use and adult ADHD symptoms. *Drug and Alcohol Dependence*, *95*(1-2), 90–96. https://doi.org/10.1016/j.drugalcdep.2007.12.012

Fluharty, M. E., Sallis, H., & Munafò, M. R. (2018). Investigating possible causal effects of externalizing behaviors on tobacco initiation: A Mendelian randomization analysis. *Drug and Alcohol Dependence*, *191*, 338–342. https://doi.org/10.1016/j.drugalcdep.2018.07.015

Goldman, D. (2017). Polygenic Risk Scores in Psychiatry. *Biological Psychiatry*, *82*(10), 698–699. https://doi.org/10.1016/j.biopsych.2017.09.002

Grant, B. F., Saha, T. D., Ruan, W. J., Goldstein, R. B., Chou, S. P., Jung, J., … Hasin, D. S. (2016). Epidemiology of DSM-5 Drug Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions-III. *JAMA Psychiatry*, *73*(1), 39–47. https://doi.org/10.1001/jamapsychiatry.2015.2132

Grant, J. D., Lynskey, M. T., Madden, P. A., Nelson, E. C., Few, L. R., Bucholz, K. K., … Agrawal, A. (2015). The role of conduct disorder in the relationship between alcohol, nicotine and cannabis use disorders. *Psychological Medicine*, *45*(16), 3505–3515. https://doi.org/10.1017/S0033291715001518

Groenman, A. P., Janssen, T., & Oosterlaan, J. (2017). Childhood Psychiatric Disorders as Risk Factor for Subsequent Substance Abuse: A Meta-Analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(7), 556–569. https://doi.org/10.1016/j.jaac.2017.05.004

Groenman, A. P., Oosterlaan, J., Rommelse, N., Franke, B., Roeyers, H., Oades, R. D., … Faraone, S. V. (2013). Substance use disorders in adolescents with attention deficit hyperactivity disorder: a 4-year follow-up study. *Addiction (Abingdon, England)*, *108*(8), 1503–1511. https://doi.org/10.1111/add.12188

Gurriarán, X., Rodríguez-López, J., Flórez, G., Pereiro, C., Fernández, J. M., Fariñas, E., … GenPol Study Group. (2019). Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. *Genes, Brain, and Behavior*, *18*(3), e12504. https://doi.org/10.1111/gbb.12504

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology*, *42*(7), 608–620. https://doi.org/10.1002/gepi.22138

Hayley, A. C., Stough, C., & Downey, L. A. (2017). DSM-5 cannabis use disorder, substance use and DSM-5 specific substance-use disorders: Evaluating comorbidity in a population-based sample. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *27*(8), 732–743. https://doi.org/10.1016/j.euroneuro.2017.06.004

Hemani, G., Bowden, J., & Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, *27*(R2), R195–R208. https://doi.org/10.1093/hmg/ddy163

Icick, R., Moggi, F., Slobodin, O., Dom, G., Mathys, F., van den Brink, W., … ICASA-group. (2020). Attention Deficit/Hyperactivity Disorder and Global Severity Profiles in Treatment-Seeking Patients with Substance Use Disorders. *European Addiction Research*, *26*(4-5), 201–210. https://doi.org/10.1159/000508546

Ilbegi, S., Groenman, A. P., Schellekens, A., Hartman, C. A., Hoekstra, P. J., Franke, B., … Buitelaar, J. K. (2018). Substance use and nicotine dependence in persistent, remittent, and late-onset ADHD: a 10-year longitudinal study from childhood to young adulthood. *Journal of Neurodevelopmental Disorders*, *10*(1), 42. https://doi.org/10.1186/s11689-018-9260-y

Jang, S. K., Saunders, G., Liu, M., 23andMe Research Team, Jiang, Y., Liu, D. J., & Vrieze, S. (2020). Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychological Medicine*, *7*, 1–11. https://doi.org/10.1017/S003329172000272X

Kolla, N. J., van der Maas, M., Toplak, M. E., Erickson, P. G., Mann, R. E., Seeley, J., & Vingilis, E. (2016). Adult attention deficit hyperactivity disorder symptom profiles and concurrent problems with alcohol and cannabis: sex differences in a representative, population survey. *BMC Psychiatry*, *16*, 50. https://doi.org/10.1186/s12888-016-0746-4

Langley, K., Heron, J., Smith, G. D., & Thapar, A. (2012). Maternal and paternal smoking during pregnancy and risk of ADHD symptoms in offspring: testing for intrauterine effects. *American Journal of Epidemiology*, *176*(3), 261–268. https://doi.org/10.1093/aje/kwr510

Lee, S. H., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2012). A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology*, *36*(3), 214–224. https://doi.org/10.1002/gepi.21614

Levin, E. D., Hawkey, A. B., Hall, B. J., Cauley, M., Slade, S., Yazdani, E., … Murphy, S. K. (2019). Paternal THC exposure in rats causes long-lasting neurobehavioral effects in the offspring. *Neurotoxicology and Teratology*, *74*, 106806. https://doi.org/10.1016/j.ntt.2019.04.003

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Martínez-Luna, N., Daigre, C., Palma-Álvarez, F., Perea-Ortueta, M., Grau-López, L., Roncero, C., … Ramos-Quiroga, J. A. (2019). Psychiatric Comorbidity and Addiction Severity Differences in Patients With ADHD Seeking Treatment for Cannabis or Cocaine Use Disorders. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054719875787.

Murphy, S. K., Itchon-Ramos, N., Visco, Z., Huang, Z., Grenier, C., Schrott, R., … Kollins, S. H. (2018). Cannabinoid exposure and altered DNA methylation in rat and human sperm. *Epigenetics*, *13*(12), 1208–1221. https://doi.org/10.1080/15592294.2018.1554521

National Institute on Drug Abuse (NIDA) (2020b). Trends & statistics: Costs of substance abuse. Retrieved from https://www.drugabuse.gov/drug-topics/trends-statistics/costs-substance-abuse

Pasman, J. A., Verweij, K., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J. L., … Vink, J. M. (2018). GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature Neuroscience*, *21*(9), 1161–1170. https://doi.org/10.1038/s41593-018-0206-1

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Rabinowitz, J. A., Musci, R. J., Milam, A. J., Benke, K., Uhl, G. R., Sisto, D. Y., … Maher, B. S. (2018). The interplay between externalizing disorders polygenic risk scores and contextual factors on the development of marijuana use disorders. *Drug and Alcohol Dependence*, *191*, 365–373. https://doi.org/10.1016/j.drugalcdep.2018.07.016

Rovira, P., Demontis, D., Sánchez-Mora, C., Zayats, T., Klein, M., Mota, N. R., … Ribasés, M. (2020). Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *45*(10), 1617–1625. https://doi.org/10.1038/s41386-020-0664-5.

Smith, A., Kaufman, F., Sandy, M. S., & Cardenas, A. (2020). Cannabis exposure during critical windows of development: Epigenetic and molecular pathways implicated in neuropsychiatric disease. *Current Environmental Health Reports*, *7*(3), 325–342. https://doi.org/10.1007/s40572-020-00275-4.

Soler Artigas, M., Sánchez-Mora, C., Rovira, P., Richarte, V., Garcia-Martínez, I., Pagerols, M., … Ribasés, M. (2019). Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. *Molecular Psychiatry*, *25*(10), 2493–2503. https://doi.org/10.1038/s41380-018-0339-3.

The 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Treur, J. L., Demontis, D., Smith, G. D., Sallis, H., Richardson, T. G., Wiers, R. W., … Munafò, M. R. (2019). Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. *Addiction Biology*, e12849. https://doi.org/10.1111/adb.12849.

United Nations Office on Drugs and Crime (UNODC) (2019). World Drug Report 2019. Retrieved from https://wdr.unodc.org/wdr2019/

van de Glind, G., Konstenius, M., Koeter, M., van Emmerik-van Oortmerssen, K., Carpentier, P. J., Kaye, S., … van den Brink, W. (2014). Variability in the prevalence of adult ADHD in treatment seeking substance use disorder patients: results from an international multi-center study exploring DSM-IV and DSM-5 criteria. *Drug and Alcohol Dependence*, *134*, 158–166. https://doi.org/10.1016/j.drugalcdep.2013.09.026

van Emmerik-van Oortmerssen, K., van de Glind, G., Koeter, M. W., Allsop, S., Auriacombe, M., Barta, C., … Schoevers, R. A. (2014). Psychiatric comorbidity in treatment-seeking substance use disorder patients with and without attention deficit hyperactivity disorder: results of the IASP study. *Addiction (Abingdon, England)*, *109*(2), 262–272. https://doi.org/10.1111/add.12370

van Emmerik-van Oortmerssen, K., van de Glind, G., van den Brink, W., Smit, F., Crunelle, C. L., Swets, M., & Schoevers, R. A. (2012). Prevalence of attention-deficit hyperactivity disorder in substance use disorder patients: a meta-analysis and meta-regression analysis. *Drug and Alcohol Dependence*, *122*(1-2), 11–19. https://doi.org/10.1016/j.drugalcdep.2011.12.007

Verbanck, M., Chen, C. Y., Neale, B., & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics*, *50*(5), 693–698. https://doi.org/10.1038/s41588-018-0099-7

Vink, J. M., Hottenga, J. J., de Geus, E. J., Willemsen, G., Neale, M. C., Furberg, H., & Boomsma, D. I. (2014). Polygenic risk scores for smoking: predictors for alcohol and cannabis use? *Addiction (Abingdon, England)*, *109*(7), 1141–1151. https://doi.org/10.1111/add.12491

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wendt, F. R., Pathak, G. A., Tylee, D. S., Goswami, A., & Polimanti, R. (2020). Heterogeneity and Polygenicity in Psychiatric Disorders: A Genome-Wide Perspective. *Chronic Stress (Thousand Oaks, Calif.)*, *4*, 2470547020924844. https://doi.org/10.1177/2470547020924844

Wetherill, L., Lai, D., Johnson, E. C., Anokhin, A., Bauer, L., Bucholz, K. K., … Agrawal, A. (2019). Genome-wide association study identifies loci associated with liability to alcohol and drug dependence that is associated with variability in reward-related ventral striatum activity in African- and European-Americans. *Genes, Brain, and Behavior*, *18*(6), e12580. https://doi.org/10.1111/gbb.12580

Wilens, T. E. (2007). The nature of the relationship between attention-deficit/hyperactivity disorder and substance use. *The Journal of Clinical Psychiatry*, *68*(Suppl 11), 4–8.

Wilens, T. E., & Morrison, N. R. (2012). Substance-use disorders in adolescents and adults with ADHD: focus on treatment. *Neuropsychiatry*, *2*(4), 301–312. https://doi.org/10.2217/npy.12.39

Wimberley, T., Agerbo, E., Horsdal, H. T., Ottosen, C., Brikell, I., Als, T. D., … Dalsgaard, S. (2020). Genetic liability to ADHD and substance use disorders in individuals with ADHD. *Addiction (Abingdon, England)*, *115*(7), 1368–1377. https://doi.org/10.1111/add.14910

Wong, C. C., & Schumann, G. (2008). Review. Genetics of addictions: strategies for addressing heterogeneity and polygenicity of substance use disorders. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1507), 3213–3222. https://doi.org/10.1098/rstb.2008.0104

World Health Organization (WHO) (2018) World Health Statistics data visualizations dashboard. Tobacco smoking. Retrieved from https://apps.who.int/gho/data/node.sdg.3-a-viz?lang=en.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Vilar-Ribó L, Sánchez-Mora C, Rovira P, et al. Genetic overlap and causality between substance use disorder and attention-deficit and hyperactivity disorder. *Am J Med Genet Part B*. 2020;1–11. https://doi.org/10.1002/ajmg.b.32827

SUPPLEMENTARY MATERIAL

Supplementary Methods: Mendelian Randomization

Causality between ADHD and SUD-related phenotypes were assessed using previously published GWAS-MA summary statistics and bidirectional two-sample Mendelian Randomization (MR) (Burgess et al., 2017).

Mendelian randomization (MR) is a technique aimed at the unbiased detection of causal effects by using genetic variants as instruments. For MR to be valid the following assumptions need to be met: (i) the genetic variant(s) need to be robustly associated with the exposure, (ii) the only way genetic variant(s) may be associated with the outcome is through the exposure, and (iii) the genetic variant(s) must be independent from unobserved confounders that may influence the exposure and the outcome.

Here, we used the inversed-variance weighted (IVW) as the main method to obtain the average effect across genetic variants. This method provides an efficient estimate when all genetic variants are valid instruments (all assumptions are meet for all variants).

Additional MR methods were implemented as sensitivity analyses: MR-Egger regression (Bowden, Smith & Burgess, 2015), weighted median regression (Bowden et al., 2016a), MR-PRESSO (Verbanck et al., 2018) and Steiger filtering (Hemani, Bowden & Davey Smith, 2018). A brief description of each method is given bellow.

MR-Egger allows all variants to have horizontal pleiotropic effects (when a variant affects the exposure and the outcome independently), violating assumption (ii), as long as an additional, weaker assumption holds: direct pleiotropic effects of the genetic variants on the outcome are distributed independently of the genetic associations with the exposure (Instrument Strength Independent of Direct Effect, InSIDE assumption). MR-Egger regression measures the average pleiotropic effect across the genetic variants by estimating the intercept and test whether its value (log OR) is different from zero; if that is the case MR-Egger can also provide a causal effect accounting for the pleiotropic effects. However, Egger causal estimates can be biased when the NOME (NO Measurement Error) assumption is violated, and I^2 (true variance of the genetic exposure

association estimates divided by the variance of the genetic exposure associations estimates) can detect whether that is the case. An I^2 value above 0.9 indicates reliable Egger estimates. When $I^2 < 0.6$, Egger results were considered not valid (Bowden et al., 2016b).

The weighted median method provides a consistent estimate when up to 50% of the genetic variants are invalid instruments (violating assumptions (ii) and/or (iii)).

MR-PRESSO assumes that at least 50% of the variants are valid instruments, there is a balanced pleiotropy and the InSIDE assumption holds; it undertakes a test to detect pleiotropy and in case of pleiotropy it corrects it by outlier detection and removal.

Steiger filtering aims to detect and remove SNPs that explain more of the variance in the outcome than in the exposure, which could imply reverse causation. When detected, those SNPs were excluded from the analysis.

Supplementary Tables

Supplementary Table 1. PRS results from pre-existing GWAS-MA datasets on five SUD-related phenotypes from the in-house ADHD cohort.

Discovery sample	Phenotype Target sample	P-value Threshold	Nagelkerke's R^2	Lee's R^2	N SNPs	P-value†	OR‡	SE‡
Smoking Initiation (Liu et al., 2019)	Smoking initiation	9.50E-04	0.01	0.01	2.766	1.50E-03	1.35	1.1
Alcohol dependence (Walters et al., 2018)	Alcohol dependence	0.26	0.01	0.01	67.827	9.00E-03	1.02	1.01
Lifetime cannabis use (Pasman et al., 2018)	Lifetime cannabis use	0.07	0.01	6.33E-03	24.012	5.00E-03	1.06	1.02
Cocaine dependence (Cabana-Domínguez et al., 2019)	Cocaine dependence	7.00E-03	2.00E-03	7.16E-04	3.405	0.29	1.02	1.01
Ever addicted to illicit drugs, UK Biobank	Ever addicted to illicit drugs	5.00E-05	3.00E-03	2.21E-03	31	0.11	4.2	2.4

†Corrected for multiple testing using 10,000 permutations

‡OR and SE presented were obtained for PRS generated as $PRS_j = \sum_i S_i \times G_{ij}$, where G_{ij} is the number of risk alleles for variant i in individual j and S_i is the discovery effect size for variant i.

Supplementary Table 2. Genetic correlation results for all possible pairs of the following traits: ADHD, smoking initiation, age of smoking initiation, cigarettes per day, smoking cessation, alcohol dependence, lifetime cannabis use and cocaine dependence using previously published GWAS-MA summary statistics datasets.

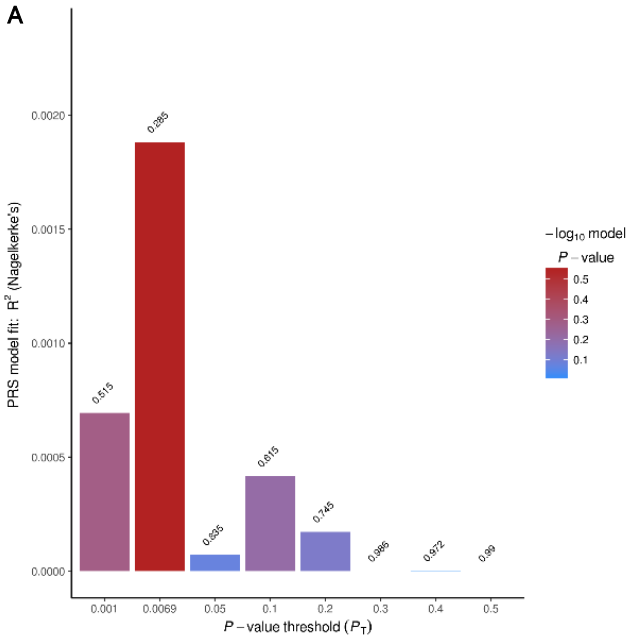
Trait 1	Trait 2	rg	SE	Z	P-value
ADHD	Lifetime cannabis use	0,15	0,04	3,57	4,00E-04
ADHD	Alcohol dependence	0,47	0,12	3,99	6,74E-05
ADHD	Cocaine dependence	0,51	0,08	6,04	1,57E-09
Lifetime cannabis use	Alcohol dependence	0,16	0,08	2	4,58E-02
Lifetime cannabis use	Cocaine dependence	0,23	0,07	3,22	1,30E-03
Alcohol dependence	Cocaine dependence	0,76	0,15	5	5,62E-07
Smoking initiation	ADHD	0,57	0,03	22,6	4,17E-113
Smoking initiation	Lifetime cannabis use	0,53	0,02	22,51	3,17E-112
Smoking initiation	Alcohol dependence	0,69	0,09	7,6	3,05E-14
Smoking initiation	Cocaine dependence	0,48	0,06	8,4	4,48E-17
Smoking initiation	Age of smoking initiation	-0,68	0,02	-30,11	3,09E-199
Smoking initiation	Cigarettes per day	0,28	0,03	8,41	3,96E-17
Age of smoking initiation	ADHD	-0,62	0,04	-14,79	1,69E-49
Age of smoking initiation	Lifetime cannabis use	-0,06	0,04	-1,44	0,15
Age of smoking initiation	Alcohol dependence	-0,68	0,1	-6,75	1,52E-11
Age of smoking initiation	Cocaine dependence	-0,49	0,08	-6,2	5,70E-10
Age of smoking initiation	Cigarettes per day	-0,37	0,04	-10,03	1,07E-23
Cigarettes per day	ADHD	0,46	0,04	11,38	5,41E-30
Cigarettes per day	Lifetime cannabis use	-0,08	0,04	-1,97	0,049
Cigarettes per day	Alcohol dependence	0,34	0,1	3,33	9,00E-04
Cigarettes per day	Cocaine dependence	0,32	0,07	4,92	8,78E-07
Smoking cessation	ADHD	0,38	0,05	7,94	2,10E-15
Smoking cessation	Lifetime cannabis use	-0,16	0,04	-4,1	4,14E-05
Smoking cessation	Alcohol dependence	0,46	0,1	4,51	6,57E-06
Smoking cessation	Cocaine dependence	0,42	0,07	5,59	2,32E-08
Smoking cessation	Smoking initiation	0,39	0,03	13,32	1,77E-40
Smoking cessation	Age of smoking initiation	-0,29	0,04	-7,06	1,71E-12
Smoking cessation	Cigarettes per day	0,44	0,03	13,64	2,29E-42

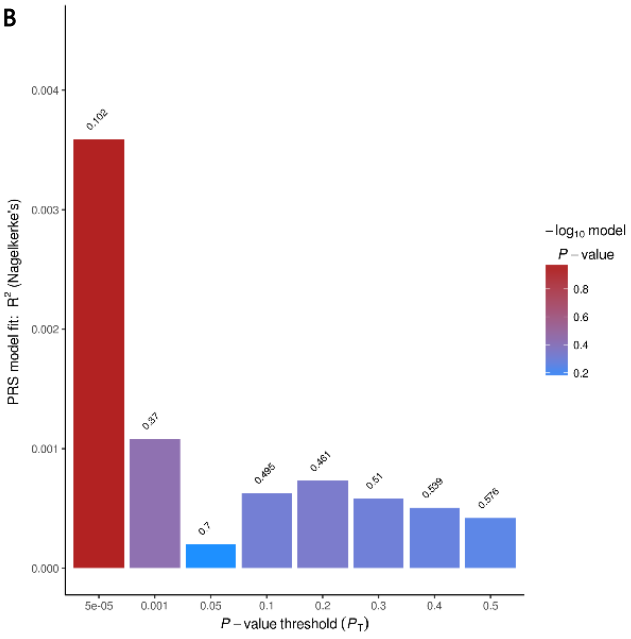
rg = genetic correlation; SE = standard error; Z = Z statistic

Supplementary table 3. I2, Cochran Q statistics and MR-PRESSO global test P-value for significant Mendelian Randomization results.

Phenotype	ADHD → SUD-related phenotype			SUD-related phenotype → ADHD		
	I^2	Cochran Q statistics P-value	MR-PRESSO global test P-value	I^2	Cochran Q statistics P-value	MR-PRESSO global test P-value
Smoking initiation	0,31	0,02	0,16	0,34	3,21E-08	1,00E-04
Age Smoking initiation	0,31	4,18E-04	1,00E-03	-	-	-
Cigarettes per day	0,31	0,23	0,24	-	-	-
Lifetime Cannabis use	0,39	0,07	0,09	0,26	0,15	0,23

Supplementary Figure 1. Bar plots showing results from the PRS analysis based on (**A**) cocaine dependence and (**B**) ever addicted to illicit drugs at broad P-value thresholds (P_T = 0.001, P_T = 0.05, P_T = 0.1, P_T = 0.2, P_T = 0.3, P_T = 0.4, P_T = 0.5) and at the best-fit PRS.





STUDY 2

Disentangling Heterogeneity in Substance Use Disorders: Insights from Genome-Wide Polygenic Scores

Laura Vilar-Ribó[1,2,3], Silvia Alemany[1,2], Judit Cabana-Domínguez[1,2,3], Natalia Llonga[1,2], Lorena Arribas[1,2], Lara Grau-López[1,2,3,4,5], Constanza Daigre[1,2,3,4,5], Bru Cormand[6,7,8,9], Noèlia Fernàndez-Castillo[6,7,8,9], Josep Antoni Ramos-Quiroga[1,2,3,4,5], María Soler Artigas[1,2,3,6 *], Marta Ribasés[1,2,3,6 *]

[1] Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain.
[2] Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain.
[3] Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain.
[4] Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain
[5] Addiction and Dual Diagnosis Unit, Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[6] Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain.
[7] Biomedical Network Research Centre on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain.
[8] Sant Joan de Déu Research Institute (IRSJD), Esplugues de Llobregat, Catalonia, Spain
[9] Institute of Biomedicine of the University of Barcelona (IBUB), Barcelona, Catalonia, Spain.

*Corresponding authors

Manuscript under review in Translational Psychiatry

ABSTRACT

Substance use disorder (SUD) is a global health problem with significant impact on individuals and society. The presentation of SUD is diverse, involving various substances, ages at onset, comorbid conditions, and disease trajectories. Current treatments for SUD struggle to address this heterogeneity, resulting in high relapse rates. SUD often co-occurs with other psychiatric and mental health-related conditions that contribute to the heterogeneity of the disorder and predispose to adverse disease trajectories. Family and genetic studies highlight the role of genetic and environmental factors in the course of SUD, and point to a shared genetic liability between SUDs and comorbid psychopathology. In this study, we aimed to disentangle SUD heterogeneity using a deeply phenotyped SUD cohort and polygenic scores (PGSs) for psychiatric disorders and related traits. We explored associations between PGSs and various SUD-related phenotypes, as well as PGS-environment interactions using information on lifetime emotional, physical and/or sexual abuse. Our results revealed different patterns of associations between the genetic liability for mental health-related traits and SUD-related phenotypes, which may help explain part of the heterogeneity observed in SUD. In our SUD sample, we found associations linking the genetic liability for ADHD with lower educational attainment, the genetic liability for PTSD with higher rates of unemployment, the genetic liability for educational attainment with lower rates of criminal records and unemployment and the genetic liability for well-being with lower rates of outpatient treatments and fewer problems related to family and social relationships. We also found evidence of PGS-environment interactions showing that genetic liability for suicide attempt worsened the psychiatric status in SUD individuals with a history of emotional physical and/or sexual abuse. Collectively, these data contribute to a better understanding of the role of the genetic liability for mental health-related conditions and adverse life experiences in SUD heterogeneity.

INTRODUCTION

Substance use disorder (SUD) is a growing global health problem impacting the individual's life and the society as a whole. In 2019, 3.2 million people died due to SUD-related causes with 300.000 deaths due to drug or alcohol overdose (Roth et al., 2018).

The presentation of SUD is highly heterogeneous across a wide range of phenotypic outcomes such as type of substance(s) (Compton et al., 2021), age at onset of SUD (Christoffersen et al., 2021; Poudel & Gautam, 2017), individual personality profiles (Bucher et al., 2019; Zilberman et al., 2018), presence of comorbid conditions (Roehrs et al., 2021) and disease trajectory (Richmond-Rakerd et al., 2017). For instance, polysubstance use, present in approximately 50% of individuals with SUD (Morley et al., 2015), has been associated with poorer treatment outcomes (Andersson, Lauvsnes, et al., 2021), higher rates premature death due to overdose (Compton et al., 2021) and higher rates of mental-health problems and risky behaviors (Morley et al., 2015). Early onset substance users are at higher risk for psychosocial problems (Poudel & Gautam, 2017), unemployment (Melchior et al., 2015), low educational attainment (Christoffersen et al., 2021) and heavier drug abuse in adulthood (Richmond-Rakerd et al., 2017). The presence of comorbid psychiatric disorders has been associated with adverse disease trajectory, such as poorer treatment adherence in individuals with comorbid major depressive disorder or Attention-Deficit Hyperactivity disorder (ADHD) (Andersson, Lauvsnes, et al., 2021; Ostacher, 2007), increased rates of suicide in individuals with comorbid schizophrenia (Lähteenvuo et al., 2021), and worse physical and mental health in individuals with comorbid Post-Traumatic Stress Disorder (PTSD) (Mills et al., 2006). In addition, behavioral traits, such as neuroticism, have been associated with lower rates of abstinence and increased symptom severity (Bucher et al., 2019). Most available inpatient and outpatient treatments for SUD, however, are not well suited to accommodate the observed heterogeneity, resulting in high rates of early treatment termination and relapse (Syan et al., 2020).

Twin and adoption studies support the role of moderate to high (30-70%) genetic influence on SUD (Agrawal & Lynskey, 2008) and genome-wide associations studies (GWASs) have identified risk loci associated with substance-specific SUDs (Deak, Zhou,

et al., 2022; E. C. Johnson, Demontis, et al., 2020; Kranzler et al., 2019; Sun et al., 2020). These studies, together with other genetic approaches, point to a shared genetic liability and a unitary genetic architecture of SUD across different substances (Hatoum et al., 2022; Schoeler et al., 2022). In addition, SUD genetic liability, which can be assessed using polygenic scores (PGSs), presents substantial overlap with psychiatric disorders and behavioral traits (Hatoum et al., 2023), and shows the strongest genetic correlations with ADHD, PTSD, anxiety, schizophrenia, depression, bipolar disorder and risk-taking behaviors (Cabana-Domínguez et al., 2019; Gelernter & Polimanti, 2021; Schoeler et al., 2022). Supporting this idea, a recent study in a deeply phenotyped SUD sample reported that PGSs for substance-specific SUDs were associated with their primary substance-related phenotypes but also with major depressive disorder, PTSD, lifetime trauma assessment, being suspended from school or family history of SUD (Kember et al., 2023). This finding suggests that the genetic liability for co-occurring psychopathology may explain part of the heterogeneity found in SUD.

In addition, there is growing evidence that the effect of genetic risk on SUD can be moderated by environmental factors, which may also contribute to the individual differences in addictive behaviors (Vink, 2016). For instance, adverse life experiences, such as trauma exposure or peer drug use, seem to moderate the effect of PGS for cannabis use on lifetime cannabis use (Meyers et al., 2019), the effect of PGS for bipolar disorder on alcohol misuse (Polimanti et al., 2018) and the effect of PGS for alcohol problems in adults on earlier alcohol problems (Salvatore et al., 2014)

In the present study, we aim to disentangle SUD heterogeneity in a SUD cohort of 1427 individuals who underwent deep phenotyping by conducting a systematic investigation of associations between 39 SUD-related phenotypes and the genetic liability for psychiatric disorders and related traits using PGSs, and to assess whether the profile of PGS associations across SUD-related phenotypes is modulated by exposure to lifetime emotional, physical and/or sexual abuse.

MATERIALS AND METHODS

Sample description

A total of 1427 individuals with SUD were recruited at the Addiction and Dual Diagnosis Unit of Hospital Universitari Vall d'Hebron, Barcelona, Spain. Inclusion criteria were age over 18 years old, substance abuse or dependence according to the DSM-IV criteria, European ancestry and a signed informed consent prior to participation. The project was approved by the Ethics Committee at the Hospital Universitari Vall d'Hebron.

Clinical assessment

The clinical assessment was conducted by trained psychiatrists and psychologists in two different steps: (i) At recruitment, a questionnaire designed ad hoc was administered to gather information on sociodemographic status (sex, age, educational attainment, employment status and criminal record), lifetime medical conditions, psychiatric and SUD family history and substance use related variables (substance(s) of use and/or abuse, age at onset of use, age at onset of SUD, years of substance use and SUD treatment history); (ii) The follow-up interviews were divided into four sessions to evaluate SUD severity, DSM-IV axis I and axis II disorders, heath-related quality of life and personality traits with different scales and questionnaires (Figure 1), detailed below.

The structured Clinical Interview for Axis I and II Disorders of the DSM-IV (SCID-I and SCID-II) (J. G. Young, 1967) and the Conners' Adult ADHD diagnostic interview for DSM-IV (CAADDID-II) (Ramos-Quiroga et al., 2012) were used to assess psychiatric comorbidity. The Spanish version of the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ) (Zuckerman et al., 1993) was used to assess personality features including neuroticism-anxiety, activity, sociability, impulsive sensation-seeking and aggression-hostility. The validated Spanish version of the European Addiction Severity Index interview (EuropASI) (Kokkevi & Hartgers, 1995) is design to provide information about aspects of an individual's life which may contribute to his/her substance abuse, specifically on the following areas: legal status, employment status, medical status, psychiatric status, drug use, alcohol use and family/social relationships. Scores ranging from 0 to 1 are estimated, with higher scores indicating greater severity. The 36-Item

Short Form Survey (SF-36) was administered to measure the self-reported health and quality of life, both physically and mentally with higher scores indicating better health (Ware & Sherbourne, 1992). After data curation, 39 phenotypes with sample size >300 were considered and classified into three categories: SUD variables (n= 8), comorbidity and personality traits (n=15) and sociodemographic and health outcomes (n=16) (Table 1). For binary traits, with n1 individuals in one group and n2 individuals in the other group, effective sample size was calculated with the formula 4/(1/n1+1/n2).

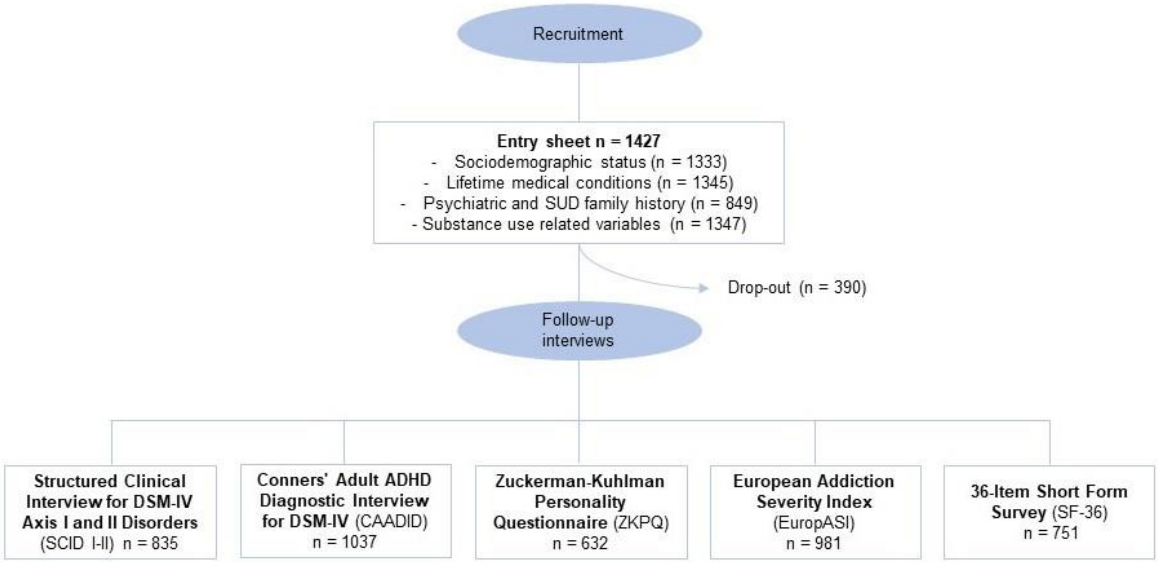


Figure 1. Flowchart. Flowchart with the different stages of the study including the recruitment and follow-up of the SUD sample. Sample size (n) refers to the number of individuals with at least one item of the interview available.

Genotyping and quality control

Genomic DNA was isolated from whole blood by the salting-out procedure and genotyped with the Illumina Infinium Global Screening Array-24 version 2 (GSA v2) (Illumina, CA, San Diego, USA) in two different waves (434 and 993 samples, respectively). Pre-imputation quality control was done with the PLINK 2.0 software (Chang et al., 2015) and included individual and variant filtering based on the following parameters: variant call rate >0.95 (before individual filtering), individual call rate >0.98, autosomal heterozygosity deviation (| Fhet | <0.2), variant call rate >0.98 (after individual filtering), difference in variant missingness between cases and controls <0.02, SNP Hardy-Weinberg equilibrium (HWE) (p >1e−06 in controls or p >1e−10 in cases) and minor allele frequency (MAF) > 0.01. Genetic outliers were identified by principal component analysis

(PCA) using PLINK 2.0 and the mixed ancestry 1000G reference panel (Auton et al., 2015). Non-European individuals were excluded if their principal component (PC) values for PC1 and PC2 were greater than 1 standard deviation from the mean-centring point for the study population. Related and duplicated samples were identified by the "KING-robust kinship estimator" analysis in PLINK 2.0 (Manichaikul et al., 2010) and one individual was excluded from each pair of subjects with kinship coefficient > 0.0442. Imputation was done with McCarthy tools, for data preparation, and the Michigan Imputation Server (Das et al., 2016), using the Haplotype Reference Consortium (HRC Version r1.1 2016) reference panel (GRCh37/hg19). Variants were excluded in case of allele mismatch between the reference panel and the study dataset (chi2>900). Post-imputation dosage files with imputation INFO score >0.8 and MAF >0.01 were considered for subsequent analyses.

Polygenic scores

PGSs were constructed in our using the PRS-CS software (Ge et al., 2019), PLINK 2.0 and available GWASs summary statistics (**Table S1**), on psychiatric disorders (ADHD (Demontis, Walters, et al., 2019) anxiety (http://www.nealelab.is/uk-biobank/), bipolar disorder (Mullins et al., 2021), depression (Howard et al., 2019), post-traumatic Stress Disorder (PTSD) (Nievergelt et al., 2019) and schizophrenia (Trubetskoy, Pardiñas, Qi, Panagiotaropoulou, van Os, et al., 2022)), behavioral traits (risk tolerance (Karlsson Linnér et al., 2019), suicide attempt (Docherty et al., 2022)) and other related traits (educational attainment (J. J. Lee et al., 2018) and well-being (Baselmans et al., 2019)). PGSs were computed and standardized to a mean of 0 and a standard deviation of 1 for all disorders and traits.

Statistical analysis

Association between polygenic scores and SUD-related phenotypes

The profile of PGSs associations across the SUD-related phenotypes were assessed with the appropriate regression models depending on the nature of the outcome variable with R: logistic regression for binary variables, linear regression for continuous variables, ordinal regression for ordinal categorical variables and negative

binominal for count variables. Prior to the analysis, logarithmic transformations were applied to continuous variables not following a normal distribution ("age at onset of substance use" and "age at onset of SUD"). Additionally, linear regression residuals were checked for continuous variables with significant results to ensure they followed a normal distribution. Age, sex, genotyping batch and the 10 first PCs were included as covariates in all analyses. Additionally, for the variables "age at onset of substance use", "age at onset of SUD", "years between substance use and SUD", and "years of substance use as a proportion of lifespan", the main drug of use, abuse or dependence was included as a covariate. *P*-values were corrected for multiple comparisons using PhenoSpD (Nyholt, 2004; Zheng et al., 2018), a command line R based tool for estimating phenotypic correlations and multiple testing correction. The effective number of independent variables estimated was 35 using the VeffLi model and the corrected *p*-value threshold was set at *p*-value < 1.46e-03 (J. Li & Ji, 2005).

Interaction between polygenic scores and emotional, physical and/or sexual abuse in SUD-related phenotypes

For those PGSs associated with any outcome, interaction with emotional, physical and/or sexual abuse was tested in a subset of 735 individuals who had completed the EuropASI family/social relationships questionnaire and information on emotional, physical and/or sexual abuse was available. Potential interaction effects were tested introducing an interaction term (PGS*abuse) in the regression model adjusted for age, sex, genotyping batch and the 10 first PCs as covariates. Multiple comparison corrected *p*-value, calculated with PhenoSpD in R, was set at $p < 2.05e-03$ (Nyholt, 2004; Zheng et al., 2018). For significant interactions, PGS-outcome associations were stratified by exposure to emotional, physical and/or sexual abuse.

RESULTS

Our cohort consisted of 1427 individuals (76.5% male), with a mean age of 38.6 years (SD = 10.3) (**Table 1**). The vast majority of subjects were polysubstance users and 47% fulfilled SUD criteria for three or more substances.

Table 1. Summary of the 39 SUD-related phenotypes

Phenotypes	*n* [a]	Summary
Sex: Male (%)	1427	76.5
Age; Mean (SD)	1427	38.6 (10.3)
SUD variables		
Age at onset of substance use; Mean (SD)	1347	17.05 (1.36)
Age at onset of SUD; Mean (SD)	1339	19.32 (1.4)
Years between substance use and SUD; Median (IQR)	1325	1 (3)
Years of substance use as proportion of lifespan; Median (IQR)	1148	34 (33)
Number of substances consumed	869	
1 (%)		23.2
2 (%)		29.5
3 or more (%)		47.3
Number of therapeutic community interventions; Median (IQR)	1314	
0 (%)		64.5
1 (%)		22.7
2 (%)		7.1
3 or more (%)		5.7
Number of inpatient detoxifications; Median (IQR)	1327	
0 (%)		70.5
1 (%)		16.0
2 (%)		6.8
3 or more (%)		6.7
Number of outpatient treatments; Median (IQR)	1252	
0 (%)		28.5
1 (%)		37.9
2 (%)		16.6
3 or more (%)		17
Comorbidity and personality traits		
Mental disorders in DSM-IV		
Borderline personality disorder (%)	447	15.9
Major depressive disorder (%)	828	37.8
Antisocial personality disorder (%)	560	21.3
Psychotic disorder (%)	593	7.8
Anxiety disorder (%)	654	24.9
Attention deficit hyperactivity disorder (%)	700	21.5
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)		
Neuroticism Anxiety personality factor; Mean (SD)	663	10.8 (4.9)
Aggression Hostility personality factor; Mean (SD)	667	8.93 (3.15)
Sociability personality factor; Mean (SD)	632	6.6 (3.4)
Impulsive sensation seeking personality factor; Mean (SD)	666	10.5 (4.3)
Activity personality factor; Mean (SD)	665	8.09 (3.5)
Suicide attempt (%)	618	47.8
Suicide ideation (%)	731	30.1
Psychotic symptoms (%)	1281	58.6

Sleeping disturbances (%)	1228	54.6
Sociodemographic and health outcomes		
EuropASI		
Legal status; Median (IQR)	984	0 (.1)
Employment status; Median (IQR)	984	.55 (.5)
Medical status; Median (IQR)	982	0 (.1)
Psychiatric status; Median (IQR)	984	.4 (.3)
Drug use; Median (IQR)	984	.2 (.2)
Alcohol use; Median (IQR)	984	.1 (.3)
Family/Social relationships; Median (IQR)	981	.4 (.5)
36-Item Short Form Survey (SF-36)		
Physical health; Mean (SD)	751	48.4 (1.8)
Mental health; Mean (SD)	751	35.5 (13.7)
Criminal record (%)	713	43.1
Unemployment (%)	1057	73.4
Number of psychiatric hospitalizations; Median (IQR)	760	
0 (%)		79.9
1 (%)		9.9
2 (%)		4.5
3 or more (%)		5.7
Psychiatric family history (%)	840	44.8
Lifetime medical conditions (%)	1334	54.4
Substance use family history (%)	818	59.6
Educational attainment	1333	
1 (Incomplete primary school) (%)		14.5
2 (Primary school) (%)		40.9
3 (Secondary/High school) (%)		35.6
4 (Bachelor's degree or higher) (%)		9.1

[a] For binary traits, with n1 individuals in one group and n2 individuals in the other group, effective sample size was calculated with the formula 4/(1/n1+1/n2).

Polygenic scores for psychiatric disorders

After multiple testing correction we found significant associations between the PGS for ADHD and lower educational attainment (OR=0.85, 95% CI [0.93, 0.77], p=1.20e-03) and between the PGS for PTSD and unemployment (OR=1.23, 95% CI [1.09, 1.4], p=1.00e-03) **(Figure 2, Table S2a and S2e)**.

Despite not surpassing the multiple testing correction threshold ($p < 0.05$), we found additional nominal associations. PGS for ADHD was associated with early-onset of first substance use (Beta (β)=-0.01, 95% CI [-0.03, -3.00e-04]), longer term substance use as a proportion of the lifespan (β=1.07, 95% CI [0.03, 2.11]), lifetime diagnosis of ADHD

(OR=1.24, 95% CI [1.06, 1.45]), and antisocial personality disorder (OR=1.24, 95% CI [1.04, 1.47]) **(Figure 2, Table S2a)**. PGS for anxiety was associated with more outpatient treatments (incidence rate ratio (IRR)=1.09, 95% CI [1.03, 1.15]), psychotic disorders across the lifetime (OR=1.49, 95% CI [1.16, 1.92]), poorer self-perceived physical health status measured with the SF-36 instrument (β=-0.93, 95% CI [-1.64, -0.22]) and psychiatric family history (OR=1.16, 95% CI [1.01, 1.33]) **(Figure 2, Table S2b)**. PGS for bipolar disorder was associated with higher rates of psychotic symptoms (OR=1.14, 95% CI [1.02, 1.28]), unemployment (OR=1.14, 95% CI [1.01, 1.29]), psychiatric hospitalizations (IRR=1.26, 95% CI [1.03, 1.53]) and substance use family history (OR=1.15, 95% CI [1.0, 1.32]) **(Figure 2, Table S2c)**. PGS for depression showed association with slower transition from substance use to SUD (IRR=1.1, 95% CI [1.01, 1.21]), more outpatient treatments (IRR=1.09, 95% CI [1.03, 1.15]) and higher rates of neuroticism-anxiety (β=0.36, 95% CI [4e-3, 0.72]) and aggression-hostility (β=0.27, 95% CI [0.04, 0.5]) according to the ZKPQ, suicide attempts (OR=1.18, 95% CI [1.02, 1.37]), criminal records (OR=1.27, 95% CI [1.09, 1.48]) and psychiatric family history (OR=1.16, 95% CI [1.01, 1.32]) **(Figure 2, Table S2d)**. PGS for PTSD was associated with early-onset of first substance use (β=-0.01, 95% CI [-0.03, -1e-3]), more inpatient detoxifications (IRR=1.13, 95% CI [1.01, 1.27]), and lower educational attainment (OR=0.87, 95% CI [0.97, 0.79]) **(Figure 2, Table S2e)**. And lastly, PGS for schizophrenia was associated with later onset of substance use (β=0.01, 95% CI [3e-4, 0.03]), psychotic disorders (OR=1.37, 95% CI [1.05, 1.78]), higher rates of psychotic symptoms (OR=1.15, 95% CI [1.03, 1.29]) and unemployment (OR=1.13, 95% CI [1, 1.28]) **(Figure 2, Table S2f)**.

Polygenic scores for behavioral traits

None of the associations between PGSs for behavioral traits and SUD-related phenotypes surpassed multiple testing correction, however, nominally significant associations ($p < 0.05$) are detailed bellow.

PGS for risk tolerance showed associations with more outpatient treatments (IRR=1.06, 95% CI [1, 1.12]), lower rates of neuroticism-anxiety (β=-0.51, 95% CI [-0.88, -0.14]) and higher rates of impulsive sensation seeking (β=0.32, 95% CI [3e-3, 0.65]) according to the ZKPQ and higher rates of legal problems measured by the EuropASI

index (OR=1.18, 95% CI [1.02, 1.36]) **(Figure 2, Table S2g)**. PGS for suicide attempt was associated with early-onset of SUD (β=-0.02, 95% CI [-0.04, -0.01]), more outpatient treatments (IRR=1.06, 95% CI [1, 1.12]), higher rates of aggression-hostility (β=0.26, 95% CI [0.02, 0.51]) and psychotic symptoms (OR=1.15, 95% CI [1.03, 1.29]), more legal (OR=1.18, 95% CI [1.01, 1.37]), medical (OR=1.14, 95% CI [1, 1.29]), psychiatric (OR=1.13, 95% CI [1.01, 1.27]) and family/social (OR=1.13, 95% CI [1, 1.26]) problems measured by the EuropASI index, more lifetime medical conditions (OR=1.14, 95% CI [1, 1.29]) and lower educational attainment (OR=0.90, 95% CI [1, 0.81]) **(Figure 2, Table S2h)**.

Figure 2. Heatmap for the results of the association between PGSs and SUD-related phenotypes. Association pattern between 10 PGSs for psychiatric disorders, behavioural and related traits with the SUD-related phenotypes; *nominal significance *p*-values; **p*-values that passed multiple testing correction using PhenoSpD (P<1.46e-03). ADHD = Attention-deficit hyperactivity disorder; PTSD = Post-traumatic Stress Disorder. Standardized coefficient corresponds to Beta for continuous variables, log(OR) for binary and ordinal variables and log(IRR) for count variables.

Polygenic scores for educational attainment and well-being

After multiple testing correction we found significant associations between PGS for educational attainment and less criminal records (OR=0.67, 95% CI [0.57, 0.78], *p*=8.03e-07), unemployment (OR=0.82, 95% CI [0.71, 0.92], *p*=1.34e-03) and higher educational attainment (OR=1.39, 95% CI [1.54, 1.25], *p*=3.34e-10), as well as associations between PGS for well-being and less outpatient treatments (OR=0.91, 95% CI [0.86, 0.96], *p*=7.00e-04) and family/social problems (OR=0.83, 95% CI [0.74, 0.93], *p*= 1.00e-03) **(Figure 2, Table S2i and S2j)**.

Nominal associations (*p* < 0.05) include the association between PGS for educational attainment and later onset of substance use (β=0.02, 95% CI [0.01, 0.03]) and SUD (β=0.02, 95% CI [4e-3, 0.04]), less therapeutic community interventions (OR=0.89, 95% CI [0.81, 0.98]) or outpatient treatments (OR=0.91, 95% CI [0.87, 0.97]), lower rates of neuroticism-anxiety (β=-0.4, 95% CI [-0.78, -0.01] 2), less social-familiar problems (OR=0.86, 95% CI [0.77, 0.96]), and substance use family history (OR=0.80, 95% CI [0.7, 0.92]) **(Figure 2, Table S2i)**. Moreover, PGS for well-being showed association with lower rates of psychiatric problems (OR=0.87, 95% CI [0.78, 0.97]), less unemployment (OR=0.88, 95% CI [0.77, 0.99]), psychiatric hospitalizations (OR=0.78, 95% CI [0.68, 0.95]), and substance use family history (OR=0.83, 95% CI [0.77, 0.99]) **(Figure 2, Table S2j)**.

Interaction between polygenic scores and emotional, physical and/or sexual abuse on SUD-related phenotypes

Information on lifetime emotional, physical and/or sexual abuse was available for a total of 735 individuals with SUD, 45.6% of which (n=335) reported having experienced some sort of abuse across their lifetime. PGS*abuse interaction analysis was performed for those PGSs nominally associated with any outcome **(Table S3).** We found one significant interaction where lifetime abuse moderates the association between PGS for suicide attempt and the psychiatric status measured by the EuropASI index (OR=1.35, 95% CI [1.03, 1.78], *p*=2.94e-02). Specifically, the genetic liability for suicide attempt was associated with worse psychiatric status scores among those having experienced lifetime emotional, physical and/or sexual abuse (OR=1.33 95% CI [0.48, 0.09], *p*= 4.67e-04), while the association was not significant for those not exposed **(Figure 3)**.

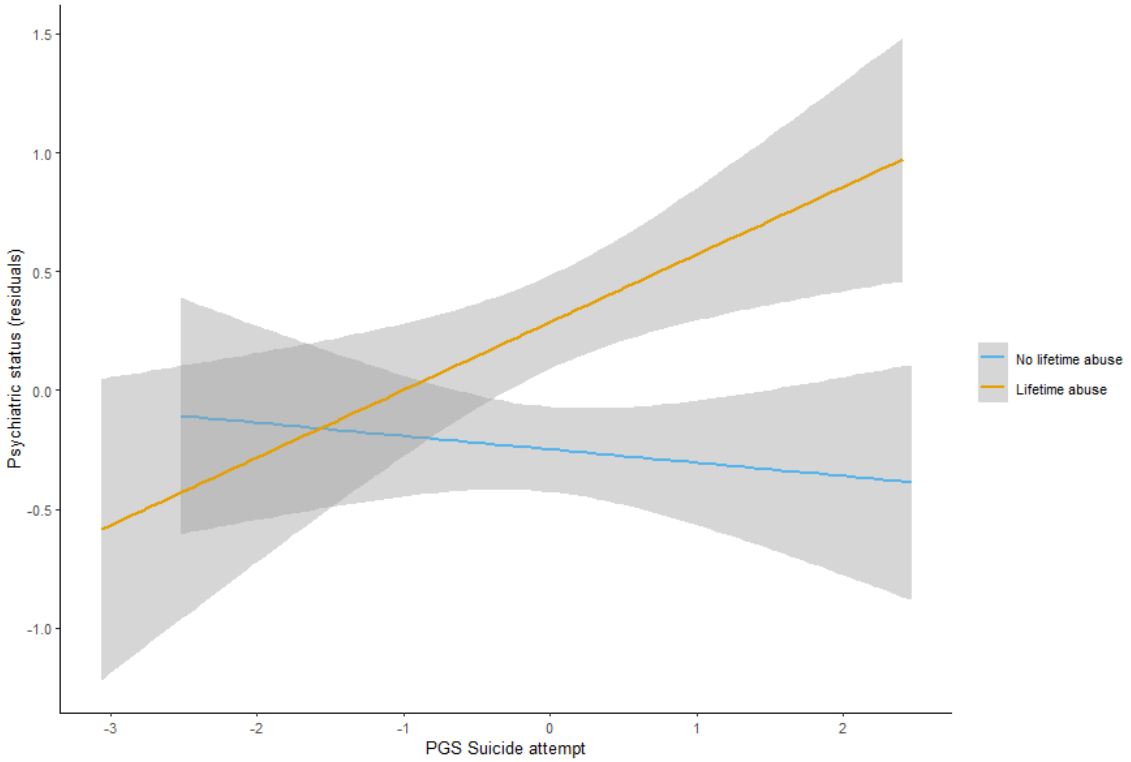


Figure 3. Statistically significant result from the interaction analysis. Interaction between PGS for suicide attempt and lifetime emotional, physical and/or sexual abuse in psychiatric status measured by the EuropASI index. The X axis presents the PGS for suicide attempt, and the Y axis shows the residuals from an ordinal regression model with psychiatric status as outcome adjusted for age, sex, genotyping batch and the 10 first principal components for those individuals who suffered lifetime abuse (in yellow) and those who did not (in green).

DISCUSSION

There is immense clinical and genetic heterogeneity among individuals with SUD, and current treatment approaches fail to accommodate this variability, resulting in poor treatment adherence and high rates of relapse (Prom-Wormley et al., 2017). In this study, we utilized multidimensional data from a deeply phenotyped SUD cohort and individual genetic liability information for a broad range of mental health-related traits using PGSs, to provide new insights into the heterogeneity of the disorder. Our approach included the systematic association of 10 PGSs for psychiatric disorders, behavioral and other related traits with 39 SUD-related phenotypes, and the assessment of PGS-environmental interactions using information on emotional, physical and/or sexual abuse. Our main

findings suggest that the genetic liability for ADHD, PTSD and suicide attempt, in conjunction with environmental factors, may underlie, at least partially, the observed heterogeneity in SUD-related phenotypes such as educational attainment, unemployment and psychiatric status.

PGSs analysis on the SUD-related phenotypes builds on previous findings supporting links between the genetic risk for psychiatric disorders and a wide variety of SUD outcomes. In line with this evidence, our results suggest that the genetic liability for mental health-related traits exhibits different patterns of associations with SUD-related phenotypes. Specifically, we replicated previous findings linking the genetic liability for ADHD with lower educational attainment (Dardani et al., 2021), the genetic liability for PTSD with higher rates of unemployment (Goldberg et al., 2014), and the genetic liability for higher educational attainment with lower rates of criminal records and unemployment (H. Liu, 2019; Wertz et al., 2018). While these associations were described in the general population, our results suggest that these patterns remain in individuals with SUD. Moreover, our findings showed that the genetic liability for well-being is associated with better outcomes, namely lower rates of outpatient treatments and fewer problems related to family and social relationships, which is consistent with the role of the genetics underlying well-being in healthy family relationships (van de Weijer et al., 2022).

Despite not surpassing multiple comparison correction, we found evidence supporting previously reported associations. For instance, PGSs for ADHD, schizophrenia and educational attainment were associated with their respective primary phenotype, confirming the validity of the approach. In addition, we identified an association between the genetic liability for depression and higher rates of suicide attempts. This is consistent with previous findings linking PGSs for depression with suicide attempt (P. H. Lee et al., 2022; Levey et al., 2019; Lim et al., 2020; Ruderfer et al., 2020), and studies suggesting an increased risk of suicide attempt and ideation among individuals with comorbid SUD and major depressive disorder (Onaemo et al., 2022; Østergaard et al., 2017). Our findings add to the evidence supporting that genetic liability for depression may have a relevant role regarding suicide attempt and ideation in the context of SUD.

Our results also shed light into the association between the genetic liability for multiple psychiatric disorders and poor SUD-related outcomes. These include early age at onset of substance use and high number of outpatient treatments, strengthening the notion that genetic susceptibility to psychiatric diseases and behavioral traits may play a role in promoting the initiation and impeding the cessation of substance use (Andersson et al., 2019; Bagot et al., 2015; Krawczyk et al., 2017; Ostacher, 2007). For instance, we found that individuals with higher PGS for ADHD showed earlier onset and more prolonged substance use, while those with higher PGS for depression showed faster transition from substance use to SUD and more outpatient treatments. Similarly, individuals with higher PGS for PTSD showed earlier onset of substance use and more inpatient treatments.

Moreover, PGSs for educational attainment or suicide attempt were associated with multiple outcomes (more than ten). Increased genetic risk for educational attainment was associated with less therapeutic interventions, late age at onset of substance use or SUD and less SUD family history or problems related with family and social relationships. These findings are consistent with previous evidence showing that the genetic liability for education attainment is linked to decreased SUD severity (Salvatore et al., 2020) and a recent study by Kinreich et al., (2021) suggesting that polygenic liability to years of education could be used to predict remission in patients with alcohol use disorder. Additionally, the genetic liability for suicide attempt showed the strongest association with early age at onset of SUD, number of outpatient treatments, higher rates of psychotic symptoms, and a wide range of medical, psychological and legal problems. Adding to this evidence, we report a significant interaction between PGS for suicide attempt and having been exposed to lifetime emotional, physical or sexual abuse in the psychiatric status of SUD individuals. While it is well established that exposure to sexual trauma and/or abuse increases the risk for substance use and mental health problems later in life (B. S. O'Brien & Sher, 2013), we found that the genetic liability for suicide attempt exacerbates the negative impact on mental health problems in individuals with a history of abuse. Similar findings have been reported for cannabis use (Meyers et al., 2019) or bipolar disorder (Park et al., 2020), where exposure to trauma and/or maltreatment potentiates the polygenic risk for these

disorders. Overall, these results highlight that focusing on exposed individuals may render genetic effects that may not be found when environmental exposures are not considered.

Although many results are well supported by prior research, we also found that for some disorders PGSs were not associated with their primary phenotype. For instance, PGSs for anxiety or depression did not show an association with anxiety disorder or major depressive disorder in the SUD dataset. The reasons for this lack of association may include selection bias, complex relationships between SUD and comorbid conditions and limited sample size for some of the outcomes. Moreover, PGS for suicide attempt was not associated with suicide behaviors, namely suicide ideation and attempt, in our SUD dataset. Suicide attempt is a clinically complex phenotype that can vary greatly in frequency and intensity (Diblasi et al., 2021). Even though the GWAS meta-analysis used to construct PGS for suicide attempt aimed to harmonize data across various cohorts by including clinical samples from major psychiatric disorders and individuals from the Million Veterans Project sample (Docherty et al., 2022), differences in population characteristics or assessment methods of the phenotype may account for the inconclusive results observed in our dataset. In previous studies, reliability of PGS-based predictions of suicide attempt has been inconsistent when applied to independent datasets (Loughnan et al., 2022; Mitjans et al., 2022; Mullins et al., 2014), and Lannoy et al., (2022) found evidence for the interaction between PGS for suicide attempt and drug use on suicide ideation. Together, these results highlight the multifactorial nature of suicide attempt and suggest that other factors, such as psychiatric comorbidity, SUD type or severity and environmental factors, should be taken into account when assessing suicide risk.

It is important to be cautious when comparing results from PGSs for the disorders and traits tested, taking into consideration the variations in statistical power between some of them. The differences in sample size among the GWAS meta-analyses used to construct PGSs, as well as among the outcomes, could have contributed to the uneven pattern of associations observed. Additionally, other factors such as environmental factors and sex differences may play a significant role in certain aspects of SUD heterogeneity. Furthermore, our results suggest that the patterns of lifetime comorbidity

in SUD result, in part, from the contribution of genetic factors. However, it is currently unknown whether substance use is a consequence of underlying psychiatric disorders or whether it increases the risk of mental health problems later in life. Access to longitudinal data would provide new and valuable information to assess causal relationships between SUD and comorbid conditions and to examine the impact of the genetic liability on disease progression.

This study supports that the genetic liability for distinct mental health-related traits plays a role in the heterogeneity of SUD and can influence disease outcome in terms of severity, comorbidity rates and socio-demographic factors. There is also evidence for PGS-environment interactions between the genetic liability for suicide attempt and lifetime emotional physical and/or sexual abuse on the psychiatric status of individuals with SUD. These results encourage the use of PGSs and gene-environment interactions to better understand the heterogeneity of SUD and complex traits.

Author Contributions: L.V.R., and L.A. participated in the DNA isolation and preparation of samples. L.V.R., M.S.A. and M.R. undertook the statistical analyses. L.V.R., S.A., J.C.D., N.L., M.S.A., and M.R. contributed to the interpretation of data for the work. L.G.L., and C.D., contributed to the clinical assessment and recruitment of patients. J.A.R.Q. participated in the study design, clinical assessment and coordination of the clinical research. M.R. and M.S.A. conceived the project, wrote the protocol and coordinated the study design and the statistical analyses. M.S.A. and M.R. supervised the project and the manuscript preparation. All authors revised the work critically for important intellectual content and have approved the final version.

Funding: This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-1461, 2021SGR-00840 and 2021-SGR-01093), the Instituto de Salud Carlos III (P19/01224, PI20/00041, PI22/00464 and FI18/00285 to L.V.R, CP22/00128 to M.S.A and CP22/00026 to S.A), the Ministry of Science, Innovation and Universities (IJC2018-035346-I to M.S.A, RYC2021-031324-I to J.C.D and PID2021-1277760B-I100 to B.C. and N.F-C.), the Ministry of Health, Social Services and Equality (PNSD-2020I042 to N.F-C.), the Network Center for Biomedical Research (CIBER) to J.C.D., the European Regional Development Fund (ERDF), the European Union H2020 Programme (H2020/2014-2020) under grant agreements no. 848228 (DISCOvERIE) and no. 2020604 (TIMESPAN), the ECNP Network 'ADHD across the Lifespan', "La Marató de TV3" (202228-30 and 202228-31) and ICREA Academia 2021.

Conflict of interest: J.A.R.Q was on the speakers' bureau and/or acted as consultant for Biogen, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Rubió, Uriach, Technofarma and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious, and Rubió. All other authors declare no biomedical financial interests or conflicts of interest.

REFERENCES

Agrawal, A., & Lynskey, M. T. (2008). Are there genetic influences on addiction: evidence from family, adoption and twin studies. Addiction (Abingdon, England), 103(7), 1069–1081. https://doi.org/10.1111/j.1360-0443.2008.02213.x

Andersson, H. W., Lauvsnes, A. D. F., & Nordfjærn, T. (2021). Emerging Adults in Inpatient Substance Use Treatment: A Prospective Cohort Study of Patient Characteristics and Treatment Outcomes. European Addiction Research, 27(3), 206–215. https://doi.org/10.1159/000512156

Andersson, H. W., Wenaas, M., & Nordfjærn, T. (2019). Relapse after inpatient substance use treatment: A prospective cohort study among users of illicit substances. Addictive Behaviors, 90, 222–228. https://doi.org/10.1016/j.addbeh.2018.11.008

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. https://doi.org/10.1038/nature15393

Bagot, K. S., Milin, R., & Kaminer, Y. (2015). Adolescent Initiation of Cannabis Use and Early-Onset Psychosis. Substance Abuse, 36(4), 524–533. https://doi.org/10.1080/08897077.2014.995332

Baselmans, B. M. L., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., Bao, Y., Smart, M., Kumari, M., Willemsen, G., Hottenga, J. J., Boomsma, D. I., de Geus, E. J. C., Nivard, M. G., & Bartels, M. (2019). Multivariate genome-wide analyses of the well-being spectrum. Nature Genetics, 51(3), 445–451. https://doi.org/10.1038/s41588-018-0320-8

Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. Clinical Psychology Review, 70(March), 51–63. https://doi.org/10.1016/j.cpr.2019.04.002

Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N., & Cormand, B. (2019). Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 94. https://doi.org/10.1016/J.PNPBP.2019.109667

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Christoffersen, L. A. N., Mortensen, E. L., Becker, U., Osler, M., Sørensen, H. J., & Flensborg-Madsen, T. (2021). Age at onset and age at treatment of alcohol use disorders: Associations with educational level and intelligence. Alcohol, 95, 7–14. https://doi.org/10.1016/j.alcohol.2021.04.005

Compton, W. M., Valentino, R. J., & DuPont, R. L. (2021). Polysubstance use in the U.S. opioid crisis. In Molecular Psychiatry (Vol. 26, Issue 1, pp. 41–50). https://doi.org/10.1038/s41380-020-00949-3

Dardani, C., Riglin, L., Leppert, B., Sanderson, E., Rai, D., Howe, L. D., Davey Smith, G., Tilling, K., Thapar, A., Davies, N. M., Anderson, E., & Stergiakouli, E. (2021). Is genetic liability to ADHD

and ASD causally linked to educational attainment? International Journal of Epidemiology, 50(6), 2011–2023. https://doi.org/10.1093/ije/dyab107

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. Nature Genetics, 48(10), 1284–1287. https://doi.org/10.1038/ng.3656

Deak, J. D., Zhou, H., Galimberti, M., Levey, D. F., Wendt, F. R., Sanchez-Roige, S., Hatoum, A. S., Johnson, E. C., Nunez, Y. Z., Demontis, D., Børglum, A. D., Rajagopal, V. M., Jennings, M. V, Kember, R. L., Justice, A. C., Edenberg, H. J., Agrawal, A., Polimanti, R., Kranzler, H. R., & Gelernter, J. (2022). Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci. Molecular Psychiatry, 27(10), 3970–3979. https://doi.org/10.1038/s41380-022-01709-1

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics, 51(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Diblasi, E., Kang, J., & Docherty, A. R. (2021). Genetic contributions to suicidal thoughts and behaviors. Psychological Medicine, 51(13), 2148–2155. https://doi.org/10.1017/S0033291721001720

Docherty, A. R., Mullins, N., Ashley-Koch, A. E., Qin, X. J., Coleman, J., Shabalin, A. A., Kang, J., Murnyak, B., McMahon, B. H., Oslin, D. W., Hauser, E. R., Hauser, M. A., Li, Q., Coon, H., Kimbrel, N. A., Beckham, J. C., & Ruderfer, D. (2022). Genome-wide association study meta-analysis of suicide attempt in 43,871 cases identifies twelve genome-wide significant loci. MedRxiv. https://doi.org/10.1101/2022.07.03.22277199

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-09718-5

Gelernter, J., & Polimanti, R. (2021). Genetics of substance use disorders in the era of big data. In Nature Reviews Genetics (Vol. 22, Issue 11, pp. 712–729). Nature Research. https://doi.org/10.1038/s41576-021-00377-1

Goldberg, J., Magruder, K. M., Forsberg, C. W., Kazis, L. E., Üstün, T. B., Friedman, M. J., Litz, B. T., Vaccarino, V., Heagerty, P. J., Gleason, T. C., Huang, G. D., & Smith, N. L. (2014). The association of PTSD with physical and mental health functioning and disability (VA Cooperative Study #569: The course and consequences of posttraumatic stress disorder in Vietnam-era Veteran twins). Quality of Life Research, 23(5), 1579–1591. https://doi.org/10.1007/s11136-013-0585-4

Hatoum, A. S., Colbert, S. M. C., Johnson, E. C., Huggett, S. B., Deak, J. D., Pathak, G., Jennings, M. V., Paul, S. E., Karcher, N. R., Hansen, I., Baranger, D. A. A., Edwards, A., Grotzinger, A., Consortium, S. U. D. W. G. of the P. G., Tucker-Drob, E. M., Kranzler, H. R., Davis, L. K., Sanchez-Roige, S., Polimanti, R., … Agrawal, A. (2022). Multivariate genome-wide association meta-

analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. MedRxiv, 1(March), 2022.01.06.22268753. https://doi.org/10.1038/s44220-023-00034-y

Hatoum, A. S., Johnson, E. C., Colbert, S. M. C., Polimanti, R., Zhou, H., Walters, R. K., Gelernter, J., Edenberg, H. J., Bogdan, R., & Agrawal, A. (2022). The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. Neuropsychopharmacology, 47(10), 1739–1745. https://doi.org/10.1038/s41386-021-01209-w

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nature Neuroscience, 22(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Johnson, E. C., Demontis, D., Thorgeirsson, T. E., Walters, R. K., Polimanti, R., Hatoum, A. S., Sanchez-Roige, S., Paul, S. E., Wendt, F. R., Clarke, T. K., Lai, D., Reginsson, G. W., Zhou, H., He, J., Baranger, D. A. A., Gudbjartsson, D. F., Wedow, R., Adkins, D. E., Adkins, A. E., … Børglum, A. D. (2020). A large-scale genome-wide association study meta-analysis of cannabis use disorder. The Lancet Psychiatry, 7(12), 1032–1045. https://doi.org/10.1016/S2215-0366(20)30339-4

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Wagner, G. G. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nature Genetics, 51(2), 245–257. https://doi.org/10.1038/s41588-018-0309-3

Kember, R. L., Hartwell, E. E., Xu, H., Rotenberg, J., Almasy, L., Zhou, H., Gelernter, J., & Kranzler, H. R. (2022). Phenome-wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped Sample. Biological Psychiatry. https://doi.org/10.1016/j.biopsych.2022.08.010

Kinreich, S., McCutcheon, V. V, Aliev, F., Meyers, J. L., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Kuang, W., Pandey, G., Viteri, S. S. S. de, Francis, M. W., Chan, G., Bourdon, J. L., Dick, D. M., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., … Porjesz, B. (2021). Predicting alcohol use disorder remission: a longitudinal multimodal multi-featured machine learning approach. Translational Psychiatry, 11(1), 166. https://doi.org/10.1038/s41398-021-01281-2

Kokkevi, A., & Hartgers, C. (1995). EuropASI: European Adaptation of a Multidimensional Assessment Instrument for Drug and Alcohol Dependence. European Addiction Research, 1(4), 208–210. https://doi.org/10.1159/000259089

Kranzler, H. R., Zhou, H., Kember, R. L., Vickers Smith, R., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C., Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-09480-8

Krawczyk, N., Feder, K. A., Saloner, B., Crum, R. M., Kealhofer, M., & Mojtabai, R. (2017). The association of psychiatric comorbidity with treatment completion among clients admitted

to substance use treatment programs in a U.S. national sample. Drug and Alcohol Dependence, 175, 157–163. https://doi.org/10.1016/j.drugalcdep.2017.02.006

Lähteenvuo, M., Batalla, A., Luykx, J. J., Mittendorfer-Rutz, E., Tanskanen, A., Tiihonen, J., & Taipale, H. (2021). Morbidity and mortality in schizophrenia with comorbid substance use disorders. Acta Psychiatrica Scandinavica, 144(1), 42–49. https://doi.org/10.1111/acps.13291

Lannoy, S., Mars, B., Heron, J., & Edwards, A. C. (2022). Suicidal ideation during adolescence: The roles of aggregate genetic liability for suicide attempts and negative life events in the past year. Journal of Child Psychology and Psychiatry and Allied Disciplines, 63(10), 1164–1173. https://doi.org/10.1111/jcpp.13653

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Turley, P. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics, 50(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lee, P. H., Doyle, A. E., Li, X., Silberstein, M., Jung, J.-Y., Gollub, R. L., Nierenberg, A. A., Liu, R. T., Kessler, R. C., Perlis, R. H., & Fava, M. (2022). Genetic Association of Attention-Deficit/Hyperactivity Disorder and Major Depression With Suicidal Ideation and Attempts in Children: The Adolescent Brain Cognitive Development Study. Biological Psychiatry, 92(3), 236–245. https://doi.org/10.1016/j.biopsych.2021.11.026

Levey, D. F., Polimanti, R., Cheng, Z., Zhou, H., Nuñez, Y. Z., Jain, S., He, F., Sun, X., Ursano, R. J., Kessler, R. C., Smoller, J. W., Stein, M. B., Kranzler, H. R., & Gelernter, J. (2019). Genetic associations with suicide attempt severity and genetic overlap with major depression. Translational Psychiatry, 9(1). https://doi.org/10.1038/s41398-018-0340-2

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity, 95(3), 221–227. https://doi.org/10.1038/sj.hdy.6800717

Lim, K. X., Rijsdijk, F., Hagenaars, S. P., Socrates, A., Choi, S. W., Coleman, J. R. I., Glanville, K. P., Lewis, C. M., & Pingault, J. B. (2020). Studying individual risk factors for self-harm in the UK Biobank: A polygenic scoring and Mendelian randomisation study. PLoS Medicine, 17(6). https://doi.org/10.1371/journal.pmed.1003137

Liu, H. (2019). Genetic architecture of socioeconomic outcomes: Educational attainment, occupational status, and wealth. Social Science Research, 82, 137–147. https://doi.org/10.1016/j.ssresearch.2019.04.008

Loughnan, R. J., Palmer, C. E., Makowski, C., Thompson, W. K., Barch, D. M., Jernigan, T. L., Dale, A. M., & Fan, C. C. (2022). Unique prediction of developmental psychopathology from genetic and familial risk. Journal of Child Psychology and Psychiatry and Allied Disciplines, 63(12), 1631–1643. https://doi.org/10.1111/jcpp.13649

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Melchior, M., Chollet, A., Elidemir, G., Galéra, C., & Younès, N. (2015). Unemployment and substance use in young adults: Does educational attainment modify the association? European Addiction Research, 21(3), 115–123. https://doi.org/10.1159/000365887

Meyers, J. L., Salvatore, J. E., Aliev, F., Johnson, E. C., McCutcheon, V. V., Su, J., Kuo, S. I. C., Lai, D., Wetherill, L., Wang, J. C., Chan, G., Hesselbrock, V., Foroud, T., Bucholz, K. K., Edenberg, H. J., Dick, D. M., Porjesz, B., & Agrawal, A. (2019). Psychosocial moderation of polygenic risk for cannabis involvement: the role of trauma exposure and frequency of religious service attendance. Translational Psychiatry, 9(1), 1–12. https://doi.org/10.1038/s41398-019-0598-z

Mills, K. L., Teesson, M., Ross, J., & Peters, L. (2006). Trauma, PTSD, and substance use disorders: Findings from the Australian National Survey of Mental Health and Well-Being. American Journal of Psychiatry, 163(4), 652–658. https://doi.org/10.1176/ajp.2006.163.4.652

Mitjans, M., Papiol, S., Barrot, C., Guardiola-Ripoll, M., Giménez-Palomo, A., Gavín, P., Oraki Kohshour, M., Acosta, M., Pomarol-Clotet, E., Castellà-García, J., Pérez-Pérez, R. M., Gutiérrez, B., Schulze, T. G., Fatjó-Vilas, M., Arias, B., & Benabarre, A. (2022). Completed suicide is associated with a higher polygenic burden for psychiatric disorders. European Archives of Psychiatry and Clinical Neuroscience, 272(3), 355–358. https://doi.org/10.1007/s00406-022-01398-5

Morley, K. I., Lynskey, M. T., Moran, P., Borschmann, R., & Winstock, A. R. (2015). Polysubstance use, mental health and high-risk behaviours: Results from the 2012 Global Drug Survey. Drug and Alcohol Review, 34(4), 427–437. https://doi.org/10.1111/dar.12263

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., Bryois, J., Charney, A. W., Drange, O. K., Gandal, M. J., Hagenaars, S. P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., … Andreassen, O. A. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nature Genetics, 53(6), 817–829. https://doi.org/10.1038/s41588-021-00857-4

Mullins, N., Perroud, N., Uher, R., Butler, A. W., Cohen-Woods, S., Rivera, M., Malki, K., Euesden, J., Power, R. A., Tansey, K. E., Jones, L., Jones, I., Craddock, N., Owen, M. J., Korszun, A., Gill, M., Mors, O., Preisig, M., Maier, W., … Lewis, C. M. (2014). Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: A genome-wide association and polygenic scoring study. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 165(5), 428–437. https://doi.org/10.1002/ajmg.b.32247

Nievergelt, C. M., Maihofer, A. X., Klengel, T., Atkinson, E. G., Chen, C. Y., Choi, K. W., Coleman, J. R. I., Dalvie, S., Duncan, L. E., Gelernter, J., Levey, D. F., Logue, M. W., Polimanti, R., Provost, A. C., Ratanatharathorn, A., Stein, M. B., Torres, K., Aiello, A. E., Almli, L. M., … Koenen, K. C. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-12576-w

Nyholt, D. R. (2004). A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. American Journal of Human Genetics, 74(4), 765–769. https://doi.org/10.1086/383251

O'Brien, B. S., & Sher, L. (2013). Child sexual abuse and the pathophysiology of suicide in adolescents and adults. In International Journal of Adolescent Medicine and Health (Vol. 25, Issue 3, pp. 201–205). Int J Adolesc Med Health. https://doi.org/10.1515/ijamh-2013-0053

Onaemo, V. N., Fawehinmi, T. O., & D'Arcy, C. (2022). Risk of suicide ideation in comorbid substance use disorder and major depression. PLoS ONE, 17(12). https://doi.org/10.1371/journal.pone.0265287

Ostacher, M. J. (2007). Comorbid Alcohol and Substance Abuse Dependence in Depression: Impact on the Outcome of Antidepressant Treatment. In Psychiatric Clinics of North America (Vol. 30, Issue 1, pp. 69–76). Elsevier. https://doi.org/10.1016/j.psc.2006.12.009

Østergaard, M. L. D., Nordentoft, M., & Hjorthøj, C. (2017). Associations between substance use disorders and suicide or suicide attempts in people with mental illness: a Danish nation-wide, prospective, register-based study of patients diagnosed with schizophrenia, bipolar disorder, unipolar depression or personal. Addiction, 112(7), 1250–1259. https://doi.org/10.1111/add.13788

Park, Y. M., Shekhtman, T., & Kelsoe, J. R. (2020). Interaction between adverse childhood experiences and polygenic risk in patients with bipolar disorder. Translational Psychiatry, 10(1). https://doi.org/10.1038/s41398-020-01010-1

Polimanti, R., Kaufman, J., Zhao, H., Kranzler, H. R., Ursano, R. J., Kessler, R. C., Stein, M. B., & Gelernter, J. (2018). Trauma exposure interacts with the genetic risk of bipolar disorder in alcohol misuse of US soldiers. Acta Psychiatrica Scandinavica, 137(2), 148–156. https://doi.org/10.1111/acps.12843

Poudel, A., & Gautam, S. (2017). Age of onset of substance use and psychosocial problems among individuals with substance use disorders. BMC Psychiatry, 17(1). https://doi.org/10.1186/s12888-016-1191-0

Prom-Wormley, E. C., Ebejer, J., Dick, D. M., & Bowers, M. S. (2017). The genetic epidemiology of substance use disorder: A review. In Drug and Alcohol Dependence (Vol. 180, pp. 241–259). NIH Public Access. https://doi.org/10.1016/j.drugalcdep.2017.06.040

Ramos-Quiroga, J. A., Bosch, R., Richarte, V., Valero, S., Gómez-Barros, N., Nogueira, M., Palomar, G., Corrales, M., Sáez-Francàs, N., Corominas, M., Real, A., Vidal, R., Chalita, P. J., & Casas, M. (2012). Criterion and concurrent validity of Conners Adult ADHD Diagnostic Interview for DSM-IV (CAADID) Spanish version. Revista de Psiquiatría y Salud Mental (English Edition), 5(4), 229–235. https://doi.org/10.1016/j.rpsmen.2012.05.002

Richmond-Rakerd, L. S., Slutske, W. S., & Wood, P. K. (2017). Age of initiation and substance use progression: A multivariate latent growth analysis. Psychology of Addictive Behaviors, 31(6), 664–675. https://doi.org/10.1037/adb0000304

Roehrs, T., Sibai, M., & Roth, T. (2021). Sleep and alertness disturbance and substance use disorders: A bi-directional relation. In Pharmacology Biochemistry and Behavior (Vol. 203). https://doi.org/10.1016/j.pbb.2021.173153

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, H. T., Abebe, M., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., Abraha, H. N., … Murray, C. J. L. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet, 392(10159), 1736–1788. https://doi.org/10.1016/S0140-6736(18)32203-7

Ruderfer, D. M., Walsh, C. G., Aguirre, M. W., Tanigawa, Y., Ribeiro, J. D., Franklin, J. C., & Rivas, M. A. (2020). Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. Molecular Psychiatry, 25(10), 2422–2430. https://doi.org/10.1038/s41380-018-0326-8

Salvatore, J. E., Aliev, F., Edwards, A. C., Evans, D. M., Macleod, J., Hickman, M., Lewis, G., Kendler, K. S., Loukola, A., Korhonen, T., Latvala, A., Rose, R. J., Kaprio, J., & Dick, D. M. (2014). Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. Genes, 5(2), 330–346. https://doi.org/10.3390/genes5020330

Salvatore, J. E., Barr, P. B., Stephenson, M., Aliev, F., Kuo, S. I. C., Su, J., Agrawal, A., Almasy, L., Bierut, L., Bucholz, K., Chan, G., Edenberg, H. J., Johnson, E. C., McCutcheon, V. V., Meyers, J. L., Schuckit, M., Tischfield, J., Wetherill, L., & Dick, D. M. (2020). Sibling comparisons elucidate the associations between educational attainment polygenic scores and alcohol, nicotine and cannabis. Addiction, 115(2), 337–346. https://doi.org/10.1111/add.14815

Schoeler, T., Baldwin, J., Allegrini, A., Barkhuizen, W., McQuillin, A., Pirastu, N., Kutalik, Z., & Pingault, J. B. (2022). Novel Biological Insights Into the Common Heritable Liability to Substance Involvement: A Multivariate Genome-wide Association Study. Biological Psychiatry, 93(6), 524–535. https://doi.org/10.1016/j.biopsych.2022.07.027

Sun, J., Kranzler, H. R., Gelernter, J., & Bi, J. (2020). A genome-wide association study of cocaine use disorder accounting for phenotypic heterogeneity and gene–environment interaction. Journal of Psychiatry and Neuroscience, 45(1), 34–44. https://doi.org/10.1503/jpn.180098

Syan, S. K., Minhas, M., Oshri, A., Costello, J., Sousa, S., Samokhvalov, A. V., Rush, B., & MacKillop, J. (2020). Predictors of premature treatment termination in a large residential addiction medicine program. Journal of Substance Abuse Treatment, 117. https://doi.org/10.1016/j.jsat.2020.108077

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., … van Os, J. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature, 604(7906), 502. https://doi.org/10.1038/S41586-022-04434-5

van de Weijer, M. P., Pelt, D. H. M., van Beijsterveldt, C. E. M., Willemsen, G., & Bartels, M. (2022). Genetic factors explain a significant part of associations between adolescent well-being and the social environment. European Child and Adolescent Psychiatry, 31(10), 1611–1622. https://doi.org/10.1007/s00787-021-01798-3

Vink, J. M. (2016). Genetics of addiction: Future focus on gene × environment interaction? Journal of Studies on Alcohol and Drugs, 77(5), 684–687. https://doi.org/10.15288/jsad.2016.77.684

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (Sf-36): I. conceptual framework and item selection. Medical Care, 30(6), 473–483. https://doi.org/10.1097/00005650-199206000-00002

Wertz, J., Caspi, A., Belsky, D. W., Beckley, A. L., Arseneault, L., Barnes, J. C., Corcoran, D. L., Hogan, S., Houts, R. M., Morgan, N., Odgers, C. L., Prinz, J. A., Sugden, K., Williams, B. S., Poulton, R., & Moffitt, T. E. (2018). Genetics and Crime: Integrating New Genomic Discoveries Into Psychological Research About Antisocial Behavior. Psychological Science, 29(5), 791–803. https://doi.org/10.1177/0956797617744542

Young, J. G. (1967). Meeting the social needs of medicare patients. Gerontologist, 7(4), 261–265. https://doi.org/10.1093/geront/7.4.261

Zheng, J., Richardson, T. G., Millard, L. A. C., Hemani, G., Elsworth, B. L., Raistrick, C. A., Vilhjalmsson, B., Neale, B. M., Haycock, P. C., Smith, G. D., & Gaunt, T. R. (2018). PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. GigaScience, 7(8), 1–10. https://doi.org/10.1093/gigascience/giy090

Zilberman, N., Yadid, G., Efrati, Y., Neumark, Y., & Rassovsky, Y. (2018). Personality profiles of substance and behavioral addictions. Addictive Behaviors, 82, 174–181. https://doi.org/10.1016/j.addbeh.2018.03.007

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A Comparison of Three Structural Models for Personality: The Big Three, the Big Five, and the Alternative Five. Journal of Personality and Social Psychology, 65(4), 757–768. https://doi.org/10.1037/0022-3514.65.4.757

SUPPLEMENTARY TABLES

Supplementary Table 1. Discovery GWASs of psychiatric diseases, behavioral and related traits used to construct the PGSs

Trait	n cases	n controls	n total [a]	ref
Attention-Deficit Hyperactivity Disorder	20,183	35,191	51,306	(Demontis, Walters, et al., 2019)
Anxiety	16,730	101,021	57,412	http://www.nealelab.is/uk-biobank/
Bipolar Disorder	41,917	371,549	150,670	(Mullins et al., 2021)
Depression	170,756	329,443	449,856	(Howard et al., 2019)
Post-Traumatic Stress Disorder	9,354	25,175	27,280	(Nievergelt et al., 2019)
Schizophrenia	67390	94015	157,013	(Trubetskoy, Pardiñas, Qi, Panagiotaropoulou, van Os, et al., 2022)
Risk tolerance	-	-	466,571	(Karlsson Linnér et al., 2019)
Suicide attempt	26,590	492,022	100,907	(Docherty et al., 2022)
Educational attainment	-	-	766,345	(J. J. Lee et al., 2018)
Well being	-	-	2,083,151	(Baselmans et al., 2019)

[a] For binary traits effective sample size was calculated with the formula 4/(1/n cases+1/n controls)

Supplementary Table 2a. Association between the PGS for ADHD and 39 clinical variables from the SUD phenome. In bold nominal significant results.

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,03 , -3,00E-04	**4.53E-02**
Age at onset of SUD [c]	1339	Linear	-0.01	-0,02 , 0,01	0.39
Years between substance use and SUD	1325	Linear	1.04	0,95 , 1,14	0.36
Years of substance use as proportion of lifespan	1148	Linear	1.07	0,03 , 2,1	**4.42E-02**
Number of substances consumed	869	Negative binomia	1.03	0,99 , 1,07	0.19
Number of therapeutic community interventions	1314	Negative binomia	1.01	0,92 , 1,11	0.80
Number of inpatient detoxifications	1327	Negative binomia	1.03	0,92 , 1,15	0.62
Number of outpatient treatments	1252	Negative binomia	1.05	0,99 , 1,11	0.09
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.12	0,92 , 1,36	0.28
Major depressive disorder	828	Logistic	1.10	0,96 , 1,27	0.17
Antisocial personality disorder	560	Logistic	1.24	1,04 , 1,47	**1.60E-02**
Psychotic disorder	593	Logistic	1.02	0,79 , 1,3	0.90
Anxiety disorder	654	Logistic	1.03	0,88 , 1,2	0.71
Attention deficit hyperactivity disorder	700	Logistic	1.24	1,06 , 1,45	**7.35E-03**
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	0.07	-0,29 , 0,43	0.72
Aggression Hostility personality factor	667	Linear	0.15	-0,08 , 0,38	0.21
Sociability personality factor	632	Linear	-0.18	-0,44 , 0,08	0.18
Impulsive sensation seeking personality factor	666	Linear	0.01	-0,3 , 0,32	0.94
Activity personality factor	665	Linear	0.02	-0,24 , 0,28	0.88
Suicide attempt	618	Logistic	1.03	0,89 , 1,19	0.67
Suicide ideation	731	Logistic	0.98	0,84 , 1,15	0.80
Psychotic symptoms	1281	Logistic	1.00	0,89 , 1,12	0.98
Sleeping disturbances	1228	Logistic	1.02	0,91 , 1,14	0.76
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.00	0,87 , 1,16	0.97
Employment status	984	Ordinal	1.08	0,97 , 1,2	0.18
Medical status	982	Ordinal	1.03	0,92 , 1,17	0.60
Psychiatric status	984	Ordinal	1.01	0,91 , 1,13	0.86
Drug use	984	Ordinal	1.06	0,95 , 1,19	0.28
Alcohol use	984	Ordinal	0.96	0,86 , 1,08	0.54
Family/Social relationships	981	Ordinal	1.10	0,99 , 1,23	0.09
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.52	-1,23 , 0,19	0.15
Mental health	751	Linear	0.49	-0,45 , 1,44	0.31
Criminal record	713	Logistic	1.13	0,98 , 1,32	0.10
Unemployment	1057	Logistic	1.03	0,91 , 1,17	0.61
Number of psychiatric hospitalizations	760	Negative binomia	1.05	0,86 , 1,29	0.61
Psychiatric family history	840	Logistic	0.97	0,85 , 1,11	0.66
Lifetime medical conditions	1334	Logistic	0.99	0,88 , 1,11	0.84
Substance use family history	818	Logistic	1.10	0,96 , 1,27	0.16
Educational attainment	1333	Ordinal	0.85	0,93 , 0,77	**1.20E-03**

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2b. Association between the PGS for anxiety and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,02 , 0,01	0.36
Age at onset of SUD [c]	1339	Linear	0.00	-0,02 , 0,01	0.77
Years between substance use and SUD	1325	Linear	1.04	0,95 , 1,13	0.45
Years of substance use as proportion of lifespan	1148	Linear	-0.03	-1,07 , 1,02	0.96
Number of substances consumed	869	Negative binomial	1.01	0,97 , 1,05	0.79
Number of therapeutic community interventions	1314	Negative binomial	1.10	1 , 1,21	0.05
Number of inpatient detoxifications	1327	Negative binomial	1.01	0,9 , 1,13	0.84
Number of outpatient treatments	1252	Negative binomial	1.09	1,03 , 1,15	**2.67E-03**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.02	0,84 , 1,24	0.86
Major depressive disorder	828	Logistic	1.03	0,9 , 1,19	0.63
Antisocial personality disorder	560	Logistic	1.09	0,92 , 1,29	0.30
Psychotic disorder	593	Logistic	1.49	1,16 , 1,92	**1.93E-03**
Anxiety disorder	654	Logistic	1.04	0,89 , 1,21	0.63
Attention deficit hyperactivity disorder	700	Logistic	0.86	0,74 , 1,01	0.06
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	0.23	-0,12 , 0,59	0.20
Aggression Hostility personality factor	667	Linear	0.07	-0,16 , 0,3	0.56
Sociability personality factor	632	Linear	-0.08	-0,34 , 0,17	0.53
Impulsive sensation seeking personality factor	666	Linear	0.02	-0,29 , 0,33	0.91
Activity personality factor	665	Linear	0.04	-0,22 , 0,29	0.79
Suicide attempt	618	Logistic	1.07	0,93 , 1,23	0.37
Suicide ideation	731	Logistic	1.01	0,87 , 1,18	0.88
Psychotic symptoms	1281	Logistic	1.06	0,95 , 1,19	0.28
Sleeping disturbances	1228	Logistic	0.98	0,88 , 1,1	0.77
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	0.92	0,8 , 1,07	0.28
Employment status	984	Ordinal	0.97	0,87 , 1,08	0.57
Medical status	982	Ordinal	1.08	0,96 , 1,22	0.23
Psychiatric status	984	Ordinal	1.08	0,97 , 1,2	0.16
Drug use	984	Ordinal	1.08	0,97 , 1,21	0.17
Alcohol use	984	Ordinal	0.95	0,85 , 1,06	0.36
Family/Social relationships	981	Ordinal	1.07	0,96 , 1,19	0.26
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.93	-1,64 , -0,22	**1.07E-02**
Mental health	751	Linear	0.06	-0,89 , 1,01	0.90
Criminal record	713	Logistic	0.94	0,81 , 1,1	0.47
Unemployment	1057	Logistic	1.05	0,93 , 1,19	0.44
Number of psychiatric hospitalizations	760	Negative binomial	1.03	0,84 , 1,27	0.77
Psychiatric family history	840	Logistic	1.16	1,01 , 1,33	**4.16E-02**
Lifetime medical conditions	1334	Logistic	1.06	0,95 , 1,19	0.31
Substance use family history	818	Logistic	1.02	0,88 , 1,17	0.80
Educational attainment	1333	Ordinal	0.92	1,02 , 0,83	0.10

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2c. Association between the PGS for bipolar disorder and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	0.01	-0,01 , 0,02	0.33
Age at onset of SUD [c]	1339	Linear	0.00	-0,02 , 0,02	0.90
Years between substance use and SUD	1325	Linear	1.01	0,93 , 1,11	0.78
Years of substance use as proportion of lifespan	1148	Linear	0.20	-0,85 , 1,24	0.71
Number of substances consumed	869	Negative binomial	1.02	0,98 , 1,07	0.26
Number of therapeutic community interventions	1314	Negative binomial	0.92	0,84 , 1,01	0.09
Number of inpatient detoxifications	1327	Negative binomial	1.03	0,92 , 1,16	0.59
Number of outpatient treatments	1252	Negative binomial	0.99	0,94 , 1,05	0.85
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.03	0,85 , 1,26	0.74
Major depressive disorder	828	Logistic	0.94	0,81 , 1,08	0.37
Antisocial personality disorder	560	Logistic	1.02	0,86 , 1,21	0.85
Psychotic disorder	593	Logistic	1.08	0,84 , 1,4	0.53
Anxiety disorder	654	Logistic	0.98	0,83 , 1,15	0.78
Attention deficit hyperactivity disorder	700	Logistic	1.04	0,89 , 1,22	0.59
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	-0.07	-0,45 , 0,3	0.71
Aggression Hostility personality factor	667	Linear	0.05	-0,18 , 0,29	0.65
Sociability personality factor	632	Linear	0.16	-0,11 , 0,43	0.24
Impulsive sensation seeking personality factor	666	Linear	0.05	-0,28 , 0,38	0.76
Activity personality factor	665	Linear	-0.09	-0,36 , 0,18	0.53
Suicide attempt	618	Logistic	1.04	0,9 , 1,2	0.62
Suicide ideation	731	Logistic	1.05	0,9 , 1,23	0.54
Psychotic symptoms	1281	Logistic	1.14	1,02 , 1,28	**2.58E-02**
Sleeping disturbances	1228	Logistic	1.00	0,9 , 1,12	0.95
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.09	0,95 , 1,26	0.22
Employment status	984	Ordinal	0.98	0,88 , 1,09	0.73
Medical status	982	Ordinal	1.08	0,95 , 1,22	0.23
Psychiatric status	984	Ordinal	1.00	0,9 , 1,11	0.94
Drug use	984	Ordinal	0.93	0,83 , 1,04	0.19
Alcohol use	984	Ordinal	1.04	0,93 , 1,16	0.48
Family/Social relationships	981	Ordinal	1.01	0,9 , 1,12	0.92
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.62	-1,36 , 0,12	0.10
Mental health	751	Linear	0.32	-0,67 , 1,31	0.53
Criminal record	713	Logistic	1.08	0,93 , 1,25	0.32
Unemployment	1057	Logistic	1.14	1,01 , 1,29	**3.72E-02**
Number of psychiatric hospitalizations	760	Negative binomial	1.26	1,03 , 1,53	**2.23E-02**
Psychiatric family history	840	Logistic	1.12	0,98 , 1,28	0.11
Lifetime medical conditions	1334	Logistic	0.99	0,89 , 1,11	0.87
Substance use family history	818	Logistic	1.15	1 , 1,32	**4.89E-02**
Educational attainment	1333	Ordinal	0.97	1,06 , 0,88	0.51

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2d. Association between the PGS for depression and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	0.01	-0,03 , 2,00E-03	0.09
Age at onset of SUD [c]	1339	Linear	0.00	-0,02 , 0,01	0.54
Years between substance use and SUD	1325	Linear	1.01	1,01 , 1,21	**3.49E-02**
Years of substance use as proportion of lifespan	1148	Linear	0.20	-1,24 , 0,86	0.72
Number of substances consumed	869	Negative binomial	1.02	0,95 , 1,04	0.80
Number of therapeutic community interventions	1314	Negative binomial	0.92	0,9 , 1,08	0.76
Number of inpatient detoxifications	1327	Negative binomial	1.03	0,91 , 1,15	0.69
Number of outpatient treatments	1252	Negative binomial	0.99	1,03 , 1,15	**3.04E-03**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.03	0,93 , 1,38	0.22
Major depressive disorder	828	Logistic	0.94	0,88 , 1,17	0.87
Antisocial personality disorder	560	Logistic	1.02	0,83 , 1,17	0.82
Psychotic disorder	593	Logistic	1.08	0,85 , 1,41	0.50
Anxiety disorder	654	Logistic	0.98	0,88 , 1,2	0.73
Attention deficit hyperactivity disorder	700	Logistic	1.04	0,92 , 1,25	0.41
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	-0.07	4,00E-03 , 0,72	**4.91E-02**
Aggression Hostility personality factor	667	Linear	0.05	0,04 , 0,5	**2.18E-02**
Sociability personality factor	632	Linear	0.16	-0,41 , 0,11	0.25
Impulsive sensation seeking personality factor	666	Linear	0.05	-0,22 , 0,41	0.56
Activity personality factor	665	Linear	-0.09	-0,29 , 0,23	0.83
Suicide attempt	618	Logistic	1.04	1,02 , 1,37	**3.04E-02**
Suicide ideation	731	Logistic	1.05	0,95 , 1,3	0.20
Psychotic symptoms	1281	Logistic	1.14	0,98 , 1,23	0.12
Sleeping disturbances	1228	Logistic	1.00	0,93 , 1,17	0.48
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.09	0,87 , 1,16	0.92
Employment status	984	Ordinal	0.98	0,96 , 1,19	0.25
Medical status	982	Ordinal	1.08	0,99 , 1,27	0.06
Psychiatric status	984	Ordinal	1.00	1 , 1,23	0.06
Drug use	984	Ordinal	0.93	0,87 , 1,09	0.68
Alcohol use	984	Ordinal	1.04	0,88 , 1,11	0.86
Family/Social relationships	981	Ordinal	1.01	0,99 , 1,23	0.08
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.62	-1,35 , 0,09	0.09
Mental health	751	Linear	0.32	-1,09 , 0,84	0.80
Criminal record	713	Logistic	1.08	1,09 , 1,48	**1.82E-03**
Unemployment	1057	Logistic	1.14	0,92 , 1,17	0.58
Number of psychiatric hospitalizations	760	Negative binomial	1.26	0,95 , 1,4	0.15
Psychiatric family history	840	Logistic	1.12	1,01 , 1,32	**3.70E-02**
Lifetime medical conditions	1334	Logistic	0.99	0,91 , 1,15	0.70
Substance use family history	818	Logistic	1.15	0,99 , 1,31	0.06
Educational attainment	1333	Ordinal	0.97	1,09 , 0,88	0.69

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2e. Association between the PGS for post-traumatic stress disorder and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,03 , -1,00E-03	**3.72E-02**
Age at onset of SUD [c]	1339	Linear	-0.01	-0,03 , 0,01	0.23
Years between substance use and SUD	1325	Linear	0.98	0,9 , 1,08	0.70
Years of substance use as proportion of lifespan	1148	Linear	0.16	-0,88 , 1,21	0.76
Number of substances consumed	869	Negative binomial	0.99	0,96 , 1,04	0.79
Number of therapeutic community interventions	1314	Negative binomial	1.01	0,92 , 1,11	0.89
Number of inpatient detoxifications	1327	Negative binomial	1.13	1,01 , 1,27	**4.15E-02**
Number of outpatient treatments	1252	Negative binomial	1.05	1 , 1,11	0.08
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.04	0,85 , 1,27	0.69
Major depressive disorder	828	Logistic	0.92	0,79 , 1,06	0.23
Antisocial personality disorder	560	Logistic	1.11	0,93 , 1,31	0.24
Psychotic disorder	593	Logistic	1.00	0,78 , 1,29	0.97
Anxiety disorder	654	Logistic	0.96	0,82 , 1,13	0.63
Attention deficit hyperactivity disorder	700	Logistic	1.15	0,99 , 1,35	0.07
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	0.15	-0,22 , 0,52	0.43
Aggression Hostility personality factor	667	Linear	0.19	-0,05 , 0,43	0.12
Sociability personality factor	632	Linear	-0.16	-0,43 , 0,11	0.24
Impulsive sensation seeking personality factor	666	Linear	-0.09	-0,41 , 0,23	0.59
Activity personality factor	665	Linear	0.05	-0,22 , 0,32	0.73
Suicide attempt	618	Logistic	1.04	0,88 , 1,22	0.68
Suicide ideation	731	Logistic	1.00	0,86 , 1,16	0.99
Psychotic symptoms	1281	Logistic	1.09	0,97 , 1,22	0.15
Sleeping disturbances	1228	Logistic	1.08	0,97 , 1,22	0.17
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.00	0,86 , 1,15	0.96
Employment status	984	Ordinal	1.00	0,9 , 1,12	0.99
Medical status	982	Ordinal	1.13	0,99 , 1,28	0.07
Psychiatric status	984	Ordinal	1.00	0,9 , 1,12	0.99
Drug use	984	Ordinal	0.96	0,86 , 1,08	0.50
Alcohol use	984	Ordinal	1.00	0,9 , 1,13	0.94
Family/Social relationships	981	Ordinal	1.06	0,95 , 1,19	0.28
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	0.32	-0,65 , 1,29	0.52
Mental health	751	Linear	-0.21	-0,94 , 0,52	0.57
Criminal record	713	Logistic	1.12	0,96 , 1,3	0.15
Unemployment	1057	Logistic	1.23	1,09 , 1,4	**1.00E-03**
Number of psychiatric hospitalizations	760	Negative binomial	0.86	0,7 , 1,06	0.15
Psychiatric family history	840	Logistic	1.15	1 , 1,32	0.05
Lifetime medical conditions	1334	Logistic	1.09	0,97 , 1,22	0.13
Substance use family history	818	Logistic	1.12	0,97 , 1,29	0.11
Educational attainment	1333	Ordinal	0.87	0,97 , 0,79	**9.48E-03**

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2f. Association between the PGS for schizophrenia and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	0.01	3,00E-04 , 0,03	**4.61E-02**
Age at onset of SUD [c]	1339	Linear	0.01	-0,01 , 0,02	0.44
Years between substance use and SUD	1325	Linear	0.99	0,9 , 1,08	0.80
Years of substance use as proportion of lifespan	1148	Linear	-0.11	-1,15 , 0,94	0.84
Number of substances consumed	869	Negative binomial	1.01	0,97 , 1,06	0.58
Number of therapeutic community interventions	1314	Negative binomial	1.04	0,95 , 1,14	0.43
Number of inpatient detoxifications	1327	Negative binomial	1.06	0,95 , 1,19	0.28
Number of outpatient treatments	1252	Negative binomial	0.98	0,93 , 1,04	0.49
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.00	0,82 , 1,22	0.99
Major depressive disorder	828	Logistic	0.93	0,8 , 1,07	0.31
Antisocial personality disorder	560	Logistic	1.04	0,87 , 1,23	0.67
Psychotic disorder	593	Logistic	1.37	1,05 , 1,78	**1.97E-02**
Anxiety disorder	654	Logistic	0.88	0,75 , 1,03	0.12
Attention deficit hyperactivity disorder	700	Logistic	0.98	0,84 , 1,15	0.84
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	0.07	-0,3 , 0,44	0.71
Aggression Hostility personality factor	667	Linear	0.18	-0,06 , 0,42	0.14
Sociability personality factor	632	Linear	0.01	-0,26 , 0,28	0.94
Impulsive sensation seeking personality factor	666	Linear	-0.17	-0,49 , 0,16	0.31
Activity personality factor	665	Linear	-0.01	-0,27 , 0,26	0.97
Suicide attempt	618	Logistic	0.90	0,78 , 1,04	0.17
Suicide ideation	731	Logistic	0.89	0,76 , 1,05	0.17
Psychotic symptoms	1281	Logistic	1.15	1,03 , 1,29	**1.40E-02**
Sleeping disturbances	1228	Logistic	0.93	0,83 , 1,04	0.21
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.06	0,92 , 1,23	0.42
Employment status	984	Ordinal	1.10	0,99 , 1,23	0.08
Medical status	982	Ordinal	1.02	0,9 , 1,16	0.71
Psychiatric status	984	Ordinal	1.03	0,92 , 1,15	0.60
Drug use	984	Ordinal	0.96	0,86 , 1,08	0.48
Alcohol use	984	Ordinal	1.03	0,92 , 1,16	0.61
Family/Social relationships	981	Ordinal	1.01	0,91 , 1,13	0.82
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.53	-1,26 , 0,21	0.16
Mental health	751	Linear	-0.10	-1,08 , 0,89	0.85
Criminal record	713	Logistic	1.07	0,92 , 1,24	0.41
Unemployment	1057	Logistic	1.13	1 , 1,28	**4.71E-02**
Number of psychiatric hospitalizations	760	Negative binomial	1.10	0,9 , 1,34	0.35
Psychiatric family history	840	Logistic	1.01	0,88 , 1,15	0.93
Lifetime medical conditions	1334	Logistic	0.99	0,88 , 1,11	0.86
Substance use family history	818	Logistic	1.03	0,9 , 1,19	0.64
Educational attainment	1333	Ordinal	1.01	1,11 , 0,91	0.87

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2g. Association between the PGS for risk tolerance and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,02 , 3,00E-03	0.12
Age at onset of SUD [c]	1339	Linear	-0.01	-0,03 , 4,00E-03	0.14
Years between substance use and SUD	1325	Linear	0.99	0,9 , 1,08	0.77
Years of substance use as proportion of lifespan	1148	Linear	0.42	-0,63 , 1,47	0.43
Number of substances consumed	869	Negative binomial	1.02	0,98 , 1,06	0.31
Number of therapeutic community interventions	1314	Negative binomial	1.02	0,93 , 1,12	0.69
Number of inpatient detoxifications	1327	Negative binomial	1.01	0,9 , 1,13	0.88
Number of outpatient treatments	1252	Negative binomial	1.06	1 , 1,12	**4.33E-02**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	0.97	0,8 , 1,18	0.76
Major depressive disorder	828	Logistic	0.89	0,77 , 1,02	0.09
Antisocial personality disorder	560	Logistic	0.99	0,83 , 1,17	0.87
Psychotic disorder	593	Logistic	1.08	0,84 , 1,38	0.55
Anxiety disorder	654	Logistic	1.03	0,89 , 1,21	0.69
Attention deficit hyperactivity disorder	700	Logistic	1.08	0,93 , 1,26	0.30
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	-0.51	-0,88 , -0,14	**6.90E-03**
Aggression Hostility personality factor	667	Linear	0.17	-0,07 , 0,41	0.16
Sociability personality factor	632	Linear	0.09	-0,18 , 0,36	0.51
Impulsive sensation seeking personality factor	666	Linear	0.32	0 , 0,65	**4.98E-02**
Activity personality factor	665	Linear	0.10	-0,17 , 0,36	0.48
Suicide attempt	618	Logistic	1.11	0,95 , 1,29	0.18
Suicide ideation	731	Logistic	1.04	0,88 , 1,23	0.62
Psychotic symptoms	1281	Logistic	1.00	0,89 , 1,12	0.98
Sleeping disturbances	1228	Logistic	1.09	0,97 , 1,22	0.16
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.18	1,02 , 1,36	**2.29E-02**
Employment status	984	Ordinal	0.95	0,85 , 1,06	0.35
Medical status	982	Ordinal	1.03	0,91 , 1,16	0.64
Psychiatric status	984	Ordinal	0.97	0,87 , 1,08	0.59
Drug use	984	Ordinal	1.01	0,91 , 1,13	0.84
Alcohol use	984	Ordinal	1.03	0,92 , 1,15	0.58
Family/Social relationships	981	Ordinal	1.01	0,9 , 1,12	0.92
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	0.08	-0,65 , 0,81	0.83
Mental health	751	Linear	0.30	-0,67 , 1,27	0.55
Criminal record	713	Logistic	1.16	1 , 1,36	0.06
Unemployment	1057	Logistic	0.99	0,87 , 1,11	0.83
Number of psychiatric hospitalizations	760	Negative binomial	0.87	0,71 , 1,07	0.18
Psychiatric family history	840	Logistic	1.01	0,88 , 1,17	0.84
Lifetime medical conditions	1334	Logistic	1.05	0,94 , 1,18	0.37
Substance use family history	818	Logistic	1.11	0,96 , 1,28	0.15
Educational attainment	1333	Ordinal	1.04	1,15 , 0,94	0.42

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2h. Association between the PGS for suicide attempt and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,03 , 2,00E-03	0.11
Age at onset of SUD [c]	1339	Linear	-0.02	-0,04 , -0,01	**1.12E-02**
Years between substance use and SUD	1325	Linear	0.92	0,84 , 1,01	0.08
Years of substance use as proportion of lifespan	1148	Linear	0.03	-1,04 , 1,09	0.96
Number of substances consumed	869	Negative binomial	1.00	0,96 , 1,05	0.85
Number of therapeutic community interventions	1314	Negative binomial	0.96	0,87 , 1,06	0.40
Number of inpatient detoxifications	1327	Negative binomial	1.02	0,9 , 1,14	0.79
Number of outpatient treatments	1252	Negative binomial	1.06	1 , 1,12	**4.10E-02**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.10	0,9 , 1,35	0.34
Major depressive disorder	828	Logistic	1.08	0,93 , 1,25	0.33
Antisocial personality disorder	560	Logistic	1.18	0,99 , 1,4	0.07
Psychotic disorder	593	Logistic	1.09	0,85 , 1,42	0.50
Anxiety disorder	654	Logistic	1.03	0,88 , 1,21	0.69
Attention deficit hyperactivity disorder	700	Logistic	1.09	0,93 , 1,27	0.30
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	0.09	-0,29 , 0,48	0.63
Aggression Hostility personality factor	667	Linear	0.26	0,02 , 0,51	**3.60E-02**
Sociability personality factor	632	Linear	0.04	-0,24 , 0,31	0.80
Impulsive sensation seeking personality factor	666	Linear	0.00	-0,33 , 0,34	0.99
Activity personality factor	665	Linear	0.24	-0,03 , 0,52	0.09
Suicide attempt	618	Logistic	1.01	0,87 , 1,18	0.85
Suicide ideation	731	Logistic	1.16	0,98 , 1,37	0.08
Psychotic symptoms	1281	Logistic	1.15	1,03 , 1,29	**1.69E-02**
Sleeping disturbances	1228	Logistic	0.97	0,87 , 1,09	0.65
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	1.18	1,01 , 1,37	**3.27E-02**
Employment status	984	Ordinal	1.03	0,92 , 1,16	0.56
Medical status	982	Ordinal	1.14	1 , 1,29	**4.93E-02**
Psychiatric status	984	Ordinal	1.13	1,01 , 1,27	**3.88E-02**
Drug use	984	Ordinal	1.05	0,93 , 1,18	0.44
Alcohol use	984	Ordinal	1.02	0,91 , 1,15	0.75
Family/Social relationships	981	Ordinal	1.13	1 , 1,26	0.05
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	-0.47	-1,21 , 0,27	0.22
Mental health	751	Linear	-0.19	-1,18 , 0,8	0.71
Criminal record	713	Logistic	1.14	0,98 , 1,32	0.11
Unemployment	1057	Logistic	1.05	0,93 , 1,19	0.40
Number of psychiatric hospitalizations	760	Negative binomial	1.08	0,88 , 1,32	0.47
Psychiatric family history	840	Logistic	0.94	0,82 , 1,08	0.40
Lifetime medical conditions	1334	Logistic	1.13	1,01 , 1,27	**3.84E-02**
Substance use family history	818	Logistic	1.11	0,96 , 1,28	0.15
Educational attainment	1333	Ordinal	0.90	1 , 0,81	**4.30E-02**

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2i. Association between the PGS for educational attainment and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	0.02	0,01 , 0,03	**8.11E-03**
Age at onset of SUD [c]	1339	Linear	0.02	4,00E-03 , 0,04	**1.48E-02**
Years between substance use and SUD	1325	Linear	1.07	0,98 , 1,17	0.15
Years of substance use as proportion of lifespan	1148	Linear	-0.73	-1,77 , 0,31	0.17
Number of substances consumed	869	Negative binomia	1.00	0,96 , 1,04	0.89
Number of therapeutic community interventions	1314	Negative binomia	0.89	0,81 , 0,98	**2.28E-02**
Number of inpatient detoxifications	1327	Negative binomia	0.89	0,8 , 1	0.06
Number of outpatient treatments	1252	Negative binomia	0.91	0,87 , 0,97	**1.64E-03**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	0.96	0,79 , 1,18	0.70
Major depressive disorder	828	Logistic	0.91	0,79 , 1,06	0.24
Antisocial personality disorder	560	Logistic	0.90	0,75 , 1,08	0.24
Psychotic disorder	593	Logistic	1.04	0,8 , 1,37	0.76
Anxiety disorder	654	Logistic	0.97	0,83 , 1,14	0.71
Attention deficit hyperactivity disorder	700	Logistic	0.89	0,76 , 1,05	0.16
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	-0.39	-0,78 , -0,01	**4.50E-02**
Aggression Hostility personality factor	667	Linear	-0.18	-0,42 , 0,07	0.16
Sociability personality factor	632	Linear	0.10	-0,18 , 0,38	0.47
Impulsive sensation seeking personality factor	666	Linear	-0.05	-0,39 , 0,28	0.75
Activity personality factor	665	Linear	0.17	-0,11 , 0,45	0.24
Suicide attempt	618	Logistic	1.03	0,89 , 1,2	0.66
Suicide ideation	731	Logistic	1.00	0,85 , 1,17	0.96
Psychotic symptoms	1281	Logistic	0.91	0,82 , 1,02	0.12
Sleeping disturbances	1228	Logistic	0.91	0,81 , 1,02	0.09
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	0.96	0,82 , 1,11	0.55
Employment status	984	Ordinal	0.90	0,81 , 1,01	0.07
Medical status	982	Ordinal	1.00	0,88 , 1,13	0.99
Psychiatric status	984	Ordinal	0.91	0,81 , 1,01	0.08
Drug use	984	Ordinal	0.95	0,84 , 1,06	0.36
Alcohol use	984	Ordinal	0.98	0,87 , 1,1	0.75
Family/Social relationships	981	Ordinal	0.86	0,77 , 0,96	**1.00E-02**
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	0.53	-0,24 , 1,3	0.18
Mental health	751	Linear	0.59	-0,43 , 1,62	0.26
Criminal record	713	Logistic	0.67	0,57 , 0,78	**8.03E-07**
Unemployment	1057	Logistic	0.82	0,72 , 0,92	**1.34E-03**
Number of psychiatric hospitalizations	760	Negative binomia	1.15	0,94 , 1,41	0.17
Psychiatric family history	840	Logistic	1.08	0,94 , 1,23	0.28
Lifetime medical conditions	1334	Logistic	0.99	0,88 , 1,11	0.88
Substance use family history	818	Logistic	0.80	0,7 , 0,92	**2.20E-03**
Educational attainment	1333	Ordinal	1.39	1,54 , 1,25	**3.34E-10**

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 2j. Association between the PGS for well-being and 39 clinical variables from the SUD phenome. In bold nominal significant results

Phenotypes	*n* [a]	Regression	Estimate [b]	95% CI	*p*
SUD variables					
Age at onset of substance use [c]	1347	Linear	-0.01	-0,02 , 3,00E-03	0.13
Age at onset of SUD [c]	1339	Linear	-0.01	-0,03 , 0,01	0.25
Years between substance use and SUD	1325	Linear	0.98	0,9 , 1,07	0.69
Years of substance use as proportion of lifespan	1148	Linear	0.65	-0,4 , 1,7	0.23
Number of substances consumed	869	Negative binomial	0.99	0,95 , 1,04	0.81
Number of therapeutic community interventions	1314	Negative binomial	0.99	0,9 , 1,08	0.77
Number of inpatient detoxifications	1327	Negative binomial	0.99	0,88 , 1,11	0.87
Number of outpatient treatments	1252	Negative binomial	0.91	0,86 , 0,96	**7.00E-04**
Comorbidity and personality traits					
Mental disorders in DSM-IV					
Borderline personality disorder	447	Logistic	1.07	0,87 , 1,31	0.52
Major depressive disorder	828	Logistic	0.99	0,86 , 1,15	0.90
Antisocial personality disorder	560	Logistic	1.06	0,9 , 1,26	0.49
Psychotic disorder	593	Logistic	0.85	0,65 , 1,11	0.23
Anxiety disorder	654	Logistic	0.90	0,76 , 1,06	0.20
Attention deficit hyperactivity disorder	700	Logistic	0.92	0,79 , 1,07	0.28
Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)					
Neuroticism Anxiety personality factor	663	Linear	-0.34	-0,72 , 0,04	0.08
Aggression Hostility personality factor	667	Linear	0.05	-0,2 , 0,29	0.71
Sociability personality factor	632	Linear	0.27	-0,01 , 0,54	0.06
Impulsive sensation seeking personality factor	666	Linear	0.25	-0,08 , 0,58	0.14
Activity personality factor	665	Linear	0.11	-0,16 , 0,38	0.43
Suicide attempt	618	Logistic	0.95	0,82 , 1,1	0.50
Suicide ideation	731	Logistic	0.93	0,79 , 1,09	0.36
Psychotic symptoms	1281	Logistic	0.93	0,83 , 1,04	0.20
Sleeping disturbances	1228	Logistic	0.91	0,82 , 1,02	0.12
Sociodemographic and health phenotypes					
EuropASI					
Legal status	984	Ordinal	0.95	0,82 , 1,1	0.51
Employment status	984	Ordinal	0.95	0,85 , 1,06	0.37
Medical status	982	Ordinal	0.95	0,83 , 1,07	0.40
Psychiatric status	984	Ordinal	0.87	0,78 , 0,97	**1.43E-02**
Drug use	984	Ordinal	0.98	0,87 , 1,09	0.67
Alcohol use	984	Ordinal	1.01	0,9 , 1,14	0.83
Family/Social relationships	981	Ordinal	0.83	0,74 , 0,93	**1.00E-03**
36-Item Short Form Survey (SF-36)					
Physical health	751	Linear	0.64	-0,1 , 1,37	0.09
Mental health	751	Linear	0.51	-0,47 , 1,49	0.31
Criminal record	713	Logistic	0.90	0,78 , 1,05	0.17
Unemployment	1057	Logistic	0.88	0,77 , 0,99	**3.59E-02**
Number of psychiatric hospitalizations	760	Negative binomial	0.78	0,64 , 0,95	**1.33E-02**
Psychiatric family history	840	Logistic	0.90	0,79 , 1,03	0.13
Lifetime medical conditions	1334	Logistic	0.98	0,87 , 1,1	0.70
Substance use family history	818	Logistic	0.83	0,73 , 0,96	**1.13E-02**
Educational attainment	1333	Ordinal	1.06	1,18 , 0,96	0.23

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2)

[b] OR is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; IRR is reported for negative binomial regression

[c] Logarithmic transformations were applied to continuous variables not following a normal distribution

Supplementary Table 3. Interaction between PGSs and lifetime emotional. physical and/or sexual abuse in the SUD phenome

PGS	SUD-Phenome	*n* [a]	Estimate	95% CI	*p*
Attention-deficit hyperactivity disorder	Age at onset of substance use [b]	701	-.002	-.04 , .04	.94
	Years of substance use as proportion of lifespan	631	.32	-2.57 , 3.2	.83
	Antisocial personality disorder	358	1.00	.66 , 1.51	.99
	Attention deficit hyperactivity disorder	434	.94	.63 , 1.4	.77
	Educational attainment	701	.96	.73 , 1.26	.77
Anxiety	Number of outpatient treatments	681	1.06	.92 , 1.22	.42
	Psychotic disorder	180	1.10	.61 , 1.99	.76
	SF36 Physical health	509	.43	-1.27 , 2.12	.62
	Psychiatric family history	524	1.03	.73 , 1.46	.86
Bipolar disorder	Psychotic symptoms	678	1.01	.74 , 1.39	.93
	Unemployment	568	1.08	.78 , 1.51	.64
	Number of psychiatric hospitalizations	504	.87	.53 , 1.42	.57
	Substance use family history	516	.98	.69 , 1.39	.91
Depression	Years between substance use and SUD	691	.96	.74 , 1.25	.78
	Number of outpatient treatments	681	1.07	.92 , 1.23	.39
	ZKPQ-Neuroticism Anxiety personality factor	492	.20	-.64 , 1.05	.64
	ZKPQ-Aggression Hostility personality factor	495	-.28	-.83 , .26	.31
	Suicide attempt	725	.93	.69 , 1.26	.65
	Criminal record	488	.76	.53 , 1.1	.14
	Psychiatric family history	524	1.31	.92 , 1.86	.13
Post-traumatic stress disorder	Age at onset of substance use [b]	701	-.02	-.06 , .03	.45
	Number of inpatient detoxifications	702	.91	.65 , 1.26	.56
	Unemployment	568	.85	.6 , 1.21	.38
	Educational attainment	701	1.08	.81 , 1.44	.58
Schizophrenia	Age at onset of substance use [b]	701	.03	-.01 , .08	.11
	Psychotic disorder	180	.85	.45 , 1.59	.60
	Psychotic symptoms	678	.97	.7 , 1.33	.83
	Unemployment	568	.76	.54 , 1.07	.12
Risk tolerance	Number of outpatient treatments	681	1.12	.96 , 1.31	.14
	ZKPQ-Neuroticism Anxiety personality factor	492	.29	-.6 , 1.18	.53

	ZKPQ-Impulsive sensation seeking personality factor	493	.39	-.37 , 1.14	.32
	EuropASI-Legal status	701	1.07	.75 , 1.52	.72
	Unemployment	568	.85	.6 , 1.21	.38
Suicide attempt	Age at onset of SUD [b]	699	.00	-.04 , .05	.89
	Number of outpatient treatments	681	.98	.84 , 1.13	.77
	ZKPQ-Aggression Hostility personality factor	495	-.02	-.58 , .55	.96
	Psychotic symptoms	678	1.03	.75 , 1.41	.86
	EuropASI-Legal status	701	1.08	.76 , 1.55	.66
	EuropASI-Medical status	699	1.13	.84 , 1.51	.41
	EuropASI-Psychiatric status	701	1.35	1.03 , 1.78	**2.94E-02**
	Lifetime medical conditions	705	.95	.7 , 1.31	.76
	Educational attainment	701	1.10	.83 , 1.46	.51
Educational attainment	Age at onset of substance use [b]	701	.01	-.03 , .06	.50
	Age at onset of SUD [b]	699	.01	-.03 , .06	.54
	Number of therapeutic community interventions	698	.86	.67 , 1.11	.26
	Number of outpatient treatments	681	1.06	.91 , 1.23	.47
	ZKPQ-Neuroticism Anxiety personality factor	492	.16	-.73 , 1.06	.72
	Criminal record	488	1.27	.86 , 1.89	.23
	Unemployment	568	.73	.51 , 1.04	.08
	Substance use family history	516	1.00	.7 , 1.45	.99
	Educational attainment	701	.94	.71 , 1.26	.69
Well being	Number of outpatient treatments	681	1.00	.86 , 1.17	.97
	EuropASI- Psychiatric status	701	1.01	.77 , 1.32	.96
	EuropASI-Family/social relationships	701	.95	.73 , 1.25	.22
	Unemployment	568	.87	.61 , 1.25	.46
	Number of psychiatric hospitalizations	504	1.23	.73 , 2.07	.44
	Substance use family history	516	1.05	.73 , 1.52	.80

Note. Odds Ratio (OR) is reported for logistic regression and ordinal regression; Beta is reported for lineal regression; Incidence Rate Ratio (IRR) is reported for negative binomial regression.

[a] For binary traits sample size was calculated with the formula 4/(1/n1+1/n2).

[b] Logarithmic transformations were applied to continuous variables not following a normal distribution.

4 DISCUSSION

SUDs are complex and multifactorial disorders, involving the interplay of both genetic and environmental factors. Furthermore, SUDs exhibit high heterogeneity across a wide range of dimensions, including the type of substance or substances consumed, the age at onset of substance use, the presence of comorbid conditions, and the treatment outcomes. Over the past decade, the field of psychiatric genetics has focused on identifying individual genes and loci that conferee a higher risk of developing psychiatric disorders, including SUDs. To pursue this goal, GWASs to date have identified various risk loci for substance-specific SUDs, and more recently, the first risk loci for the general addiction risk factor. However, given the substantial polygenic complexity that characterizes SUDs, it became clear the importance of employing larger samples and post-GWAS methodologies to understand the genome-wide genetic architecture of SUDs. In addition, there is a growing body of evidence supporting the role of gene-environment interactions in the development of SUDs, contributing to the complexity of the disorder. Furthermore, information gathered from GWASs has evidenced the high genetic overlap between SUDs and a wide-range of psychiatric disorders and other related traits, and current genomic methodologies aim to unravel the complex mechanisms involved in the co-occurrence of these conditions.

The present thesis comprises two studies that utilize in-house clinical cohorts with phenotypical and genetic data available and state-of-the-art genomic techniques to provide genetic insights into the heterogeneity and comorbidity of SUDs. The first study particularly focuses on the relationship between SUDs and ADHD, revealing evidence of a shared genetic background underlying substance-specific SUDs and substance use phenotypes between the general population and individuals with ADHD, as well as evidence of bidirectional causal relationships for some of the substance use phenotypes tested. Furthermore, the second study leverages deep phenotyping information from a SUDs cohort, providing an overview of the pattern of associations of the genetic liability for multiple psychiatric, behavioral and related traits with various SUD-related phenotypes, including sociodemographic and health outcomes, comorbidity and personality traits and SUD variables. This study also uncovers evidence of GxE, indicating that genetic liability for suicide attempt worsened the psychiatric status in SUDs individuals with a history of emotional physical and/or sexual abuse. Collectively, the

results presented in this thesis support the current literature highlighting the strong genetic relationship between ADHD and SUDs and contribute to a better understanding of the role of the genetic liability for mental health-related conditions and adverse life experiences in the heterogeneity observed in SUDs. In this section, the results derived in this thesis are discussed in detail in the context of the current literature.

1. Current Approaches for Evaluating SUDs in Genetic Studies

The complex nature of SUDs and its heterogeneity pose challenges in accurately characterizing the phenotype, which in turn can limit the success of genetic studies of SUDs. Traditionally, SUDs phenotypes in GWASs has been defined based on diagnostic criteria from psychiatric manuals, such as the DSM or the International Classification of Diseases (ICD). However, relying solely on these criteria might oversimplify the phenotype and fail to capture the full complexity of SUDs. The diagnosis assessment of SUDs is subject to substantial variability due to many factors, including substance classification (e.g., illicit vs. legal substances or polysubstance use), definition (e.g., problematic use vs. dependence) and diagnosis guidelines used (e.g., DSM vs. ICD).

Refining the phenotype of SUDs is an essential point in order to make significant progress on the field. In order to do so, there are several important factors to take into account, including dimensional phenotyping (e.g., mild vs. severe SUD), the use of intermediate phenotypes (e.g., frequency/quantity of substance use) and the presence of comorbid traits and/or environmental risk factors. In this section, the different diagnostic tools used to characterize SUDs will be discussed and compared to the benefits of the minimal phenotyping approach used by large biobanks, and the deep phenotyping approach used in clinical settings. Moreover, two additional sources of heterogeneity in the assessment of SUDs, mainly substance-specific vs. polysubstance and illicit vs. legal substances, will be discussed.

1.1. Differences in Diagnoses Guidelines: Transition from DSM-IV to DSM-5

The DSM is a manual detailing diagnostic criteria for mental health disorders, including SUDs. In 2013, the DSM-5 was released, replacing the DSM-IV, which had been used for over a decade (released in 1994) (American Psychiatric Association, 2013). This

major revision implicated numerous changes in diagnostic criteria for nearly every DSM-IV disorder, for some more extensive than for others.

The release of the DSM-5 overlapped with the recruitment process for Study 1 and Study 2 of the present thesis, which started around 2005. However, the *Hospital Universitari Vall d'Hebron* has not yet implemented the transition to the revised edition. As a result, all participants have been evaluated based on the DSM-IV criteria. These participants include the cohort of 989 individuals from Study 1, who underwent assessment for DSM-IV-based ADHD and SUDs, as well as the cohort of 1,427 individuals from Study 2 who underwent assessment for all DSM-IV Axis I and II Disorders (including mental health disorders, SUDs, and personality disorders).

Even though in the DSM-5 the basic definition of SUDs remains unchanged, the distinction between substance abuse or dependence has been removed in favor of a single diagnosis of SUDs. This has resulted in changes in the diagnostic threshold. Under the DSM-IV, the diagnostic criteria differ between abuse (at least one symptom out of four, at any time) and dependence (at least three symptoms out of seven, in a 12-month period). In addition, people meeting criteria for dependence does not receive diagnosis for abuse for that class of substance. However, in the DSM-5 people meeting criteria for a SUD (at least two symptoms out of 11, in a 12-month period) can be sub-classified into mild (two-three symptoms), moderate (four-five symptoms) or severe (six or more symptoms) SUD, based on the number of criteria met. One of the reasons for this change is the evidence that the hierarchical structure between abuse and dependence does not seem to follow the anticipated relationship, meaning that abuse fails to encapsulate a less severe disease presentation of dependence (Kahler & Strong, 2006). In addition, distinguishing between abuse or dependence creates diagnostic orphans, where individuals who exhibit two dependence symptoms and no abuse symptoms do not meet any diagnostic criteria (D. S. Hasin, O'Brien, et al., 2013). Importantly, within the studies present in this thesis, individuals diagnosed with substance abuse or dependence according to DSM-IV-based criteria were categorized into a single SUDs diagnosis. This methodology was adopted to ensure an up-to-date and more coherent classification of the disorder.

Other additions to the DSM-5 criteria are the craving (or strong desire or urge to use the substance) symptom, and the withdrawal symptom for cannabis (which was not included for this substance in the DSM-IV). Importantly, studies of the general population suggest that craving does not add to the general information offered by other SUDs criteria (e.g., tolerance, withdrawal, and continuing use despite health problems) (D. S. Hasin et al., 2012). Therefore, this overlap may diminish the contribution of the craving criteria for the identification of people who already meet the threshold for a SUD through other criteria. Despite this, craving is also a component of a different diagnostic system, the ICD, and its addition to the DSM-5 can improve consistency across classifications systems (Degenhardt, Bharat, Bruno, et al., 2019).

The sum of changes applied to SUDs diagnostic criteria can impact on the disorder's prevalence estimate. Although there is very little information regarding the assessment of the new DSM-5-based SUDs, available data suggest that the revised version, which appears to have more inclusive threshold criteria, will estimate a higher SUDs prevalence than the DSM-IV (Agrawal, Heath, et al., 2011). Thus, the clinical assessment based on the DSM-IV used to diagnose the clinical cohorts from our studies may have underestimated SUDs diagnosis, especially for cannabis (due to the addition of the craving symptom), and excluded a percentage of diagnostic orphans. Nevertheless, combining abuse and dependence diagnostic criteria may increase instrument reliability by providing a more accurate assessment of the underlying disorder, with the limitation that this will also introduce additional variability in the measure (D. S. Hasin, Auriacombe, et al., 2013).

Notably, the use of the revised DMS-5 edition for clinical assessment is rarely reported in any big GWAS of SUDs, or other mental-health disorders, which is expected given that all large-scale studies are predominantly composed by participants recruited previous to the release of DSM-5. In addition, it has been reported that studies relying on DSM-IV or DSM-5 SUDs diagnostic criteria offer similar information and thus can be compared when accumulating a body of evidence (Livne et al., 2021).

However, the DSM is not the only diagnostic manual employed to assess mental-health disorders. The ICD is the official world classification of most medical disorders

(World Health Organization, 2004). This tool is widely used for the diagnosis of mental and behavioral disorders of participants included in large-scale GWAS given its straightforward applicability and implementation in electronic health records databases. However, there are several significant differences in the classification of SUDs between the latest editions of the ICD and the DSM (e.g., the distinction of harmful use vs. dependence in the ICD) and the ICD system has proven to be less accurate and reliable for mental disorders, compared to the DSM system (Tyrer, 2014). Nevertheless, previous studies have found a high concordance between the two diagnostic systems when comparing ICD-10 or ICD-11 vs. DSM-IV (Degenhardt, Bharat, Bruno, et al., 2019; D. Hasin et al., 1997, 2006), although a recent study from the World Mental Health Surveys, reported lower concordances when comparing ICD-11 to DSM-5 (Degenhardt, Bharat, Bruno, et al., 2019). Furthermore, other tools frequently used in general population studies to assess the risk of alcohol use disorder and nicotine dependence, namely the Alcohol Use Disorders Identification Test (AUDIT) and FTND, respectively, have been validated and show high rates of concordance with the DSM system (Agrawal, Scherrer, et al., 2011; Moehring et al., 2019). It is essential to ensure the consistency of within-subject diagnostic findings across countries, languages and cultures in order to produce translational and scalable research.

1.2. Phenotyping Strategies for Advancing Genetic Studies on SUDs: Dimensional vs. Minimal Phenotyping

This section offers an overview of phenotyping strategies employed to assess SUDs within the context of the two studies presented in this thesis. Relying solely on categorical diagnoses of SUDs, such as the ones provided by tools like the DSM or ICD, where diagnosis is predominantly based on the presence of a minimal number of symptoms from a list, may not capture the full complexity of the disorder. SUDs are characterized by substantial phenotypic heterogeneity, it involves different substances and it can manifest on a spectrum with varying severity levels (Beseler et al., 2006). Clinical and genetic variability can arise from differences in sex, age at onset, developmental course, treatment response, symptomatology and co-occurrence of comorbid conditions. In addition, the contribution of environmental risk factors may vary across different manifestations of the disorder (Prom-Wormley et al., 2017). The combination of

individuals with different symptoms profiles into a single group (e.g., the case group in a case/control GWAS) reduces statistical power to detect important etiological information (Huckins, 2022). The variety of methodologies and tools employed to assess SUDs in genetic studies are reflected in both studies from the present thesis.

In Study 1, our cohort was assessed for SUDs under the DSM-IV criteria, classifying individuals in case/control groups. In addition, in Study 1 we utilized publicly available data on GWASs for substance-specific SUDs. Given the lack of large-scale GWASs based on DSM diagnoses for some of the SUD phenotypes at the time of the study, we utilized GWASs on intermediate phenotypes (frequency/quantitative traits) for tobacco and cannabis. These included smoking initiation, age of smoking initiation, cigarettes per day and smoking cessation for tobacco, and lifetime cannabis use for cannabis. In addition, to address the lack of a large-scale GWAS combining all substances we utilized self-reported questionnaire data from the UK Biobank ("have you ever been addicted to illicit drugs?"). Alcohol and cocaine were the only SUD phenotypes with available GWASs on dependence, but not abuse, assessed with the DSM-IV. Since publication of Study 1 in 2020, the field has done outstanding progress on increasing sample sizes for clinically diagnosed SUDs population in GWAS. Such is the case for tobacco use disorder and cannabis use disorder, were large GWASs have been recently conducted, capturing genetic signatures that correlate with those of psychiatric disorders (E. C. Johnson, Demontis, et al., 2020; Toikumo et al., 2023). Furthermore, the largest multivariable GWAS combining data from various major substances has been published, allowing the study of a unified addiction factor (Hatoum et al., 2023).

In Study 2 we were able to benefit from the deep and dimensional phenotyping assessment of our SUDs cohort. Dimensional phenotyping offers an alternative or complementary approach to gain a more comprehensive characterization of the phenotype (Tiego et al., 2023). This approach does not simply assess SUDs as a dichotomic phenotype, as having or not having the disorder, but rather as a range of behaviors and symptoms able to capture the heterogeneity and complexity of the disorder. Several factors can be taken into account when assessing SUDs as a dimensional phenotype, such as substance use patterns, addiction severity and type(s) of substance(s) used. In our study, the clinical assessment was conducted by trained psychiatrists and

psychologists to gather information on sociodemographic status (sex, age, educational attainment, employment status and criminal record), lifetime medical conditions, psychiatric and SUDs family history and substance use related variables (substance(s) of use and/or abuse, age at onset of use, age at onset of SUDs, years of substance use and SUDs treatment history). In addition, a variety of scales and questionnaires were conducted to evaluate SUD individuals in terms of SUD severity (European version of the Addiction Severity Index (EuropASI)), psychiatric comorbidity (Structured clinical interview for DSM-IV axis I and axis II disorders), personality traits (Zuckerman–Kuhlman Personality Questionnaire (ZKPQ)) and health-related quality of life (36-Item Short Form Survey (SF-36)). For instance, the EuropASI (Kokkevi & Hartgers, 2009) evaluates the severity of addiction across seven main problematic areas typically affected by substance abuse or dependence. Another relevant tool in the context of SUDs is the AUDIT, which captures various dimensions of alcohol use disorder, including the severity and patterns of alcohol consumption, as well as the associated problems and consequences (de Meneses-Gaya et al., 2009). By utilizing information on severity, as well as behavioral and environmental risk factors, genetic studies can identify the underlying genetic architecture of specific dimensions of SUDs, explore shared genetic influences across substances, and provide insights into the biological pathways involved in addiction vulnerability (Tiego et al., 2023).

However, dimensional phenotyping requires significant resources, including time and qualified psychologists and psychiatrists. The extensive assessments and interviews can be time-consuming and costly, making it challenging to achieve sample sizes for large-scale studies (Sanchez-Roige & Palmer, 2020). On the opposite extreme of the phenotypic assessment of complex diseases, minimal phenotyping refers to a simplified approach to capture only a few key characteristics associated with SUDs. This approach has been widely used by large-scale biobanks and genomic resources and has allowed to increase rapidly GWASs sample sizes in recent years. By conducting minimal phenotyping, studies characterize case/control status of participants based on self-reported data from a single question (e.g., "In the past week, how many alcoholic beverages did you have?"). The phenotypes acquired by these questions usually refer to frequency and/or quantity of substance use rather than clinical diagnoses of SUDs. This

traits, which are known to be heritable (Saunders et al., 2022), can serve as intermediate phenotypes for the disorder. However, such phenotyping strategy introduces substantial in-sample heterogeneity, by classifying as cases individuals that likely diverge on important unmeasured characteristics, lowering effect sizes and power (Feczko et al., 2019). Moreover, the genetic architecture captured by substance use might not be an accurate proxy for studying the underlying genetics of SUDs. The validity of such phenotyping strategies is still an important consideration for the genetic studies of SUDs.

The minimal phenotyping approach has had great success in discovering hundreds of risk loci for tobacco and alcohol-related phenotypes (M. Liu et al., 2019). However, prior studies suggest that consumption measures (e.g., alcohol drinking frequency and cannabis initiation) have divergent patterns of genetic correlation relative to their respective SUDs. For example, alcohol drinking frequency and quantity show opposite directions of genetic correlations with measures related to socio-economic status and different patterns of correlations with psychopathology (e.g., quantity of alcohol consumption is genetically correlated with psychopathology whereas frequency of alcohol consumption is not) (Kranzler et al., 2019). This suggests that alcohol consumption metrics measure different aspects of drinking behavior with different genetic risk profiles (Marees et al., 2020). Minimal phenotyping has also shown some success, although to a lesser degree, for cannabis use (Pasman et al., 2018). Similar than with alcohol, the moderate genetic correlation between cannabis use disorder and various psychiatric disorders, particularly ADHD (rg = 0.53) (E. C. Johnson, Demontis, et al., 2020), does not correspond with the relatively low genetic correlations observed with cannabis use (rg = 0.15) (Pasman et al., 2018). This discrepancy may be due to differences in how these intermediate measures are assessed compared to the clinical diagnosis of SUDs. For instance, when diagnosing SUDs according to the DSM, the assessment considers both present and past symptoms of the disease. Meanwhile, measures of substance use are assessed in a limited timeframe, such as the past week or the past year. In addition, large biobank samples, widely employed to study substance use related phenotypes, are often not representative of the general population, showing higher education and socio-economic status and older age (e.g., UK Biobank and 23andme) or being predominantly male (e.g., Million Veteran Program) (Sanchez-Roige & Palmer,

2020). Moreover, a recent study showed that alcohol consumption and tobacco smoking assessed in UK Biobank are subject to misreports and longitudinal changes, causing bias in gene discovery and follow-up analyses (Xue et al., 2021).

Given the complex interplay between genetic and sociological factors in the context of substance use and the development of SUDs, these biases can generate false positives in genetic studies, and complicate efforts to examine distinctions between the genetics of substance use and SUDs. Therefore, minimal phenotypes may result in an incomplete and biased understanding of the genetic architecture of the disorder (Tiego et al., 2023). In order to avoid this, appropriate phenotyping strategies are needed, particularly when assessing self-reported frequency and quantity of substance use. Future genetic research should focus on clinically defined phenotypes in addition to broad phenotypes in order to maximize the clinical applicability from genetic studies of psychiatric traits.

Emerging phenotyping strategies, such as the Hierarchical Taxonomy of Psychopathology (HiTOP), offer a dimensional and hierarchical approach to classifying mental disorders, overcoming limitations of current categorical systems like the DSM or the ICD (Waszczuk et al., 2019). The HiTOP model suggests that psychopathological symptoms and disorders can be conceptualized as existing on a continuum, ranging from normal variation to severe psychopathology. It proposes a hierarchical structure, with broad higher-order dimensions (e.g., externalizing and internalizing) that encompass more specific lower-order dimensions and individual symptoms (e.g., impulsivity and neuroticism). This approach offers alternatives for genetic studies, addressing misclassification and differentiating between broad psychopathology and dimension-specific risk factors and has the potential to increase statistical power for gene discovery (Waszczuk et al., 2019).

1.3. Heterogeneity in SUDs Assessment: Substance Specific vs. Combined Approaches and Legal vs. Illegal Substance Use

This section will address two key sources that contribute to heterogeneity in genetic studies on SUDs. First, the differences of substance-specific SUDs and the combination of all substances into a unified disorder. Second, the variations arising from the use of legal substances versus illegal substances.

The choice between studying individual substances or combining them into a unitary SUD depends on the specific research aims, available resources and balance between sample size and heterogeneity. For instance, in Study 1, we were able to assess substance-specific SUDs in an ADHD population, including alcohol, tobacco, cocaine, cannabis, and all illicit drugs combined. This allowed us to examine potential differences in the association of genetic liability for SUD subtypes with ADHD. Additionally, we performed bidirectional Mendelian randomization analysis for these SUD subtypes and ADHD. There were two main reasons for adopting this approach. First, at the time of the study, there was no large-scale GWASs combining all SUD subtypes. Therefore, we leveraged GWASs data on substance-specific SUDs from publicly available sources and assessed the corresponding SUD-related phenotypic information in our in-house ADHD cohort. Secondly, ADHD has demonstrated distinct patterns of associations across SUD subtypes (S. S. Lee et al., 2011), with the strongest genetic correlations observed with tobacco smoking-related phenotypes (Abdellaoui et al., 2021; Jang et al., 2022). Our study was able to account for this variability by analyzing substance-specific SUDs and a self-reported data from the UK biobank encompassing addiction to all illicit drugs.

However, polysubstance use, present in a large percentage of the population, presents a challenge when attempting to isolate the genetic liability of substance-specific SUDs and to assess causal relationships with comorbid psychopathology, such as ADHD. This challenge became apparent in Study 2, where nearly 50% of the SUD population under study exhibited polysubstance use (defined as the use of 3 or more substances). Given the limited sample size of our cohort, we were unable to examine substance-specific SUDs individually. Instead, we combined all illicit SUDs into a unified SUD phenotype, which included the most prevalent substances, namely cocaine, cannabis, opioids, and sedatives. This approach is supported by substantial evidence for a common addiction factor among SUDs (Palmer et al., 2012, 2015), which is also associated with psychopathology (Hatoum et al., 2022). Our study design, however, prevented us from investigating potential differences in the association between the use of individual substances and the genetic liability to co-occurring psychopathology.

It is known that both common and substance-specific genetic and environmental factors contribute to individual differences in the development of SUDs (Bhalla et al.,

2017; Palmer et al., 2012). When assessing SUDs separately per substance, we can obtain valuable insights into the distinct genetic underpinnings associated with each specific substance. However, GWASs on specific substances may face limitations in terms of sample size, which can impact the statistical power of the study. Conversely, polysubstance, which is present in approximately 50% of SUD individuals (Morley et al., 2015), poses a challenge in identifying unique substance-specific GWAS signals, and can reduce the reliability of the findings (E. C. Johnson, Chang, et al., 2020). Combining all substances into a unified disorder may facilitate larger sample sizes but also enable a comprehensive exploration of shared genetic factors among different substances (Hatoum et al., 2022). However, it may also introduce heterogeneity within the population, mainly because individuals with different substances preferences, patterns of use and underlying genetic liability are grouped together (Bhalla et al., 2017; Carroll, 2021). To address this issue, one potential approach is to adopt a hybrid strategy. This involves conducting a GWAS combining multiple substances while simultaneous performing sensitivity analysis for individual substances. As seen in the latest multivariate GWAS of SUD, this approach can help gain insights into both substance-specific and shared genetic factors underlying SUDs (Hatoum et al., 2023). Additionally, methods that aim to identify homogeneous subgroups within polysubstance samples can allow for a better understanding of the heterogeneity observed in SUD (Feczko et al., 2019).

Another source of heterogeneity when assessing the genetic etiology of SUDs is the differences between legal and illegal substances. In this thesis, the inclusion of legal and/or illegal substances varied between Study 1 and Study 2. While Study 1 included tobacco and alcohol as individual substances of interest, these were excluded as the primary substance of abuse or dependence in Study 2. However, it should be noted that individuals assessed in Study 2 could have received a diagnosis of a SUD related to tobacco or alcohol, but only in conjunction with a SUD diagnosis for an illegal substance. By focusing primarily on illegal substances, Study 2 aimed to create a more homogeneous group, facilitating a more targeted analysis of SUD-related phenotypes and outcomes. Therefore, the exclusion of legal substances minimized potential biases arising from differences in disease presentation and sociodemographic variables between legal and illegal substances. These differences are primarily driven by variations

in accessibility and overall patterns of consumption within the social context (Andersson, Lilleeng, et al., 2021; Martz et al., 2022). Legal substances, like alcohol and tobacco, are widely available due to their regulatory status and social acceptance. They often follow established patterns of consumption and can have distinct impacts on individuals and communities. In contrast, illegal substances, such as cocaine or opioids, are typically obtained through illicit markets, which often leads to limited availability and more restricted access. Additionally, the social context surrounding illegal substances tends to involve stigmatization and marginalization (L. H. Yang et al., 2017).

2. Exploring Post-GWAS Results in SUDs Through PGSs

Over the past 15 years, common genetic variants associated with SUDs have been extensively studied through GWASs. The proportion of heritability explained by these common variants ranges from 12% to 67% for illicit substances and alcohol use disorder, and 86% for tobacco use disorder. The latest multivariable GWAS on SUD identified 17 risk loci in a sample of over 1 million individuals (Hatoum et al., 2023). In polygenic disorders, such as SUDs, a single variant is not informative for assessing disease risk. Instead, PGSs can estimate an individual's genetic predisposition to traits and diseases by aggregating information across multiple genetic variants identified in GWASs. In this section, the different methods used to construct PGSs within this thesis is discussed, as well as the current utility of genetic risk profiling for clinical prediction of disease vulnerability.

2.1. Differences in PGSs Construction Methods

In both studies included in the present thesis, we leveraged information from large GWASs on SUDs, psychiatric disorders and other related traits to build PGSs on clinical cohorts where deep clinical assessment and genetic data were available. In Study 1, we utilized PGSs to investigate whether the genetic liability for substance-specific SUDs could predict SUDs diagnosis in a clinical cohort of ADHD subjects. Additionally, in Study 2 PGSs were employed to examine the pleiotropic effects of the genetic liability for comorbid psychopathology and related traits on a variety of SUD-related phenotypes (more on section 3.1).

There are substantial methodological differences regarding the construction of the PGSs between the two studies, which are worth mentioning. Both studies focused on PGSs methods that only require GWASs summary statistics, an LD reference panel and a validation or target sample with genotype data to calculate the scores. The methods used were PRSice and PRScs for Study 1 and Study 2, respectively (Choi & O'Reilly, 2019; Ge et al., 2019). The primary difference between PRSice and PRScs lies in their criteria for inclusion of variants in the resulting score (SNP preselection or genome-wide methods) and the corresponding weights assigned to them. In 2020, when Study 1 was conducted, SNP-preselection methods, such as PRSice (Choi & O'Reilly, 2019), were commonly employed for PGSs calculations. This approach involves clumping and thresholding (C+T) (Privé et al., 2019). Clumping first removes correlated SNPs, retaining only independent signals within a genomic region using a fixed LD r^2 threshold. Thresholding is then applied by selecting variants within a range of pre-defined p-value thresholds and selecting the one that yields the highest prediction accuracy in a target sample with both genotype and phenotype data available. C+T assumes that the selected SNPs are largely independent and can be fitted additively. This approach has been widely used for its computational and conceptual simplicity. However, researchers typically choose an arbitrary r^2 threshold for SNPs removal, introducing some variability (Wang et al., 2022). Additionally, there is a risk of overfitting when performing multiple tests and selecting from a wide range of p-value thresholds (Choi et al., 2020).

However, in a relatively short period, by the time Study 2 was conducted, more advanced methods for PGSs calculations emerged, such as PRScs, which utilizes genome-wide information by simultaneously modeling all SNPs for the PGS computation (Ge et al., 2019). Some major advantages of fitting genome-wide SNPs simultaneously, rather than relying only on the most predictive p-value, is that it improves substantially the predictive power and reduces the risk of overfitting, a concern in the previously mentioned method (Ge et al., 2019). This model employs a Bayesian regression framework and applies a continuous shrinkage via prior distribution on SNP effect sizes. In other words, it adjusts the estimated effect sizes of all SNPs from the given summary statistics based on LD patterns between them, so that the amount of shrinkage applied to each SNP is adaptive to the strength of its association signal in GWAS. By doing so,

this approach can accommodate diverse underlying genetic architectures, an essential feature for complex traits such as SUDs. The transition from SNP-preselection methods to genome-wide modeling in PGSs calculations represents a progression toward more sophisticated approaches that maximize the signal captured by genetic variants, thereby enhancing accuracy and mitigating potential overfitting issues (Pain et al., 2021).

Another notable distinction in describing PGSs between the two studies within this thesis relates to the terminology utilized to denote to the PGS itself. A large number of studies, including Study 1, commonly employ the term PRS ("polygenic risk score") to denote the assessment of genetic risk for a specific disease or trait. However, in Study 2, we adopt the term PGS ("genome-wide polygenic score") to refer to the polygenic scores due to two main reasons. First, by avoiding the term "risk", which implies that genetic influences are associated with negative outcomes, we are able to assess genetic liability for "positive" outcomes, such as educational attainment or well-being. Second, as previously explained, the methodology utilized to calculate the PGSs takes into account the effects of genome-wide SNPs. Therefore, by employing the term PGS, we encompass a broader range of outcomes beyond disease risk and acknowledge the methodology's incorporation of genome-wide SNP effects.

2.2. The Clinical Utility of PGSs

Many studies have demonstrated the validity of PGSs in predicting diseases status in various research settings, including research-based case-control studies, population-based cohort studies and electronic health records-based studies (Hatoum et al., 2023; Musliner et al., 2019; Zheutlin et al., 2019). However, the prediction accuracy at individual level is still limited. This limitation can be explained by two main reasons. First, genetic factors only account for a proportion of a trait's variance, meaning that the maximum accuracy of genetic prediction is limited by the heritability of the disorder. Second, current PGSs are not equipped to capture all genetic variation, but only the additive effect of common genetic variants captured by GWAS, which often have small effects. The SNP-heritability sets the upper limit for the variance that can be explained by PGSs. Therefore, the applicability of PGSs in the clinical practice may not be as straightforward as it could have been intended to be (Wray et al., 2021).

PGSs have shown some promise for the stratification of individuals at risk by their polygenic load for some health conditions, particularly for coronary artery disease, type 2 diabetes or breast cancer (Khera et al., 2018). For psychiatric diseases, however, the polygenic, heterogeneous and multifactorial nature of the disorders poses significant challenges and potential pitfalls in genetic prediction efforts (Murray et al., 2021). In a clinical cohort diagnosed with depression, PGSs for bipolar disorder and schizophrenia were found to predict the risk of developing bipolar disorder and psychotic disorders, respectively (Musliner et al., 2020). However, prediction was still stronger when using information on family history (Musliner et al., 2020). Additionally, schizophrenia PGS was able to differentiate schizophrenia from other psychosis diagnosis in first-episode psychosis individuals (Vassos et al., 2017). Current SUDs PGSs explain a relative small proportion of the phenotypic variance (2.6-6.6%) in SUD-related outcomes (Hatoum et al., 2023), limiting their current clinical utility (Barr et al., 2020). Supporting this idea, a recent study found that when combined with clinical and environmental risk factors, SUDs PGSs made only minimal contributions to disease prediction (Barr et al., 2022; Nurnberger et al., 2022). Therefore, the use of these data for SUDs is not informative enough for clinical and therapeutic settings as they do not improve upon factors already assessed in the clinic for diagnoses and prediction purposes (Barr et al., 2022; Nurnberger et al., 2022).

Prognostic ability of PGSs can be difficult to interpret. In order to translate PGSs to clinical tools, relative risk that compares individuals across the PGS continuum (highest percentile versus lowest percentile) needs to be transformed to absolute risk for diseases and disorders (Chatterjee et al., 2016; Wang et al., 2022). For instance, based on an approximation of currently available data, the schizophrenia PGS explains a liability variance of 10%, and the 10% of the population at the highest risk for schizophrenia based on PGSs have an approximate 4-fold increase in risk compared to the rest of the population (relative risk) (Trubetskoy, Pardiñas, Qi, Panagiotaropoulou, O'Donovan, et al., 2022). But because schizophrenia prevalence in the general population is of 1%, only 4% of the people in this high-risk group are expected to develop the disease (absolute risk) (Murray et al., 2021). These values are still not impactful enough to be considered in a clinical setting. As GWASs sample sizes increase, the generation of PGSs capable of

identifying individuals at higher risk than the general population will be enhanced. Although it is unlikely that this approach will achieve clinical utility for predicting individual-level outcomes for psychiatric diseases on its own, PGSs could significantly enhance risk assessment when integrated with other measures of risk. These measures may include family history, exposure to stressful life events or trauma, brain imaging and neurocognitive performance, to name a few (Wang et al., 2022). From this perspective, PGSs should be viewed similarly to many tests used in health care with the purpose of identifying individuals at higher risk of disease for stratification. Furthermore, PGSs may provide relevant information for precision medicine purposes, such as early intervention strategies, treatment response and prognostic on disease course (Murray et al., 2021). For instance, they could guide interventions such as advising against the use of recreational drugs in individuals at the highest percentile of schizophrenia PGS.

One of the biggest challenges for the implementation of PGSs into the clinic is ensuring that they can be equally applicable to all ethnic groups. Transferability of PGSs across populations, however, is still limited. Current GWAS studies, from which effect sizes are taken for PGSs calculations, are predominantly European-ancestry-based (Fatumo et al., 2022). Difference in patterns of LD and allele frequencies at disease-associated loci between populations with distinct ancestry result in attenuated predictive accuracy of PGSs across ancestry populations (M. S. Kim et al., 2018). Promising advances have been made in developing approaches that incorporate information from diverse populations to improve prediction performance, especially in underrepresented non-European populations (Wang et al., 2022).

Several private companies offer direct-to-costumer genotyping, and online tools allow individuals to upload their data and generate personal PGSs for a wide number of diseases and traits. In the free online tool impute.me (http://Impute.me) schizophrenia and alcohol use disorder are among the top searched conditions (Murray et al., 2021). However, an increase in psychological distress was seen in individuals after receiving an above-average alcohol use disorder PGSs (Driver et al., 2023). This raises concerns about how this information should be appropriately delivered to the individual and what prevention strategies should be implemented alongside it. Genetic counselling is crucial

when it comes to genetic testing, particularly in the field on mental disorders, due to the risk for misinterpretation and stigmatizing assumptions (Palk et al., 2019).

Over the past decade, PGSs have emerged as a rapidly advancing field of study within complex traits. The ability to directly assess an individual's genetic predisposition has revolutionized research by enabling the inclusion of genetic predictors in various studies. The predictive capability of PGSs does not require knowledge of the intermediate processes linking genes to behavior. While the ultimate goal is to achieve a comprehensive bottom-up explanation of psychiatric diseases from genes to the brain, the ability to make accurate predictions is also a valuable accomplishment as it has immediate practical implications for identifying individuals at risk and serves as a crucial initial step towards eventual explanation. Therefore, a pressing research priority is to enhance the predictive power of PGSs, enabling their utilization as an early warning system for prevention and prognostic strategies.

3. Exploring the Shared Genetic Architecture Between SUDs and Comorbid Conditions

One of the main focuses of psychiatric genetic research to date is to unravel the shared genetic architecture across psychiatric disorders. The data acquired from GWASs has paved the way into the estimation of the genetic overlap and the assessment of causal relationships between SUDs and other psychiatric disorder and related traits. The following sections discuss the three main methodological approaches employed within this thesis to explore the shared genetic architecture between SUDs and comorbid conditions in the context of the current literature, namely genetic correlation, PGS and Mendelian randomization analyses.

3.1. Investigating Pleiotropy Between SUDs and Comorbid Psychopathology Through Genetic Correlation and PGS Analyses

The study of the genetic overlap, also referred to as pleiotropy, between SUDs and other phenotypes can be assessed through several methodologies. A widely used approach is genetic correlation analysis, which gives an estimate of the average correlation of genetic effects across the genome of two phenotypes. Another approach

is PGS analysis, which, a part from the clinical applications discussed in the previous section, can be used to examine the genetic architecture of psychopathology by providing information of the shared genetic liability among different traits and disorders.

In Study 1, genetic correlation analyses were employed to explore pairwise genetic overlap between substance-specific SUDs and substance use phenotypes and ADHD. Our results revealed that smoking-related phenotypes yield the strongest genetic correlation with ADHD. Specifically, we detected a substantial genetic correlation between smoking initiation and ADHD (rg = 0.57). Additionally, smoking initiation and cannabis use yielded the strongest genetic correlation among substance use phenotypes (rg = 0.53). However, it is noteworthy that the genetic correlation between cannabis use and ADHD was relatively low (rg = 0.15). This finding is particularly intriguing, as reports utilizing a more recent and larger GWAS for cannabis use disorder, reported a higher genetic correlation with ADHD (rg = 0.53) (E. C. Johnson, Demontis, et al., 2020). This suggests that while there may be a genetic link between cannabis use and ADHD, this correlation becomes more pronounced when considering cannabis use disorder specifically. These findings highlight the importance of differentiating between substance use and SUD phenotypes in order to obtain accurate insights into the pleiotropic nature of SUDs and its association with psychiatric disorders.

Furthermore, in Study 1, PGSs were employed to test whether the genetic liability to SUDs shares a common genetic background between the general population and ADHD individuals. PGSs were constructed for smoking initiation, alcohol or cocaine dependence, lifetime cannabis use and ever addicted to illicit drugs, using data from pre-existing GWAS datasets. Then, their association was tested with these SUDs phenotypes in an in-house ADHD sample of 989 individuals. Our results supported a common genetic background between lifetime cannabis use, alcohol dependence and smoking initiation in the general population and in subjects with ADHD. These findings are in agreement with epidemiological studies showing increased risk and higher severity of substance use, abuse and dependence in ADHD subjects (Groenman et al., 2013; van de Glind et al., 2014), and add to the existing evidence of the shared genetic vulnerability between ADHD and SUDs (Du Rietz et al., 2018; Soler Artigas et al., 2020; Wilens, 2007; Wimberley et al., 2020).

Study 2 aimed to explore the genetic overlap between major psychiatric, behavioral and other related traits with a wide range of SUD-related phenotypic outcomes, including sociodemographic and health outcomes, comorbidity and personality traits and SUD variables. PGSs for 10 comorbid and related traits were constructed in an in-house SUD cohort of 1,427 individuals, and their association was tested with 39 SUD-related phenotypes. Our findings uncovered significant pleiotropic effects of the genetic liability for several psychiatric, behavioral and related conditions on various SUD-related phenotypes. Our main findings suggest that the genetic liability for psychiatric conditions, particularly ADHD and PTSD, as well as other related traits, mainly educational attainment and well-being, may underlie, at least partially, the observed heterogeneity in SUD-related phenotypes, including educational attainment, unemployment, familiar relationship status and treatment adherence. These findings build upon existing literature providing further evidence for the presence of shared genetic factors that contribute to both the vulnerability and manifestation of SUDs and other psychiatric disorders (Dardani et al., 2021; Goldberg et al., 2014; H. Liu, 2019; van de Weijer et al., 2022; Wertz et al., 2018). In addition, our study supports that the genetic liability for distinct mental health-related traits has a substantial role in the heterogeneity that characterizes SUDs. By identifying these genetic factors, we may enhance our understanding of the underlying etiological structure of psychopathology.

Conducting studies that explore the shared genetic architecture of SUDs across mental health disorders is highly important. Other studies have employed similar methodologies to identify cross-trait associations of the genetic liability to SUDs across multiple phenotypic domains (Hartwell et al., 2022; Kember et al., 2023). Following a PheWAS approach, these studies constructed PGSs for substance-specific SUDs and explored their associations in deeply phenotyped samples with available information on psychiatric and medical conditions, family environment and early childhood experiences (Hartwell et al., 2022; Kember et al., 2023). This perspective allows to move beyond symptom-based categorizations of diseases and focuses on the underlying biological mechanisms that drive the development and progression of mental health disorders. Additionally, unraveling the genetic underpinnings that contribute to the co-occurrence

of SUDs and other psychiatric traits can move the field towards developing novel therapeutic targets and personalized treatment interventions.

3.2. Exploring Pleiotropy Across Psychiatric Disorders Through Cross-Disorder Analyses

In the field of psychiatry, the presence of comorbidity is not an exception, but rather the prevailing norm. Recent cross-disorder studies have attempted to uncover the shared biological processes that contribute to the observed phenotypic and genetic similarities among psychiatric and neurological disorders (Anttila et al., 2018; Grotzinger et al., 2022; P. H. Lee et al., 2019; Lu et al., 2021; Romero et al., 2022). While this particular line of research is not directly aligned with the focus of this thesis, this section will discuss recent findings from cross-disorder studies, shedding light on the complexities of the relationships among these disorders and the challenges researchers face in unravelling their shared genetic underpinnings.

In a 2018 study, the Brainstorm Consortium found significant genetic correlations among psychiatric disorders, particularly ADHD, major depressive disorder, bipolar disorder, anxiety disorders, and schizophrenia (Anttila et al., 2018). These correlations also extended to related traits like neuroticism and years of education. However, there was minimal overlap observed between psychiatric and neurological disorders, with the exception of migraine. Other pioneer studies used extensive and powerful genetic methodologies, such as genomic structural equation modeling (gSEM), which revealed interesting models of the genetic substructure within psychiatric disorders with high genetic overlap (e.g., mood, externalizing, psychotic or neurodevelopmental disorders). These models suggest that some genetic variants appear to have transdiagnostic influences on psychopathology, meaning that they have an effect in more than one disorder, and that the common genetic signal for psychiatric disorders is enriched in evolutionary conserved regions (Grotzinger et al., 2022; Romero et al., 2022).

However, these studies also shed light on the current challenges faced in cross-trait meta-analyses for psychiatric disorders. Despite consistently observing moderate-to-high genetic correlations (ranging from -0.13 to 0.83), the identification of shared SNPs, genes or biological mechanisms remains inconclusive. In the study conducted by Romero et al. (2022), which involved a cross-trait meta-analysis of 12 psychiatric

disorders, the authors observed substantial heterogeneity in the contribution of each individual disorder to the meta-analytic signal. The signal was strongly determined by schizophrenia and not necessarily representative of genetic variance common to multiple psychiatric disorders. Additionally, they observed that genetic overlap primarily occurs within pairs of disorders rather than across multiple disorders (Romero et al., 2022). These findings emphasize the existing differences in current GWASs of psychiatric disorders across various domains, including sample sizes, diagnostic assessment methods (e.g., minimal phenotyping versus clinical assessment), and the underlying genetic architecture (e.g., amount of polygenicity) for each disorder. These divergent factors may prevent the accurate assessment of these disorders jointly (Newson et al., 2020; Nishino et al., 2018). To address this issue, future research would benefit from the application of detailed and standardized assessment protocols on large and representative samples (Newson et al., 2020).

The shared genetic liability to SUDs and the rest of psychiatric disorders has been extensively reported, however, in the aforementioned studies, there is a noticeable absence of substance use and SUD phenotypes. The study conducted by the Brainstorm Consortium included smoking as a risk factor for other conditions (Anttila et al., 2018). However, smoking is an addictive trait on its own, and powerful GWASs on smoking phenotypes and tobacco use disorder can be used to explore the genetic correlations between smoking and other psychiatric disorders. Moreover, in the first cross-disorder GWAS meta-analysis of neuropsychiatric disorders, conducted by Lee et al. (2019), which combined eight psychiatric disorders, no data from SUDs GWASs was included in the analyses. In posterior studies, conducted by Grotzinger et al. (2022) and Romero et al. (2022), alcohol use disorder was the only SUD phenotype included. However, despite the important relationship between illicit substance use and some psychiatric disorders, especially psychotic disorders (Polimanti et al., 2017), no other SUDs was taken into account. To expand our understanding of the shared biological processes between SUDs and psychiatric disorders, it is crucial to include a broader range of substance use and SUD phenotypes into cross-disorder analyses. Building upon this line of thinking, Abdellaoui et al. (2021) extended the scope of cross-disorder studies by including three SUD phenotypes, namely alcohol dependence, nicotine dependence and cannabis use

disorder. Their study revealed that the inclusion of SUDs in a genetic factor analysis lead to changes in the underlying genetic factor structure of psychiatric disorders. This result can be attributed to the high sensitivity of factor models to the input data, and highlights the need to incorporate data on the widest range of psychiatric disorders possible to accurately estimate the underlying genetic structure.

4. Unravelling The Causal Relationship Between SUDs and ADHD

Genetic correlation and PGS association analyses, among others, are valid methodologies to test whether pleiotropy exists, but do not really distinguish between types of pleiotropy (van Rheenen et al., 2019). It is crucial to differentiate between correlation and causation when examining the relationship between SUDs and psychiatric disorders, as well as to determine the directionality of the association. It is still unclear whether psychiatric disorders increase the risk for the development of SUDs (e.g., self-medication hypothesis), whether SUDs are a risk factor for later mental health conditions, or both. Although there is probably not a unique explanation for the co-occurrence of these conditions, genetic methodologies provide relevant insights into the causal mechanisms underlying the association between SUDs and comorbid psychiatric disorders. In this thesis, we applied Mendelian randomization (Sanderson et al., 2022), a method that has been extensively employed to assess the causal relationship among psychiatric disorders, as well as many other health conditions.

In Study 1, we performed MR to assess the bidirectional causal relationship between various substance use and SUD phenotypes (smoking initiation, cannabis use, and alcohol, cocaine and illicit drug dependence) and ADHD. We found evidence of a causal effect of the genetic liability to ADHD on the risk for smoking initiation, age of smoking initiation and cigarettes per day. This goes in line with evidence reported by epidemiological and twin studies (Treur et al., 2015), and further supports evidence that ADHD medication reduces early smoking initiation and alleviates symptoms of smoking withdrawal (Schoenfelder et al., 2014). Other studies reporting MR findings also support that the genetic liability to ADHD increases smoking initiation, smoking heaviness and lifetime smoking, and decreases smoking cessation (Artigas et al., 2023; Jang et al., 2022; Treur et al., 2021). However, this association remains unclear with regards to tobacco use

disorder. Recent studies have failed to detect significant evidence supporting a causal relationship between the genetic liability to ADHD and the risk of nicotine dependence (Vink et al., 2021) or tobacco use disorder (Toikumo et al., 2023), which contradicts the previous observations concerning various smoking-related phenotypes. One explanation for this discrepancy may be that the genetic liability to ADHD, as manifested through its characteristic symptoms like impulsiveness, plays distinct roles in the initiation of substance use (e.g., smoking initiation) compared to the subsequent transition to the disorder (e.g., tobacco use disorder). However, these lack of associations could also be explained by the limited power in the MR analyses undertaken, either due to the small samples size in the outcome GWAS (Vink et al., 2021), or the small number of genetic instruments included in the analysis (Toikumo et al., 2023).

Additionally, in our MR analysis, we found evidence of a causal effect of the genetic liability to ADHD on an increased risk of cannabis use, which is supported by twin studies (Elkins et al., 2018) and genetically-informed studies (Soler Artigas et al., 2020; Treur et al., 2021). Our study design does not allow to identify the extract mechanisms by which ADHD precedes smoking-behaviors and cannabis use. Nevertheless, these causal relationships could be explained through the undercontrol/disinhibition pathway, whereby the high levels of impulsivity that characterizes individuals with ADHD leads to substance use without considering the associated negative consequences (Jean et al., 2022; Molinero & Hinckley, 2023). Another plausible mechanism is the self-medication pathway, whereby the substance is used for its alleviating effects on ADHD symptoms (van Amsterdam et al., 2018).

Furthermore, we found some evidence of reverse causation, where the genetic liability to both smoking initiation and cannabis use have a casual effect on the risk for ADHD. For smoking initiation, however, our sensitivity analysis, similarly than in previous MR studies (Treur et al., 2021), suggested that this link may be the result of horizontal pleiotropy. Under this context, genetic variants appear to influence both phenotypes simultaneously, rather than suggesting vertical pleiotropy, indicative of causation. In addition, twin studies have reported conflictive evidence on the causal effect of smoking on increased ADHD symptoms (e.g., attention problems) later in life (Elkins et al., 2020; Treur et al., 2015), adding to the inconclusive relationship between smoking initiation

and ADHD. For cannabis use, however, evidence of reverse causation had not been previously reported. Nevertheless, these results raise some questions about the appropriate temporal sequence by which the risk factor (smoking initiation or cannabis use) precedes the outcome (ADHD).

On one hand, this causal relationship could reflect the effects of parental exposure to substances of abuse on psychiatric outcomes in offspring. Several studies have demonstrated that maternal cannabis use during pregnancy is associated with offspring externalizing problems in childhood (Ikeda et al., 2022; Paul et al., 2021). This hypothesis is supported by the knowledge linking the endocannabinoid system with neurogenesis and the association between cannabis exposure (in uterus) and neuronal functioning (Smith et al., 2020). This suggests that the exposure to cannabis during sensitive and critical gestational phases may impact the risk of ADHD development through epigenetic modifications and subsequent gene expression alterations related to neurodevelopment (Smith et al., 2020). In addition, animal and human studies have found that pre-gestational cannabis exposure can induce epigenetic alterations in the sperm, which could promote germline epigenetic inheritance in the offspring (Mazzeo & Meccariello, 2023; Murphy et al., 2018) and cause deleterious long-term behavioral effects in the offspring (Levin et al., 2019). In line with these findings, a population study found that paternal and maternal cannabis use before pregnancy was also associated with offspring externalizing problems (El Marroun et al., 2019).

In addition to causal intrauterine effects, the association between parental smoking or cannabis use and increased risk of ADHD symptoms in the offspring may be influenced by shared genetic and familial confounding factors that were not considered in our MR analyses. For instance, dynastic effects can occur when the parental trait, influenced by parental genetics, impacts the offspring's outcome trait (Brumpton et al., 2020). In such cases, parental smoking or cannabis use might affect the household environment, thereby contributing to the manifestation of externalizing and inattentive behaviors in the offspring. Furthermore, assortative mating in SUDs and other psychiatric conditions could be another plausible mechanism that may lead to an increased risk for externalizing problems in offspring of parents who engage in substance use behaviors (Nordsletten et al., 2016). When individuals with specific genetic predispositions choose

partners who possess similar genetically influenced traits, known as assortative mating, it can induce spurious genetic associations which can result in biased estimates from MR studies (Hartwig et al., 2018; Howe et al., 2019). The sensitivity analyses conducted in our MR analysis are insufficient to account for uncontrolled confounding arising from familial effects. Within-family MR analysis, utilizing siblings or parent-offspring trios, could potentially address this bias (Brumpton et al., 2020). However, this type of analysis requires large number of genotyped family data, which is currently insufficient in cohorts assessed for psychiatric conditions (Brumpton et al., 2020).

Our MR analysis failed to detect significant evidence supporting a causal effect between ADHD and the rest of the SUD phenotypes tested (alcohol and cocaine dependence or addicted to illicit drugs). While a previous study found weak evidence for the causal effect of the genetic liability to ADHD on the risk of alcohol dependence (Treur et al., 2021), this was not supported by the latest extensive study of alcohol use disorder (Zhou, Sealock, et al., 2020). Furthermore, the lack of large-scale GWAS for cocaine dependence still poses a challenge to properly assess its relationship with major psychiatric conditions.

In summary, the current body of evidence regarding the causal relationship between SUDs and psychiatric conditions remains limited. Interesting causal relationships have been identified, including the genetic liability to cannabis use and the risk for schizophrenia (Vaucher et al., 2018) and the genetic liability for prescription opioid use and the risk for major depressive disorder (Rosoff et al., 2021). Nevertheless, future studies should employ multivariate MR techniques to validate these observed associations. This approach will allow the consideration of potential confounders, which could otherwise lead to false-positive causal findings, as well as the identification of traits that may mediate the observed causal effects among psychiatric conditions (i.e., factors that lie on the causal pathway between exposure and outcome) (Burgess & Thompson, 2015). Additionally, it is crucial to complement MR results with other genetically informative methods, such as latent causal variant (LCV) analysis (O'Connor & Price, 2018) or causal analysis using summary effect estimates (CAUSE) (Morrison et al., 2020), and, when possible, non-genetically informed methods, such as longitudinal and family-based

designs (Pingault et al., 2018). The triangulation of evidence across these methodologies will aid in unraveling the complex interplay between SUDs and mental health conditions.

5. Exploring the Implications of Gene-Environment Interactions on the Etiology of SUDs: The Role of Emotional, Physical and Sexual Abuse

In recent years, there has been a growing body of evidence highlighting the role of GxE in the development of SUDs (Pasman et al., 2019). Adding to this evidence, in Study 2, we report a significant interaction between the suicide attempt PGS and having been exposed to lifetime emotional, physical or sexual abuse on the mental health status of SUD individuals. While it is well established that exposure to sexual trauma and/or abuse increases the risk for substance use and mental health problems later in life (B. S. O'Brien & Sher, 2013), we found that a history of abuse exacerbates the negative impact of the genetic liability for suicide attempt on mental health problems of SUD individuals. Similar findings have been reported for cannabis use (Meyers et al., 2019) or bipolar disorder (Park et al., 2020), where exposure to trauma and/or maltreatment potentiates the polygenic risk for these disorders on disease development. This section discusses previous evidence regarding the genetic correlation between SUDs and suicide behaviors, explores the complexity of GxE and discusses challenges faced in GxE research methodologies and the potential for future advancements in the field.

SUDs and suicide-related behaviors exhibit a substantial genetic correlation, ranging from 0.31 to 0.62, (Colbert et al., 2021), which goes in line with the increased risk of suicide behaviors in SUD individuals (Lynch et al., 2020). In addition, prior studies suggest that both familiar and non-familiar environmental factors jointly influence liability to SUDs and suicide-related behaviors (Edwards et al., 2023; Kendler et al., 2016). A recent study on suicidality identified five risk loci that interacted with a range of environmental factors related to traumatic experiences, social support, and socioeconomic status, on the risk of suicidality (Wendt et al., 2021). Interestingly, results from functional annotation pointed to the *OPRM1* gene, which has also been linked to the risk for SUDs (Schwantes-An et al., 2016), specially opioid use disorder and alcohol consumption (Taqi et al., 2019; Weerts et al., 2017), as well as to an increased risk for suicide behaviors in depressed individuals (Nobile et al., 2019). Additionally, a GxE study reported that stressful life

events interacted with genetic variants encoded in the *OPRM1* gene modulating the risk for depression (Swann et al., 2014). Overall, these findings reinforce the shared genetic and environmental liability between suicide behaviors and SUDs.

One of the primary challenges encountered in GxE studies is the collection of high-quality measures of the environment and phenotypes in large-scale samples. These challenge was present in our GxE analysis, where only half of the individuals from our in-house SUD cohort had available information on lifetime emotional, physical and/or sexual abuse. Moreover, the identification and selection of appropriate environmental variables for assessment poses another challenge. SUDs are potentially influenced by numerous environmental factors which can overlap during an individual's life time. However, in our study we focused on a single environmental exposure. This limitation is commonly seen in most current GxE studies for SUDs, which focus on environmental exposures separately, such as childhood trauma (Meyers et al., 2019), peer substance use (Bountress et al., 2017) or parental support (Su et al., 2021).

An interesting aspect regarding GxE interactions is that the influence of the environment can vary throughout different stages of the lifespan, so that GxE effects tend to be more pronounced during early developmental periods (Keers & Pluess, 2017). However, in our study we lacked information regarding the specific timeframes in which adverse environmental exposures took place. Instead, we adopted a broader perspective, encompassing exposure across an individual's entire lifetime. Recent data delves into the advantages of adopting a life-course approach, with a specific focus on early environmental exposures, in order to fully understand how GxE contribute to the trajectory of mental health conditions (Keers & Pluess, 2017).

Given the substantial limitations that current GxE studies face, it is crucial to implement some measures, such as using more sophisticated methods, better phenotyping measures, larger sample sizes and attempting replication of reported GxE findings. Furthermore, in order to obtain large cohorts that can be meta-analysed, assessment of phenotypic outcomes and environmental factors need to be homogenized across cohorts. These efforts will increase statistical power of meta-analysis and

replication studies and yield robust findings in the investigation of how adverse life experiences interact with genetic factors to modulate the risk for SUDs.

6. Methodological Considerations in the Context of This Thesis

6.1. Power and Sample Size

A major limitation of GWASs is the need to achieve a high level of significance to account for multiple testing correction, which is based on the assumption of 1 million independent test for common genetic variation (Dudbridge & Gusnanto, 2008). The primary goal to overcome this limitation is to increase the sample size. In order to achieve this goal, large international consortia, such as the PGC-SUD working group, are meta-analysing data available on SUD phenotypes across multiple cohorts. However, assembling large sample sizes is more complicated for clinically diagnosed SUD phenotypes, such as the ones obtained with the DSM, mainly because of the cost and resources needed to obtain such data. Large-scale biobanks (e.g., UK BioBank, 23andme) have contributed towards obtaining large sample sizes by performing minimal phenotyping on substance use phenotypes. This strategy has been especially successful for tobacco and alcohol, which are phenotypes very well represented in large biobanks. For example, the GWAS from GSCAN on smoking initiation has over 1 million individuals, being one of the largest GWAS performed to date (M. Liu et al., 2019). However, for illegal SUDs, data is still very limited in biobanks, and information is mainly available through clinical diagnoses or ICD-codes from electronic health records.

The analyses presented in this thesis are based on the largest GWASs on SUD phenotypes, psychiatric disorders and related traits available at the time that the studies were conducted. The statistical power of GWASs, also quantified by the number of GWS variants identified, conditions the power of secondary and follow-up analyses, including PGSs or Mendelian randomization analyses (Abdellaoui et al., 2023). As GWASs increase in sample size, so does the power to capture the polygenicity of a trait though PGSs analyses. Moreover, Mendelian randomization analyses utilize GWS genetic variants discovered in GWASs as instruments to study and uncover causal relationships between complex traits. Therefore, at the time that Study 1 was conducted, the smaller sample

sizes and limited-powered GWAS available for illegal SUDs, especially for cocaine and alcohol dependence, may have not provided enough statistical power to accurately investigate the polygenic genetic architecture for these traits. Similarly, in Study 2, the differences in sample sizes among the discovery GWASs used to construct PGSs for psychiatric disorders and related traits (ranging from 27,280 to 2,083,151 individuals) may have contributed to the variability in the results of the study and the reduced ability to detect significant associations for those traits with smaller sample sizes.

As proposed by Hatoum et al. (2023) an alternative strategy to enhance sample size in SUDs GWASs is to combine all substances and investigate the genetic architecture of the addiction risk-factor (addiction-rf). By employing this approach, a sample size exceeding 1 million SUD individuals was achieved, without incorporating substance use phenotypes. However, despite the substantial sample size, only 17 GWS loci were associated with the addiction-rf, while 47 loci were specific to individual substances. This may reflect the phenotypic heterogeneity due to variability in the measurement of SUDs across samples and the genetic heterogeneity arising from assessing multiple SUDs with distinct underlying genetic architecture.

Moreover, we acknowledge the modest sample size of the in-house clinical cohorts presented in this work (N=989 in Study 1, and N=1,427 in Study 2). The individuals were recruited from a restricted geographic area in Barcelona by specialized clinical teams on ADHD and SUDs, respectively, by following a detailed assessment protocol. Importantly, PGSs studies rely on large sample sizes within the target sample to identify significant associations with a particular trait (Choi et al., 2020). In Study 2, a significant limitation was the strict multiple testing correction applied to account for the extensive number of tests conducted, combined with the limited sample size of the study, which may have resulted in decreased statistical power to detect true associations. However, as the recruitment of SUD individuals continues, larger sample sizes will mitigate this limitation and allow for a more in-depth assessment of SUDs heterogeneity.

6.2. Genomic Research Is Predominantly Focused on European-Ancestry Populations

The restriction of our studies to clinical cohorts exclusively composed of individuals of European ancestry has posed a significant limitation. As a consequence, all

analyses employing publicly available GWAS data were conducted solely within populations of European ancestry. This limitation is extensive to many current genetic research of complex diseases, where European ancestries are overrepresented. A 2016 study of the GWAS catalogue found that participants of Asian, African or Latin American ancestry represented only 19% of individuals studied in GWASs (Popejoy and Fullerton 2016). In 2021, and despite recent efforts made to include diverse populations, 86% of GWASs participants where from European ancestry (Fatumo et al. 2022). As a result, transferability of GWASs findings to other ancestry groups is uncertain.

The lack of transferability of GWASs findings between ancestry groups is explained by the differences in allelic frequencies and LD patterns between variants, as well as environmental differences altering effect sizes due to GxE interactions (Auton et al., 2015). Consequently, genetic associations as well as polygenic predictors may not replicate or may underperform in other ancestry groups because of reasons other than statistical power. Increasing diversity in genetic research is crucial in order to reach a comprehensive understanding of the genetic etiology of mental health and substance use. Large-scale biobanks have emerged with the aim to perform GWASs in populations of non-European ancestries (e.g., Biobank Japan, China Kandoori Biobank, Taiwan Biobank) and novel statistical methods are being developed with the main purpose to improve transferability (Ruan et al., 2022).

6.3. The Need for Sex-Stratified Genetic Studies

In our studies, we lacked statistical power to include a sex division of our results. However, sex differences in substance use and in SUDs are well-documented in epidemiological and family-based studies, are a major source of heterogeneity in the clinical presentation of SUDs and can influence disease outcome (Datta et al., 2020). Prevalence of substance use, the particular substance involved, the age at onset of substance use and substance use patterns are some of the aspects where sex differences are observed. Specifically, the gender gap is characterized by a greater prevalence and earlier onset of SUDs in men, but a faster course of substance use problems and transition to SUDs in women. In addition, women in SUD treatment are known to present more severe problems related to employment, social and family relationships, psychiatric

status and overall quality of life (McHugh et al., 2018). There are also sex differences in comorbidity rates with other psychiatric disorders, with higher rates of anxiety and depressive disorders in women and higher rates of other externalizing disorders such as conduct disorder and antisocial personality disorder in men (McHugh et al., 2018). In light of these evidence, in Study 2, it would have been interesting to perform a sex-stratified PGS association analysis. These secondary analyses could have uncovered different patterns of associations between the genetic liability for psychiatric and behavioral traits and a wide range of SUD-related phenotypes among men and women.

Current GWASs largely lack the phenotypic depth and power needed to discover the underlying molecular mechanisms driving sex differences in SUDs. Therefore, only a few GWASs have succeeded in identifying sex-specific loci for SUDs. A recent GWAS from Kranzler et al. (2019) found two female-specific loci for alcohol consumption and one for alcohol use disorder, and another from Yang et al. (2019) identified a sex-specific variant for opioid dependence in African American males. In addition, a recent PheWAS found sex-specific associations between the genetic liability to SUDs and multiple phenotypes, particularly regarding somatic and psychiatric conditions (Hartwell et al., 2022). Studies conducted in other psychiatric disorders, such as depression, have been successful in identifying sex-specific loci and genes through GWASs (Silveira et al., 2023). Characterizing the role that the genetic underpinnings have on the sex differences observed in SUD-related outcomes may help in the design of treatment and prevention strategies. A key next step in understanding these variables is increasing the representation of women in research studies.

7. Future Directions for Advancing in the Genetic Research of SUDs

In the next few years, the increasing availability of large-scale datasets for SUDs, especially for illegal substances, along with advancements in computational methods, will help make substantial progress in advancing the understanding of the genetic architecture of SUDs and other substance use behaviors. These continued advances will contribute to the development of clinically meaningful implications for SUDs prevention and treatment in the future. With special emphasis on the studies conducted in this thesis, here are some of the future prospects that should be addressed.

Study 1 included publicly available data on five substance use behavior phenotypes, namely smoking initiation, age of smoking initiation, smoking cessation, cigarettes per day and lifetime cannabis use, and three SUD phenotypes, including alcohol dependence, cocaine dependence and ever addicted to illicit drugs. Since the publication of Study 1 larger GWASs haven been published, with particular interest regarding tobacco use disorder and cannabis use disorder. The inclusion of more powerful GWASs that focus on SUDs rather than substance use could help reinforce the associations seen in the PGS and MR analyses between substance use phenotypes and ADHD.

Furthermore, future studies using complementary methods to examine the causal relationship between SUDs and ADHD should aim to clarify associations that do not follow an appropriate temporal sequence between exposure and outcome (e.g., cannabis use preceding ADHD). This could be done in multiple ways: First, MR analyses should control for familial confounding factors, such as dynastic effects and assortative mating. Within-family MR analysis, utilizing siblings or parent-offspring trios, could potentially address this bias (Brumpton et al., 2020). Second, by performing stratified analyses on childhood versus adult ADHD, as well as additional research on the effect of cannabis use on late-onset ADHD, where ADHD symptoms arise in the late adolescence or adulthood. This would help clarify whether our results reflect an effect of cannabis use on the risk on the persistent form of ADHD. This hypothesis is supported by evidence of the association between hazardous cannabis use and ADHD symptoms in adulthood (Estévez et al., 2016; Fergusson & Boden, 2008; Kolla et al., 2016). Finally, triangulation of MR results with other research methods is crucial to provide increased confidence in the inferences drawn. This includes comparison to other genetically informative methods such as twin studies, LCV analysis (O'Connor & Price, 2018), or CAUSE (Morrison et al., 2020), as well as longitudinal analyses of cohort data.

Regarding Study 2, future directions focus on expanding our SUD cohort, increasing sample size and acquiring high-quality information on a wider range of phenotypical outcomes. With larger samples, the assessment of phenotypic information on substance-specific SUDs within our cohort will allow us to explore potential differences in patterns of associations of the genetic liability to psychiatric diseases and behavioral traits between stimulant (e.g., cocaine) versus sedative (e.g., opioids) SUDs or between illegal

versus legal SUDs. Larger samples would also enable additional analysis, such as unsupervised cluster analysis, aiming to identify homogeneous subgroups of individuals based on clinical variables. This approach would help determine if the genetic liability to psychiatric and behavioral conditions exhibits stronger associations with clusters characterized by higher symptom severity.

Furthermore, access to longitudinal data from our patients would provide valuable information to assess the impact of the genetic liability on disease progression, the individual effectiveness of treatments and interventions, and other factors contributing to the heterogeneity of SUDs. Lastly, it is crucial to focus future efforts in collecting information on environmental risk factors. This will allow the exploration of a broader range of GxE and their consequences on the course of SUDs. Overall, these advancements have the potential to greatly advance our understanding of SUDs heterogeneity and the complex interplay between genetics and environmental factors in the development and progression of SUDs.

5 CONCLUSIONS

1. There is a strong genetic correlation among the studied SUD phenotypes as well as between SUDs and ADHD, with expected direction of effect.

2. There are shared genetic factors underlying SUDs in the general population and individuals with ADHD, in particular for lifetime cannabis use, alcohol dependence, and smoking initiation.

3. There is a causal effect for the genetic liability to ADHD on the risk for smoking initiation, age of smoking initiation, cigarettes per day and lifetime cannabis use.

4. There is evidence of reverse causation for the genetic liability to smoking initiation and lifetime cannabis use on an increased risk for ADHD.

5. The genetic liability for psychiatric disorders, especially for ADHD and PTSD, are associated with poorer outcomes in SUD individuals, including lower educational attainment and higher rates of unemployment, but also earlier onset of substance use and higher number of outpatient treatments.

6. The genetic liability for educational attainment and well-being are associated with better outcomes in SUD individuals, especially with lower rates of criminal records, unemployment and outpatient treatments and fewer problems related to family and social relationships.

7. The genetic liability for psychiatric disorders and other mental health-related traits exhibit different patterns of associations with SUD-related phenotypes, which may partially explain the heterogeneity observed in SUDs.

8. The genetic liability for suicide attempt aggravates the negative impact of having been exposed to emotional, physical and/or sexual abuse on the mental health status of SUD individuals.

6 REFERENCES

Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., Adekanmbi, V., Adeoye, A. M., Adetokunboh, O. O., … Amini, S. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Abdellaoui, A., Smit, D. J. A., van den Brink, W., Denys, D., & Verweij, K. J. H. (2021). Genomic relationships across psychiatric disorders including substance use disorders. *Drug and Alcohol Dependence*, *220*. https://doi.org/10.1016/j.drugalcdep.2021.108535

Abdellaoui, A., Yengo, L., Verweij, K. J. H., & Visscher, P. M. (2023). 15 years of GWAS discovery: Realizing the promise. *American Journal of Human Genetics*, *110*(2), 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011

Agrawal, A., Chou, Y. L., Carey, C. E., Baranger, D. A. A., Zhang, B., Sherva, R., Wetherill, L., Kapoor, M., Wang, J. C., Bertelsen, S., Anokhin, A. P., Hesselbrock, V., Kramer, J., Lynskey, M. T., Meyers, J. L., Nurnberger, J. I., Rice, J. P., Tischfield, J., Bierut, L. J., … Nelson, E. C. (2018). Genome-wide association study identifies a novel locus for cannabis dependence. *Molecular Psychiatry*, *23*(5), 1293–1302. https://doi.org/10.1038/mp.2017.200

Agrawal, A., Heath, A. C., & Lynskey, M. T. (2011). DSM-IV to DSM-5: The impact of proposed revisions on diagnosis of alcohol use disorders. *Addiction*, *106*(11), 1935–1943. https://doi.org/10.1111/j.1360-0443.2011.03517.x

Agrawal, A., & Lynskey, M. T. (2008). Are there genetic influences on addiction: Evidence from family, adoption and twin studies. *Addiction*, *103*(7), 1069–1081. https://doi.org/10.1111/j.1360-0443.2008.02213.x

Agrawal, A., Scherrer, J. F., Pergadia, M. L., Lynskey, M. T., Madden, P. A. F., Sartor, C. E., Grant, J. D., Duncan, A. E., Haber, J. R., Jacob, T., Bucholz, K. K., & Xian, H. (2011). A latent class analysis of DSM-IV and fagerström (FTND) criteria for nicotine dependence. *Nicotine and Tobacco Research*, *13*(10), 972–981. https://doi.org/10.1093/ntr/ntr105

Agrawal, A., Verweij, K. J. H., Gillespie, N. A., Heath, A. C., Lessov-Schlaggar, C. N., Martin, N. G., Nelson, E. C., Slutske, W. S., Whitfield, J. B., & Lynskey, M. T. (2012). The genetics of addictiona translational perspective. *Translational Psychiatry*, *2*(May). https://doi.org/10.1038/tp.2012.54

Akalın, B. Y., & Bilici, R. (2022). The Effect of Attention Deficit and Hyperactivity Disorder on Treatment Compliance in Individuals Undergoing Probation Addiction Program Treatment. *Anadolu Psikiyatri Dergisi*, *23*(1), 33–39. https://doi.org/10.5152/alphapsychiatry.2021.21304

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). *Diagnostic and Statistical Manual of Mental Disorders*.

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders (4th ed.). *Diagnostic and Statistical Manual of Mental Disorders*.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). *Diagnostic and Statistical Manual of Mental Disorders*, 947. https://doi.org/10.1176/appi.books.9780890425596

Andersson, H. W., Lauvsnes, A. D. F., & Nordfjærn, T. (2021). Emerging Adults in Inpatient Substance Use Treatment: A Prospective Cohort Study of Patient Characteristics and Treatment Outcomes. *European Addiction Research*, *27*(3), 206–215. https://doi.org/10.1159/000512156

Andersson, H. W., Lilleeng, S. E., & Ose, S. O. (2021). Comparison of social and sociodemographic characteristics and treatment goals of persons with alcohol versus drug use disorders: Result from a national census of inpatients in specialized treatment for substance use. *Addictive Behaviors Reports*, *13*, 100340. https://doi.org/10.1016/j.abrep.2021.100340

Andersson, H. W., Wenaas, M., & Nordfjærn, T. (2019). Relapse after inpatient substance use treatment: A prospective cohort study among users of illicit substances. *Addictive Behaviors*, *90*, 222–228. https://doi.org/10.1016/j.addbeh.2018.11.008

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G. J., Gormley, P., Malik, R., Patsopoulos, N. A., Ripke, S., Wei, Z., Yu, D., Lee, P. H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*(6395). https://doi.org/10.1126/science.aap8757

Armoon, B., Fleury, M. J., Bayat, A. H., Bayani, A., Mohammadi, R., & Griffiths, M. D. (2022). Quality of life and its correlated factors among patients with substance use disorders: a systematic review and meta-analysis. *Archives of Public Health*, *80*(1). https://doi.org/10.1186/s13690-022-00940-0

Artigas, M. S., Sánchez-Mora, C., Rovira, P., Vilar-Ribó, L., Ramos-Quiroga, J. A., & Ribases, M. (2023). Mendelian randomization analysis for attention deficit/hyperactivity disorder: studying a broad range of exposures and outcomes. *International Journal of Epidemiology*, *52*(2), 386–402. https://doi.org/10.1093/ije/dyac128

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Bagot, K. S., Milin, R., & Kaminer, Y. (2015). Adolescent Initiation of Cannabis Use and Early-Onset Psychosis. *Substance Abuse*, *36*(4), 524–533. https://doi.org/10.1080/08897077.2014.995332

Balhara, Y. P. S., Kuppili, P. P., & Gupta, R. (2017). Neurobiology of Comorbid Substance Use Disorders and Psychiatric Disorders: Current State of Evidence. *Journal of Addictions Nursing*, *28*(1), 11–26. https://doi.org/10.1097/JAN.0000000000000155

Barr, P. B., Driver, M. N., Kuo, S. I. C., Stephenson, M., Aliev, F., Linnér, R. K., Marks, J., Anokhin, A. P., Bucholz, K., Chan, G., Edenberg, H. J., Edwards, A. C., Francis, M. W., Hancock, D. B., Harden, K. P., Kamarajan, C., Kaprio, J., Kinreich, S., Kramer, J. R., … Dick, D. M. (2022). Clinical, environmental, and genetic risk factors for substance use disorders: characterizing combined effects across multiple cohorts. *Molecular Psychiatry*, *27*(11), 4633–4641. https://doi.org/10.1038/s41380-022-01801-6

Barr, P. B., Ksinan, A., Su, J., Johnson, E. C., Meyers, J. L., Wetherill, L., Latvala, A., Aliev, F., Chan, G., Kuperman, S., Nurnberger, J., Kamarajan, C., Anokhin, A., Agrawal, A., Rose, R. J., Edenberg, H. J., Schuckit, M., Kaprio, J., & Dick, D. M. (2020). Using polygenic scores for identifying

individuals at increased risk of substance use disorders in clinical and population samples. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-00865-8

Basedow, L. A., Wiedmann, M. F., Kuitunen-Paul, S., Roessner, V., & Golub, Y. (2023). Attenuated psychotic symptoms, substance use and self-reported PTSD in adolescence. *European Journal of Psychotraumatology*, *14*(2), 2193327. https://doi.org/10.1080/20008066.2023.2193327

Baselmans, B. M. L., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., Bao, Y., Smart, M., Kumari, M., Willemsen, G., Hottenga, J. J., Boomsma, D. I., de Geus, E. J. C., Nivard, M. G., & Bartels, M. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature Genetics*, *51*(3), 445–451. https://doi.org/10.1038/s41588-018-0320-8

Belsky, J., & Pluess, M. (2009). Beyond Diathesis Stress: Differential Susceptibility to Environmental Influences. *Psychological Bulletin*, *135*(6), 885–908. https://doi.org/10.1037/a0017376

Berrettini, W. (2017). A brief review of the genetics and pharmacogenetics of opioid use disorders. *Dialogues in Clinical Neuroscience*, *19*(3), 229–236. https://doi.org/10.31887/dcns.2017.19.3/wberrettini

Beseler, C., Jacobson, K. C., Kremen, W. S., Lyons, M. J., Glatt, S. J., Faraone, S. V, Gillespie, N. A., & Tsuang, M. T. (2006). Is there heterogeneity among syndromes of substance use disorder for illicit drugs? *Addictive Behaviors*, *31*(6), 929–947. https://doi.org/10.1016/j.addbeh.2006.03.037

Bhalla, I. P., Stefanovics, E. A., & Rosenheck, R. A. (2017). Clinical Epidemiology of Single Versus Multiple Substance Use Disorders. *Medical Care*, *55*(9), S24–S32. https://doi.org/10.1097/MLR.0000000000000731

Bierut, L. J., Dinwiddie, S. H., Begleiter, H., Crowe, R. R., Hesselbrock, V., Nurnberger, J. I., Porjesz, B., Schuckit, M. A., & Reich, T. (1998). Familial transmission of substance dependence: Alcohol, marijuana, cocaine, and habitual smoking: A report from the collaborative study on the genetics of alcoholism. *Archives of General Psychiatry*, *55*(11), 982–988. https://doi.org/10.1001/archpsyc.55.11.982

Blanco, C., Alegría, A. A., Liu, S. M., Secades-Villa, R., Sugaya, L., Davies, C., & Nunes, E. V. (2012). Differences among major depressive disorder with and without co-occurring substance use disorders and substance-induced depressive disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry*, *73*(6), 865–873. https://doi.org/10.4088/JCP.10m06673

Boschloo, L., Vogelzangs, N., Van den Brink, W., Smit, J. H., Beekman, A. T. F., & Penninx, B. W. J. H. (2012). Predictors of the 2-year recurrence and persistence of alcohol dependence. *Addiction*, *107*(9), 1639–1640. https://doi.org/10.1111/j.1360-0443.2012.03860.x

Bountress, K., Chassin, L., & Lemery-Chalfant, K. (2017). Parent and peer influences on emerging adult substance use disorder: A genetically informed study. *Development and Psychopathology*, *29*(1), 121–142. https://doi.org/10.1017/S095457941500125X

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., … Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian

randomization through within-family analyses. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-17117-4

Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review*, *70*(March), 51–63. https://doi.org/10.1016/j.cpr.2019.04.002

Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, *37*(7), 658–665. https://doi.org/10.1002/gepi.21758

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., Hartwig, F. P., Holmes, M. V., Minelli, C., Relton, C. L., & Theodoratou, E. (2020). Guidelines for performing Mendelian randomization investigations. *Wellcome Open Research*, *4*, 186. https://doi.org/10.12688/wellcomeopenres.15555.2

Burgess, S., & Thompson, S. G. (2015). Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *American Journal of Epidemiology*, *181*(4), 251–260. https://doi.org/10.1093/aje/kwu283

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, *8*(12). https://doi.org/10.1371/journal.pcbi.1002822

Busse, A., Gerra, G., Koutsenok, I., & Saenz, E. (2015). International standards on the treatment of drug use disorders. In *Drug and Alcohol Dependence* (Vol. 156). World Health Organization. https://doi.org/10.1016/j.drugalcdep.2015.07.1006

Butelman, E. R., Chen, C. Y., Brown, K. G., Lake, K. J., & Kreek, M. J. (2021). Age of onset of heaviest use of cannabis or alcohol in persons with severe opioid or cocaine use disorders. *Drug and Alcohol Dependence*, *226*, 108834. https://doi.org/10.1016/j.drugalcdep.2021.108834

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Cabana-Domínguez, J., Roncero, C., Grau-López, L., Rodríguez-Cintas, L., Barral, C., Abad, A. C., Erikson, G., Wineinger, N. E., Torrico, B., Arenas, C., Casas, M., Ribasés, M., Cormand, B., & Fernàndez-Castillo, N. (2016). A Highly Polymorphic Copy Number Variant in the NSF Gene is Associated with Cocaine Dependence. *Scientific Reports*, *6*, 31033. https://doi.org/10.1038/srep31033

Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N., & Cormand, B. (2019). Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *94*, 109667. https://doi.org/10.1016/j.pnpbp.2019.109667

Cadoret, R. J., Yates, W. R., Troughton, E., Woodworth, G., & Stewart, M. A. (1995). Adoption Study Demonstrating Two Genetic Pathways to Drug Abuse. *Archives of General Psychiatry*, *52*(1), 42–52. https://doi.org/10.1001/archpsyc.1995.03950130042005

Capusan, A. J., Gustafsson, P. A., Kuja-Halkola, R., Igelström, K., Mayo, L. M., & Heilig, M. (2021).

Re-examining the link between childhood maltreatment and substance use disorder: a prospective, genetically informative study. *Molecular Psychiatry*, *26*(7), 3201–3209. https://doi.org/10.1038/s41380-021-01071-8

Carey, C. E., Agrawal, A., Bucholz, K. K., Hartz, S. M., Lynskey, M. T., Nelson, E. C., Bierut, L. J., & Bogdan, R. (2016). Associations between polygenic risk for psychiatric disorders and substance involvement. *Frontiers in Genetics*, *7*(AUG). https://doi.org/10.3389/fgene.2016.00149

Carliner, H., Mauro, P. M., Brown, Q. L., Shmulewitz, D., Rahim-Juwel, R., Sarvet, A. L., Wall, M. M., Martins, S. S., Carliner, G., & Hasin, D. S. (2017). The widening gender gap in marijuana use prevalence in the U.S. during a period of economic change, 2002–2014. In *Drug and Alcohol Dependence* (Vol. 170). https://doi.org/10.1016/j.drugalcdep.2016.10.042

Carroll, K. M. (2021). The Profound Heterogeneity of Substance Use Disorders: Implications for Treatment Development. *Current Directions in Psychological Science*, *30*(4), 358–364. https://doi.org/10.1177/09637214211026984

Castelpietra, G., Knudsen, A. K. S., Agardh, E. E., Armocida, B., Beghi, M., Iburg, K. M., Logroscino, G., Ma, R., Starace, F., Steel, N., Addolorato, G., Andrei, C. L., Andrei, T., Ayuso-Mateos, J. L., Banach, M., Bärnighausen, T. W., Barone-Adesi, F., Bhagavathula, A. S., Carvalho, F., … Monasta, L. (2022). The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990–2019: Findings from the Global Burden of Disease Study 2019. *The Lancet Regional Health - Europe*, *16*, 100341. https://doi.org/10.1016/j.lanepe.2022.100341

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Charach, A., Yeung, E., Climans, T., & Lillie, E. (2011). Childhood attention-deficit/hyperactivity disorder and future substance use disorders: Comparative meta-analyses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(1), 9–21. https://doi.org/10.1016/j.jaac.2010.09.019

Charney, D. A., Palacios-Boix, J., Negrete, J. C., Dobkin, P. L., & Gill, K. J. (2005). Association between concurrent depression and anxiety and six-month outcome of addiction treatment. *Psychiatric Services*, *56*(8), 927–933. https://doi.org/10.1176/appi.ps.56.8.927

Chassin, L., Flora, D. B., & King, K. M. (2004). Trajectories of alcohol and drug use and dependence from adolescence to adulthood: The effects of familial alcoholism and personality. *Journal of Abnormal Psychology*, *113*(4), 483–498. https://doi.org/10.1037/0021-843X.113.4.483

Chatterjee, N., Shi, J., & García-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, *17*(7), 392–406. https://doi.org/10.1038/nrg.2016.27

Cheng, Z., Zhou, H., Sherva, R., Farrer, L. A., Kranzler, H. R., & Gelernter, J. (2018). Genome-wide Association Study Identifies a Regulatory Variant of RGMA Associated With Opioid Dependence in European Americans. *Biological Psychiatry*, *84*(10), 762–770. https://doi.org/10.1016/j.biopsych.2017.12.016

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk

score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, *8*(7), giz082. https://doi.org/10.1093/gigascience/giz082

Christoffersen, L. A. N., Mortensen, E. L., Becker, U., Osler, M., Sørensen, H. J., & Flensborg-Madsen, T. (2021). Age at onset and age at treatment of alcohol use disorders: Associations with educational level and intelligence. *Alcohol*, *95*, 7–14. https://doi.org/10.1016/j.alcohol.2021.04.005

Cicchetti, D., & Handley, E. D. (2019). Child maltreatment and the development of substance use and disorder. *Neurobiology of Stress*, *10*. https://doi.org/10.1016/j.ynstr.2018.100144

Clarke, T. K., Adams, M. J., Davies, G., Howard, D. M., Hall, L. S., Padmanabhan, S., Murray, A. D., Smith, B. H., Campbell, A., Hayward, C., Porteous, D. J., Deary, I. J., & McIntosh, A. M. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112117). *Molecular Psychiatry*, *22*(10), 1376–1384. https://doi.org/10.1038/mp.2017.153

Colbert, S. M. C., Hatoum, A. S., Shabalin, A., Li, Q. S., Coon, H., Nelson, E. C., Agrawal, A., Docherty, A. R., & Johnson, E. C. (2021). Exploring the genetic overlap of suicide-related behaviors and substance use disorders. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *186*(8), 445–455. https://doi.org/10.1002/ajmg.b.32880

Compton, W. M., Valentino, R. J., & DuPont, R. L. (2021). Polysubstance use in the U.S. opioid crisis. In *Molecular Psychiatry* (Vol. 26, Issue 1, pp. 41–50). https://doi.org/10.1038/s41380-020-00949-3

Covault, J., Tennen, H., Armeli, S., Conner, T. S., Herman, A. I., Cillessen, A. H. N., & Kranzler, H. R. (2007). Interactive Effects of the Serotonin Transporter 5-HTTLPR Polymorphism and Stressful Life Events on College Student Drinking and Drug Use. *Biological Psychiatry*, *61*(5), 609–616. https://doi.org/10.1016/j.biopsych.2006.05.018

Crum, R. M., Ford, D. E., Storr, C. L., & Chan, Y. F. (2004). Association of sleep disturbance with chronicity and remission of alcohol dependence: Data from a population-based prospective study. *Alcoholism: Clinical and Experimental Research*, *28*(10), 1533–1540. https://doi.org/10.1097/01.ALC.0000141915.56236.40

Dardani, C., Riglin, L., Leppert, B., Sanderson, E., Rai, D., Howe, L. D., Davey Smith, G., Tilling, K., Thapar, A., Davies, N. M., Anderson, E., & Stergiakouli, E. (2021). Is genetic liability to ADHD and ASD causally linked to educational attainment? *International Journal of Epidemiology*, *50*(6), 2011–2023. https://doi.org/10.1093/ije/dyab107

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284–1287. https://doi.org/10.1038/ng.3656

Datta, U., Schoenrock, S. E., Bubier, J. A., Bogue, M. A., Jentsch, J. D., Logan, R. W., Tarantino, L. M., & Chesler, E. J. (2020). Prospects for finding the mechanisms of sex differences in addiction with human and model organism genetic analysis. *Genes, Brain and Behavior*, *19*(3), e12645.

https://doi.org/10.1111/gbb.12645

Davies, N. M., von Hinke Kessler Scholder, S., Farbmacher, H., Burgess, S., Windmeijer, F., & Smith, G. D. (2015). The many weak instruments problem and Mendelian randomization. *Statistics in Medicine*, *34*(3), 454–468. https://doi.org/10.1002/sim.6358

Dawson, D. A., Goldstein, R. B., & Grant, B. F. (2007). Rates and correlates of relapse among individuals in remission from DSM-IV alcohol dependence: A 3-year follow-up. *Alcoholism: Clinical and Experimental Research*, *31*(12), 2036–2045. https://doi.org/10.1111/j.1530-0277.2007.00536.x

de Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., & Crippa, J. A. S. (2009). Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. *Psychology & Neuroscience*, *2*(1), 83–97. https://doi.org/10.3922/j.psns.2009.1.12

Deak, J. D., & Johnson, E. C. (2021). Genetics of substance use disorders: A review. In *Psychological Medicine* (Vol. 51, Issue 13, pp. 2189–2200). Cambridge University Press. https://doi.org/10.1017/S0033291721000969

Deak, J. D., Levey, D. F., Wendt, F. R., Zhou, H., Galimberti, M., Kranzler, H. R., Gaziano, J. M., Stein, M. B., Polimanti, R., & Gelernter, J. (2022). Genome-Wide Investigation of Maximum Habitual Alcohol Intake in US Veterans in Relation to Alcohol Consumption Traits and Alcohol Use Disorder. *JAMA Network Open*, *5*(10), E2238880. https://doi.org/10.1001/jamanetworkopen.2022.38880

Deak, J. D., Zhou, H., Galimberti, M., Levey, D. F., Wendt, F. R., Sanchez-Roige, S., Hatoum, A. S., Johnson, E. C., Nunez, Y. Z., Demontis, D., Børglum, A. D., Rajagopal, V. M., Jennings, M. V., Kember, R. L., Justice, A. C., Edenberg, H. J., Agrawal, A., Polimanti, R., Kranzler, H. R., & Gelernter, J. (2022). Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci. *Molecular Psychiatry*, *27*(10), 3970–3979. https://doi.org/10.1038/s41380-022-01709-1

Degenhardt, L., Bharat, C., Bruno, R., Glantz, M. D., Sampson, N. A., Lago, L., Aguilar-Gaxiola, S., Alonso, J., Andrade, L. H., Bunting, B., Caldas-de-Almeida, J. M., Cia, A. H., Gureje, O., Karam, E. G., Khalaf, M., McGrath, J. J., Moskalewicz, J., Lee, S., Mneimneh, Z., … Kessler, R. C. (2019). Concordance between the diagnostic guidelines for alcohol and cannabis use disorders in the draft ICD-11 and other classification systems: analysis of data from the WHO's World Mental Health Surveys. *Addiction*, *114*(3), 534–552. https://doi.org/10.1111/add.14482

Degenhardt, L., Bharat, C., Glantz, M. D., Sampson, N. A., Al-Hamzawi, A., Alonso, J., Andrade, L. H., Bunting, B., Cia, A., De Girolamo, G., De Jonge, P., Demyttenaere, K., Gureje, O., Haro, J. M., Harris, M. G., He, Y., Hinkov, H., Karam, A. N., Karam, E. G., … Kessler, R. C. (2019). Association of Cohort and Individual Substance Use with Risk of Transitioning to Drug Use, Drug Use Disorder, and Remission from Disorder: Findings from the World Mental Health Surveys. *JAMA Psychiatry*, *76*(7), 708–720. https://doi.org/10.1001/jamapsychiatry.2019.0163

Degenhardt, L., Bharat, C., Glantz, M. D., Sampson, N. A., Scott, K., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L. H., Bromet, E. J., Bruffaerts, R., Bunting, B., de Girolamo, G., Gureje, O., Haro, J. M., Harris, M. G., He, Y., de Jonge, P., … Kessler, R. C. (2019). The epidemiology of drug use disorders cross-nationally: Findings from the WHO's World Mental Health Surveys. *International Journal of Drug Policy*, *71*, 103–112. https://doi.org/10.1016/j.drugpo.2019.03.002

Demontis, D., Rajagopal, V. M., Thorgeirsson, T. E., Als, T. D., Grove, J., Leppälä, K., Gudbjartsson, D. F., Pallesen, J., Hjorthøj, C., Reginsson, G. W., Tyrfingsson, T., Runarsdottir, V., Qvist, P., Christensen, J. H., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Huckins, L. M., Stahl, E. A., Timmermann, A., … Børglum, A. D. (2019). Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nature Neuroscience*, *22*(7), 1066–1074. https://doi.org/10.1038/s41593-019-0416-1

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., … Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. https://doi.org/10.1038/s41588-018-0269-7

Diblasi, E., Kang, J., & Docherty, A. R. (2021). Genetic contributions to suicidal thoughts and behaviors. *Psychological Medicine*, *51*(13), 2148–2155. https://doi.org/10.1017/S0033291721001720

Dick, D. M., & Kendler, K. S. (2012). The impact of gene-environment interaction on alcohol use disorders. In *Alcohol research : current reviews* (Vol. 34, Issue 3, pp. 318–324).

Docherty, A. R., Mullins, N., Ashley-Koch, A. E., Qin, X. J., Coleman, J., Shabalin, A. A., Kang, J., Murnyak, B., McMahon, B. H., Oslin, D. W., Hauser, E. R., Hauser, M. A., Li, Q., Coon, H., Kimbrel, N. A., Beckham, J. C., & Ruderfer, D. (2022). Genome-wide association study meta-analysis of suicide attempt in 43,871 cases identifies twelve genome-wide significant loci. *MedRxiv*. https://doi.org/10.1101/2022.07.03.22277199

Dodge, R., Sindelar, J., & Sinha, R. (2005). The role of depression symptoms in predicting drug abstinence in outpatient substance abuse treatment. *Journal of Substance Abuse Treatment*, *28*(2), 189–196. https://doi.org/10.1016/j.jsat.2004.12.005

Driver, M. N., Kuo, S. I. C., Dron, J. S., Austin, J., & Dick, D. M. (2023). The impact of receiving polygenic risk scores for alcohol use disorder on psychological distress, risk perception, and intentions to reduce drinking. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *192*(5–6), 93–101. https://doi.org/10.1002/ajmg.b.32933

Du Rietz, E., Coleman, J., Glanville, K., Choi, S. W., O'Reilly, P. F., & Kuntsi, J. (2018). Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(7), 635–643. https://doi.org/10.1016/j.bpsc.2017.11.013

Dube, S. R., Miller, J. W., Brown, D. W., Giles, W. H., Felitti, V. J., Dong, M., & Anda, R. F. (2006). Adverse childhood experiences and the association with ever using alcohol and initiating alcohol use during adolescence. *Journal of Adolescent Health*, *38*(4), 444.e1-444.e10. https://doi.org/10.1016/j.jadohealth.2005.06.006

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology*, *32*(3), 227–234. https://doi.org/10.1002/gepi.20297

Edwards, A. C., Ohlsson, H., Lannoy, S., Stephenson, M., Crump, C., Sundquist, J., Sundquist, K., & Kendler, K. S. (2023). Shared genetic and environmental etiology between substance use disorders and suicidal behavior. *Psychological Medicine*, *53*(6), 2380–2388.

https://doi.org/10.1017/S0033291721004256

El Marroun, H., Bolhuis, K., Franken, I. H. A., Jaddoe, V. W. V., Hillegers, M. H., Lahey, B. B., & Tiemeier, H. (2019). Preconception and prenatal cannabis use and the risk of behavioural and emotional problems in the offspring; A multi-informant prospective longitudinal study. *International Journal of Epidemiology*, *48*(1), 287–296. https://doi.org/10.1093/ije/dyy186

Elkins, I. J., Saunders, G. R. B., Malone, S. M., Keyes, M. A., McGue, M., & Iacono, W. G. (2018). Associations between childhood ADHD, gender, and adolescent alcohol and marijuana involvement: A causally informative design. *Drug and Alcohol Dependence*, *184*, 33–41. https://doi.org/10.1016/j.drugalcdep.2017.11.011

Elkins, I. J., Saunders, G. R. B., Malone, S. M., Wilson, S., McGue, M., & Iacono, W. G. (2020). Differential implications of persistent, remitted, and late-onset ADHD symptoms for substance abuse in women and men: A twin study from ages 11 to 24. *Drug and Alcohol Dependence*, *212*, 107947. https://doi.org/10.1016/j.drugalcdep.2020.107947

Elliott, J. C., Carey, K. B., & Bonafide, K. E. (2012). Does family history of alcohol problems influence college and university drinking or substance use? A meta-analytical review. *Addiction*, *107*(10), 1774–1785. https://doi.org/10.1111/j.1360-0443.2012.03903.x

Elliott, J. C., Stohl, M., Wall, M. M., Keyes, K. M., Goodwin, R. D., Skodol, A. E., Krueger, R. F., Grant, B. F., & Hasin, D. S. (2014). The risk for persistent adult alcohol and nicotine dependence: The role of childhood maltreatment. *Addiction*, *109*(5), 842–850. https://doi.org/10.1111/add.12477

Elliott, J. C., Stohl, M., Wall, M. M., Keyes, K. M., Skodol, A. E., Eaton, N. R., Shmulewitz, D., Goodwin, R. D., Grant, B. F., & Hasin, D. S. (2016). Childhood maltreatment, personality disorders and 3-year persistence of adult alcohol and nicotine dependence in a national sample. *Addiction*, *111*(5), 913–923. https://doi.org/10.1111/add.13292

Ellis, M. S., Kasper, Z. A., & Cicero, T. J. (2020). The impact of opioid use disorder on levels of educational attainment: Perceived benefits and consequences. *Drug and Alcohol Dependence*, *206*, 107618. https://doi.org/10.1016/j.drugalcdep.2019.107618

Estévez, N., Dey, M., Eich-Höchli, D., Foster, S., Gmel, G., & Mohler-Kuo, M. (2016). Adult attention-deficit/hyperactivity disorder and its association with substance use and substance use disorders in young men. *Epidemiology and Psychiatric Sciences*, *25*(3), 255–266. https://doi.org/10.1017/S2045796015000360

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., Rohde, L. A., Sonuga-Barke, E. J. S., Tannock, R., & Franke, B. (2015). Attention-deficit/hyperactivity disorder. *Nature Reviews Disease Primers*, *1*, 15020. https://doi.org/10.1038/nrdp.2015.20

Farmer, R. F., Gau, J. M., Seeley, J. R., & Kosty, D. B. (2022). Family-Based Predictors of Alcohol Use Disorder (AUD) Recurrence and New Non-Alcohol Substance Use Disorder Onset Following Initial AUD Recovery. *Journal of Studies on Alcohol and Drugs*, *83*(2), 239–247. https://doi.org/10.15288/jsad.2022.83.239

Farmer, R. F., Seeley, J. R., Kosty, D. B., Gau, J. M., Duncan, S. C., Lynskey, M. T., & Lewinsohn, P. M. (2015). Internalizing and externalizing psychopathology as predictors of cannabis use disorder onset during adolescence and early adulthood. *Psychology of Addictive Behaviors*,

29(3), 541–551. https://doi.org/10.1037/adb0000059

Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., & Kuchenbaecker, K. (2022). A roadmap to increase diversity in genomic studies. In *Nature Medicine* (Vol. 28, Issue 2, pp. 243–250). https://doi.org/10.1038/s41591-021-01672-4

Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A. M., Nigg, J. T., & Fair, D. A. (2019). The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends in Cognitive Sciences*, *23*(7), 584–601. https://doi.org/10.1016/j.tics.2019.03.009

Fenton, M. C., Keyes, K., Geier, T., Greenstein, E., Skodol, A., Krueger, B., Grant, B. F., & Hasin, D. S. (2012). Psychiatric comorbidity and the persistence of drug use disorders in the United States. *Addiction*, *107*(3), 599–609. https://doi.org/10.1111/j.1360-0443.2011.03638.x

Fergusson, D. M., & Boden, J. M. (2008). Cannabis use and adult ADHD symptoms. In *Drug and Alcohol Dependence* (Vol. 95, Issues 1–2, pp. 90–96). Elsevier Science. https://doi.org/10.1016/j.drugalcdep.2007.12.012

Flanagan, J. C., Korte, K. J., Killeen, T. K., & Back, S. E. (2016). Concurrent Treatment of Substance Use and PTSD. *Current Psychiatry Reports*, *18*(8), 70. https://doi.org/10.1007/s11920-016-0709-y

Fleury, M. J., Djouini, A., Huỳnh, C., Tremblay, J., Ferland, F., Ménard, J. M., & Belleville, G. (2016). Remission from substance use disorders: A systematic review and meta-analysis. *Drug and Alcohol Dependence*, *168*, 293–306. https://doi.org/10.1016/j.drugalcdep.2016.08.625

Fluharty, M. E., Sallis, H., & Munafò, M. R. (2018). Investigating possible causal effects of externalizing behaviors on tobacco initiation: A Mendelian randomization analysis. *Drug and Alcohol Dependence*, *191*, 338–342. https://doi.org/10.1016/j.drugalcdep.2018.07.015

Fonseca, F., Robles-Martínez, M., Tirado-Muñoz, J., Alías-Ferri, M., Mestre-Pintó, J. I., Coratu, A. M., & Torrens, M. (2021). A Gender Perspective of Addictive Disorders. In *Current Addiction Reports* (Vol. 8, Issue 1, pp. 89–99). https://doi.org/10.1007/s40429-021-00357-9

Fowler, T., Lifford, K., Shelton, K., Rice, F., Thapar, A., Neale, M. C., McBride, A., & Van Den Bree, M. B. M. (2007). Exploring the relationship between genetic and environmental influences on initiation and progression of substance use. *Addiction*, *102*(3), 413–422. https://doi.org/10.1111/j.1360-0443.2006.01694.x

Fridell, M., Bäckström, M., Hesse, M., Krantz, P., Perrin, S., & Nyhlén, A. (2019). Prediction of psychiatric comorbidity on premature death in a cohort of patients with substance use disorders: A 42-year follow-up. *BMC Psychiatry*, *19*(1), 1–13. https://doi.org/10.1186/s12888-019-2098-3

Gage, S. H., Jones, H. J., Burgess, S., Bowden, J., Davey Smith, G., Zammit, S., & Munafò, M. R. (2017). Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychological Medicine*, *47*(5), 971–980. https://doi.org/10.1017/S0033291716003172

Gage, S. H., Jones, H. J., Taylor, A. E., Burgess, S., Zammit, S., & Munafò, M. R. (2017). Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomization. *Scientific Reports*, *7*(1), 40653. https://doi.org/10.1038/srep40653

Ganguli, M., Dodge, H. H., & Mulsant, B. H. (2002). Drug use and the risk of major depressive disorder, alcohol dependence, and substance use disorders. *Archives of General Psychiatry*, *59*(11), 1039–1044. https://doi.org/10.1001/archpsyc.59.11.1039

Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., Lafleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., … O'Leary, T. J. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, *70*, 214–223. https://doi.org/10.1016/j.jclinepi.2015.09.016

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1), 1776. https://doi.org/10.1038/s41467-019-09718-5

Gelernter, J., & Polimanti, R. (2021). Genetics of substance use disorders in the era of big data. In *Nature Reviews Genetics* (Vol. 22, Issue 11, pp. 712–729). https://doi.org/10.1038/s41576-021-00377-1

Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H. R., & Farrer, L. (2014). Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Molecular Psychiatry*, *19*(6), 717–723. https://doi.org/10.1038/mp.2013.99

George, O., & Koob, G. F. (2017). Individual differences in the neuropsychopathology of addiction. *Dialogues in Clinical Neuroscience*, *19*(3), 217–229. https://doi.org/10.31887/dcns.2017.19.3/gkoob

Goldberg, J., Magruder, K. M., Forsberg, C. W., Kazis, L. E., Üstün, T. B., Friedman, M. J., Litz, B. T., Vaccarino, V., Heagerty, P. J., Gleason, T. C., Huang, G. D., & Smith, N. L. (2014). The association of PTSD with physical and mental health functioning and disability (VA Cooperative Study #569: The course and consequences of posttraumatic stress disorder in Vietnam-era Veteran twins). *Quality of Life Research*, *23*(5), 1579–1591. https://doi.org/10.1007/s11136-013-0585-4

Goldstein, B. I., & Bukstein, O. G. (2010). Comorbid substance use disorders among youth with bipolar disorder: Opportunities for early identification and prevention. *Journal of Clinical Psychiatry*, *71*(3), 348–358. https://doi.org/10.4088/JCP.09r05222gry

Goodwin, D. W., Schulsinger, F., Hermansen, L., Guze, S. B., & Winokur, G. (1973). Alcohol Problems in Adoptees Raised Apart From Alcoholic Biological Parents. *Archives of General Psychiatry*, *28*(2), 238–243. https://doi.org/10.1001/archpsyc.1973.01750320068011

Grant, J. D., Scherrer, J. F., Lynskey, M. T., Agrawal, A., Duncan, A. E., Haber, J. R., Heath, A. C., & Bucholz, K. K. (2012). Associations of Alcohol, Nicotine, Cannabis, and Drug Use/Dependence with Educational Attainment: Evidence from Cotwin-Control Analyses. *Alcoholism: Clinical and Experimental Research*, *36*(8), 1412–1420. https://doi.org/10.1111/j.1530-0277.2012.01752.x

Groenman, A. P., Oosterlaan, J., Rommelse, N., Franke, B., Roeyers, H., Oades, R. D., Sergeant, J. A., Buitelaar, J. K., & Faraone, S. V. (2013). Substance use disorders in adolescents with attention deficit hyperactivity disorder: A 4-year follow-up study. *Addiction*, *108*(8), 1503–1511. https://doi.org/10.1111/add.12188

Groenman, A. P., Schweren, L. J. S., Weeda, W., Luman, M., Noordermeer, S. D. S., Heslenfeld, D. J., Franke, B., Faraone, S. V., Rommelse, N., Hartman, C. A., Hoekstra, P. J., Buitelaar, J., & Oosterlaan, J. (2019). Stimulant treatment profiles predicting co-occurring substance use disorders in individuals with attention-deficit/hyperactivity disorder. *European Child and Adolescent Psychiatry*, *28*(9), 1213–1222. https://doi.org/10.1007/s00787-019-01283-y

Grotzinger, A. D., Mallard, T. T., Akingbuwa, W. A., Ip, H. F., Adams, M. J., Lewis, C. M., McIntosh, A. M., Grove, J., Dalsgaard, S., Lesch, K. P., Strom, N., Meier, S. M., Mattheisen, M., Børglum, A. D., Mors, O., Breen, G., Mors, O., Meier, S. M., Lee, P. H., … Nivard, M. G. (2022). Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics*, *54*(5), 548–559. https://doi.org/10.1038/s41588-022-01057-4

Gujska, J. H., Silczuk, A., Madejek, R., & Szulc, A. (2023). Exploring the Link Between Attention-Deficit Hyperactivity Disorder and Cannabis Use Disorders: A Review. *Medical Science Monitor*, *29*, e939749. https://doi.org/10.12659/MSM.939749

Gurriarán, X., Rodríguez-López, J., Flórez, G., Pereiro, C., Fernández, J. M., Fariñas, E., Estévez, V., Arrojo, M., & Costas, J. (2019). Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. *Genes, Brain and Behavior*, *18*(3), e12504. https://doi.org/10.1111/gbb.12504

Hancock, D. B., Guo, Y., Reginsson, G. W., Gaddis, N. C., Lutz, S. M., Sherva, R., Loukola, A., Minica, C. C., Markunas, C. A., Han, Y., Young, K. A., Gudbjartsson, D. F., Gu, F., McNeil, D. W., Qaiser, B., Glasheen, C., Olson, S., Landi, M. T., Madden, P. A. F., … Johnson, E. O. (2018). Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Molecular Psychiatry*, *23*(9), 1911–1919. https://doi.org/10.1038/mp.2017.193

Hancock, D. B., Reginsson, G. W., Gaddis, N. C., Chen, X., Saccone, N. L., Lutz, S. M., Qaiser, B., Sherva, R., Steinberg, S., Zink, F., Stacey, S. N., Glasheen, C., Chen, J., Gu, F., Frederiksen, B. N., Loukola, A., Gudbjartsson, D. F., Brüske, I., Landi, M. T., … Stefansson, K. (2015). Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Translational Psychiatry*, *5*, 651. https://doi.org/10.1038/tp.2015.149

Hannah Ritchie and Pablo Arriagada and Max Roser. (2022). *Opioids, cocaine, cannabis and other illicit drugs*. Our World in Data. https://ourworldindata.org/illicit-drug-use

Harris, M. G., Bharat, C., Glantz, M. D., Sampson, N. A., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., Caldas de Almeida, J. M., Cia, A. H., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Hinkov, H., Karam, E. G., Karam, G., Lee, S., Lépine, J. P., Levinson, D., … Degenhardt, L. (2019). Cross-national patterns of substance use disorder treatment and associations with mental disorder comorbidity in the WHO World Mental Health Surveys. *Addiction*, *114*(8), 1446–1459. https://doi.org/10.1111/add.14599

Hartwell, E. E., Merikangas, A. K., Verma, S. S., Ritchie, M. D., Kranzler, H. R., & Kember, R. L. (2022). Genetic liability for substance use associated with medical comorbidities in electronic health records of African- and European-ancestry individuals. *Addiction Biology*, *27*(1), e13099. https://doi.org/10.1111/adb.13099

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology*, *42*(7), 608–620. https://doi.org/10.1002/gepi.22138

Hasin, D., Fenton, M. C., Skodol, A., Krueger, R., Keyes, K., Geier, T., Greenstein, E., Blanco, C., & Grant, B. (2011). Personality disorders and the 3-year course of alcohol, drug, and nicotine use disorders. *Archives of General Psychiatry*, *68*(11), 1158–1167. https://doi.org/10.1001/archgenpsychiatry.2011.136

Hasin, D., Grant, B. F., Cottler, L., Blaine, J., Towle, L., Üstün, B., & Sartorius, N. (1997). Nosological comparisons of alcohol and drug diagnoses: A multisite, multi-instrument international study. *Drug and Alcohol Dependence*, *47*(3), 217–226. https://doi.org/10.1016/S0376-8716(97)00092-6

Hasin, D., Hatzenbuehler, M. L., Keyes, K., & Ogburn, E. (2006). Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and International Classification of Diseases, tenth edition (ICD-10). *Addiction*, *101*(SUPPL. 1), 59–75. https://doi.org/10.1111/j.1360-0443.2006.01584.x

Hasin, D. S., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Crowley, T., Grant, B. F., O'Brien, C., Petry, N. M., Schuckit, M., & Wall, M. M. (2013). The DSM-5 field trials and reliability of alcohol use disorder. *American Journal of Psychiatry*, *170*(4), 442–443. https://doi.org/10.1176/appi.ajp.2013.13010032

Hasin, D. S., Fenton, M. C., Beseler, C., Park, J. Y., & Wall, M. M. (2012). Analyses related to the development of DSM-5 criteria for substance use related disorders: 2. Proposed DSM-5 criteria for alcohol, cannabis, cocaine and heroin disorders in 663 substance abuse patients. *Drug and Alcohol Dependence*, *122*(1–2), 28–37. https://doi.org/10.1016/j.drugalcdep.2011.09.005

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Compton, W. M., Crowley, T., Ling, W., Petry, N. M., Schuckit, M., & Grant, B. F. (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry*, *170*(8), 834–851. https://doi.org/10.1176/appi.ajp.2013.12060782

Hatoum, A. S., Colbert, S. M. C., Johnson, E. C., Huggett, S. B., Deak, J. D., Pathak, G. A., Jennings, M. V., Paul, S. E., Karcher, N. R., Hansen, I., Baranger, D. A. A., Edwards, A., Grotzinger, A. D., Adkins, D. E., Adkins, A. E., Alanne-Kinnunen, M., Alexander, J. C., Aliev, F., Bacanu, S.-A., … Agrawal, A. (2023). Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health*, *1*(3), 210–223. https://doi.org/10.1038/s44220-023-00034-y

Hatoum, A. S., Johnson, E. C., Colbert, S. M. C., Polimanti, R., Zhou, H., Walters, R. K., Gelernter, J., Edenberg, H. J., Bogdan, R., & Agrawal, A. (2022). The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology*, *47*(10), 1739–1745. https://doi.org/10.1038/s41386-021-01209-w

Hellwege, J. N., Keaton, J. M., Giri, A., Gao, X., Velez Edwards, D. R., & Edwards, T. L. (2017). Population Stratification in Genetic Association Studies. *Current Protocols in Human Genetics*, *95*(1), 1.22.1-1.22.23. https://doi.org/10.1002/cphg.48

Henkel, D. (2011). Unemployment and substance use: A review of the Literature (1990-2010). *Current Drug Abuse Reviews*, *4*(1), 4–27. https://doi.org/10.2174/1874473711104010004

Hesse, M. (2009). Integrated psychological treatment for substance use and co-morbid anxiety or depression vs. treatment for substance use alone. A systematic review of the published

literature. In *BMC Psychiatry* (Vol. 9). https://doi.org/10.1186/1471-244X-9-6

Hicks, B. M., Johnson, W., Durbin, C. E., Blonigen, D. M., Iacono, W. G., & McGue, M. (2013). Gene-environment correlation in the development of adolescent substance abuse: Selection effects of child personality and mediation via contextual risk factors. *Development and Psychopathology*, *25*(1), 119–132. https://doi.org/10.1017/S0954579412000946

Hopfer, C. J., Crowley, T. J., & Hewitt, J. K. (2003). Review of twin and adoption studies of adolescent substance use. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(6), 710–719. https://doi.org/10.1097/01.CHI.0000046848.56865.54

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*(3), 343–352. https://doi.org/10.1038/s41593-018-0326-7

Howe, L. J., Lawson, D. J., Davies, N. M., St. Pourcain, B., Lewis, S. J., Davey Smith, G., & Hemani, G. (2019). Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. *Nature Communications*, *10*(1), 5039. https://doi.org/10.1038/s41467-019-12424-x

Huckins, L. M. (2022). Thoughtful Phenotype Definitions Empower Participants and Power Studies. *Complex Psychiatry*, *8*(3–4), 57–62. https://doi.org/10.1159/000527022

Huizink, A. C., Levälahti, E., Korhonen, T., Dick, D. M., Pulkkinen, L., Rose, R. J., & Kaprio, J. (2010). Tobacco, cannabis, and other illicit drug use among finnish adolescent twins: Causal relationship or correlated liabilities? *Journal of Studies on Alcohol and Drugs*, *71*(1), 5–14. https://doi.org/10.15288/jsad.2010.71.5

Hunt, G. E., Large, M. M., Cleary, M., Lai, H. M. X., & Saunders, J. B. (2018). Prevalence of comorbid substance use in schizophrenia spectrum disorders in community and clinical settings, 1990–2017: Systematic review and meta-analysis. *Drug and Alcohol Dependence*, *191*, 234–258. https://doi.org/10.1016/j.drugalcdep.2018.07.011

Hunt, G. E., Malhi, G. S., Cleary, M., Lai, H. M. X., & Sitharthan, T. (2016). Prevalence of comorbid bipolar and substance use disorders in clinical settings, 1990–2015: Systematic review and meta-analysis. *Journal of Affective Disorders*, *206*, 331–349. https://doi.org/10.1016/j.jad.2016.07.011

Hunt, G. E., Malhi, G. S., Lai, H. M. X., & Cleary, M. (2020). Prevalence of comorbid substance use in major depressive disorder in community and clinical settings, 1990–2019: Systematic review and meta-analysis. *Journal of Affective Disorders*, *266*, 288–304. https://doi.org/10.1016/j.jad.2020.01.141

Hussong, A. M., Jones, D. J., Stein, G. L., Baucom, D. H., & Boeding, S. (2011). An internalizing pathway to alcohol use and disorder. *Psychology of Addictive Behaviors*, *25*(3), 390–404. https://doi.org/10.1037/a0024519

Icick, R., Moggi, F., Slobodin, O., Dom, G., Mathys, F., Van Den Brink, W., Levin, F. R., Blankers, M., Kaye, S., Demetrovics, Z., Van De Glind, G., Velez-Pastrana, M. C., & Schellekens, A. S. A. (2020). Attention Deficit/Hyperactivity Disorder and Global Severity Profiles in Treatment-Seeking Patients with Substance Use Disorders. *European Addiction Research*, *26*(4–5), 201–

210. https://doi.org/10.1159/000508546

Ikeda, A. S., Knopik, V. S., Bidwell, L. C., Parade, S. H., Goodman, S. H., Emory, E. K., & Palmer, R. H. C. (2022). A Review of Associations between Externalizing Behaviors and Prenatal Cannabis Exposure: Limitations & Future Directions. *Toxics*, *10*(1). https://doi.org/10.3390/toxics10010017

Ilbegi, S., Groenman, A. P., Schellekens, A., Hartman, C. A., Hoekstra, P. J., Franke, B., Faraone, S. V., Rommelse, N. N. J., & Buitelaar, J. K. (2018). Substance use and nicotine dependence in persistent, remittent, and late-onset ADHD: A 10-year longitudinal study from childhood to young adulthood. *Journal of Neurodevelopmental Disorders*, *10*(1). https://doi.org/10.1186/s11689-018-9260-y

Jaffee, S. R., & Price, T. S. (2007). Gene-environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*(5), 432–442. https://doi.org/10.1038/sj.mp.4001950

Jang, S. K., Saunders, G., Liu, M. Z., Jiang, Y., Liu, D. J., & Vrieze, S. (2022). Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychological Medicine*, *52*(5), 968–978. https://doi.org/10.1017/S003329172000272X

Jean, F. A. M., Arsandaux, J., Montagni, I., Collet, O., Fatséas, M., Auriacombe, M., Ramos-Quiroga, J. A., Côté, S. M., Tzourio, C., & Galéra, C. (2022). Attention deficit hyperactivity disorder symptoms and cannabis use after 1 year among students of the i-Share cohort. *European Psychiatry*, *65*(1), 1–18. https://doi.org/10.1192/j.eurpsy.2022.14

Johnson, E. C., Chang, Y., & Agrawal, A. (2020). An Update on the Role of Common Genetic Variation Underlying Substance Use Disorders. *Current Genetic Medicine Reports*, *8*(2), 35–46. https://doi.org/10.1007/s40142-020-00184-w

Johnson, E. C., Demontis, D., Thorgeirsson, T. E., Walters, R. K., Polimanti, R., Hatoum, A. S., Sanchez-Roige, S., Paul, S. E., Wendt, F. R., Clarke, T. K., Lai, D., Reginsson, G. W., Zhou, H., He, J., Baranger, D. A. A., Gudbjartsson, D. F., Wedow, R., Adkins, D. E., Adkins, A. E., … Børglum, A. D. (2020). A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry*, *7*(12), 1032–1045. https://doi.org/10.1016/S2215-0366(20)30339-4

Johnson, E. O., Chen, L. S., Breslau, N., Hatsukami, D., Robbins, T., Saccone, N. L., Grucza, R. A., & Bierut, L. J. (2010). Peer smoking and the nicotinic receptor genes: An examination of genetic and environmental risks for nicotine dependence. *Addiction*, *105*(11), 2014–2022. https://doi.org/10.1111/j.1360-0443.2010.03074.x

Kahler, C. W., & Strong, D. R. (2006). A Rasch model analysis of DSM-IV Alcohol abuse and dependence items in the National Epidemiological Survey on Alcohol and Related Conditions. *Alcoholism, Clinical and Experimental Research*, *30*(7), 1165–1175. https://doi.org/10.1111/j.1530-0277.2006.00140.x

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Wagner, G. G. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 245–257. https://doi.org/10.1038/s41588-018-0309-3

Katikireddi, S. V., Green, M. J., Taylor, A. E., Davey Smith, G., & Munafò, M. R. (2018). Assessing causal relationships using genetic proxies for exposures: an introduction to Mendelian randomization. *Addiction*, *113*(4), 764–774. https://doi.org/10.1111/add.14038

Katzman, M. A., Bilkey, T. S., Chokka, P. R., Fallu, A., & Klassen, L. J. (2017). Adult ADHD and comorbid disorders: Clinical implications of a dimensional approach. *BMC Psychiatry*, *17*(1), 302. https://doi.org/10.1186/s12888-017-1463-3

Kaufman, J., Yang, B. Z., Douglas-Palumberi, H., Crouse-Artus, M., Lipschitz, D., Krystal, J. H., & Gelernter, J. (2007). Genetic and Environmental Predictors of Early Alcohol Use. *Biological Psychiatry*, *61*(11), 1228–1234. https://doi.org/10.1016/j.biopsych.2006.06.039

Kaye, S., Ramos-Quiroga, J. A., van de Glind, G., Levin, F. R., Faraone, S. V, Allsop, S., Degenhardt, L., Moggi, F., Barta, C., Konstenius, M., Franck, J., Skutle, A., Bu, E.-T., Koeter, M. W. J., Demetrovics, Z., Kapitány-Fövény, M., Schoevers, R. A., van Emmerik-van Oortmerssen, K., Carpentier, P.-J., … van den Brink, W. (2019). Persistence and Subtype Stability of ADHD Among Substance Use Disorder Treatment Seekers. *Journal of Attention Disorders*, *23*(12), 1438–1453. https://doi.org/10.1177/1087054716629217

Keers, R., & Pluess, M. (2017). Childhood quality influences genetic sensitivity to environmental influences across adulthood: A life-course Gene × Environment interaction study. *Development and Psychopathology*, *29*(5), 1921–1933. https://doi.org/10.1017/S0954579417001493

Kember, R. L., Hartwell, E. E., Xu, H., Rotenberg, J., Almasy, L., Zhou, H., Gelernter, J., & Kranzler, H. R. (2023). Phenome-wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped Sample. *Biological Psychiatry*, *93*(6), 536–545. https://doi.org/10.1016/j.biopsych.2022.08.010

Kember, R. L., Vickers-Smith, R., Xu, H., Toikumo, S., Niarchou, M., Zhou, H., Hartwell, E. E., Crist, R. C., Rentsch, C. T., Davis, L. K., Justice, A. C., Sanchez-Roige, S., Kampman, K. M., Gelernter, J., & Kranzler, H. R. (2022). Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nature Neuroscience*, *25*(10), 1279–1287. https://doi.org/10.1038/s41593-022-01160-z

Kendler, K. S., Davis, C. G., & Kessler, R. C. (1997). The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity survey: A family history study. *British Journal of Psychiatry*, *170*(JUNE), 541–548. https://doi.org/10.1192/bjp.170.6.541

Kendler, K. S., Jacobson, K. C., Prescott, C. A., & Neale, M. C. (2003). Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *American Journal of Psychiatry*, *160*(4), 687–695. https://doi.org/10.1176/appi.ajp.160.4.687

Kendler, K. S., Lönn, S. L., Maes, H. H., Lichtenstein, P., Sundquist, J., & Sundquist, K. (2016). A Swedish Population-Based Multivariate Twin Study of Externalizing Disorders. *Behavior Genetics*, *46*(2), 183–192. https://doi.org/10.1007/s10519-015-9741-7

Kendler, K. S., Myers, J., & Prescott, C. A. (2007). Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Archives of General Psychiatry*, *64*(11), 1313–1320. https://doi.org/10.1001/archpsyc.64.11.1313

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, *20*(4), 359–364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations 06 Biological Sciences 0604 Genetics. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1561-7

Kim, S. Y., Jeong, S. H., & Park, E. C. (2023). Age at onset of alcohol consumption and its association with alcohol misuse in adulthood. *Neuropsychopharmacology Reports*, *43*(1), 40–49. https://doi.org/10.1002/npr2.12302

Kinreich, S., McCutcheon, V. V, Aliev, F., Meyers, J. L., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Kuang, W., Pandey, G., Viteri, S. S. S. de, Francis, M. W., Chan, G., Bourdon, J. L., Dick, D. M., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., … Porjesz, B. (2021). Predicting alcohol use disorder remission: a longitudinal multimodal multi-featured machine learning approach. *Translational Psychiatry*, *11*(1), 166. https://doi.org/10.1038/s41398-021-01281-2

Kirsch, D. E., & Lippard, E. T. C. (2022). Early life stress and substance use disorders: The critical role of adolescent substance use. *Pharmacology Biochemistry and Behavior*, *215*. https://doi.org/10.1016/j.pbb.2022.173360

Kokkevi, A., & Hartgers, C. (1995). EuropASI: European Adaptation of a Multidimensional Assessment Instrument for Drug and Alcohol Dependence. *European Addiction Research*, *1*(4), 208–210. https://doi.org/10.1159/000259089

Kokkevi, A., & Hartgers, C. (2009). EuropASI: European Adaptation of a Multidimensional Assessment Instrument for Drug and Alcohol Dependence. *European Addiction Research*, *1*(4), 208–210. https://doi.org/10.1159/000259089

Kolla, N. J., van der Maas, M., Toplak, M. E., Erickson, P. G., Mann, R. E., Seeley, J., & Vingilis, E. (2016). Adult attention deficit hyperactivity disorder symptom profiles and concurrent problems with alcohol and cannabis: Sex differences in a representative, population survey. *BMC Psychiatry*, *16*(1), 50. https://doi.org/10.1186/s12888-016-0746-4

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129. https://doi.org/10.1016/S0893-133X(00)00195-0

Koob, G. F., & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, *59*, 29–53. https://doi.org/10.1146/annurev.psych.59.103006.093548

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: a neurocircuitry analysis. *The Lancet Psychiatry*, *3*(8), 760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Kranzler, H. R., Zhou, H., Kember, R. L., Vickers Smith, R., Justice, A. C., Damrauer, S., Tsao, P. S., Klarin, D., Baras, A., Reid, J., Overton, J., Rader, D. J., Cheng, Z., Tate, J. P., Becker, W. C.,

Concato, J., Xu, K., Polimanti, R., Zhao, H., & Gelernter, J. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09480-8

Krawczyk, N., Feder, K. A., Saloner, B., Crum, R. M., Kealhofer, M., & Mojtabai, R. (2017). The association of psychiatric comorbidity with treatment completion among clients admitted to substance use treatment programs in a U.S. national sample. *Drug and Alcohol Dependence*, *175*, 157–163. https://doi.org/10.1016/j.drugalcdep.2017.02.006

Ku, C. S., Loy, E. Y., Salim, A., Pawitan, Y., & Chia, K. S. (2010). The discovery of human genetic variations and their use as disease markers: Past, present and future. *Journal of Human Genetics*, *55*(7), 403–415. https://doi.org/10.1038/jhg.2010.55

Lähteenvuo, M., Batalla, A., Luykx, J. J., Mittendorfer-Rutz, E., Tanskanen, A., Tiihonen, J., & Taipale, H. (2021). Morbidity and mortality in schizophrenia with comorbid substance use disorders. *Acta Psychiatrica Scandinavica*, *144*(1), 42–49. https://doi.org/10.1111/acps.13291

Lannoy, S., Mars, B., Heron, J., & Edwards, A. C. (2022). Suicidal ideation during adolescence: The roles of aggregate genetic liability for suicide attempts and negative life events in the past year. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *63*(10), 1164–1173. https://doi.org/10.1111/jcpp.13653

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Turley, P. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lee, P. H., Anttila, V., Won, H., Feng, Y. C. A., Rosenthal, J., Zhu, Z., Tucker-Drob, E. M., Nivard, M. G., Grotzinger, A. D., Posthuma, D., Wang, M. M. J., Yu, D., Stahl, E. A., Walters, R. K., Anney, R. J. L., Duncan, L. E., Ge, T., Adolfsson, R., Banaschewski, T., … Smoller, J. W. (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*, *179*(7), 1469-1482.e11. https://doi.org/10.1016/j.cell.2019.11.020

Lee, P. H., Doyle, A. E., Li, X., Silberstein, M., Jung, J.-Y., Gollub, R. L., Nierenberg, A. A., Liu, R. T., Kessler, R. C., Perlis, R. H., & Fava, M. (2022). Genetic Association of Attention-Deficit/Hyperactivity Disorder and Major Depression With Suicidal Ideation and Attempts in Children: The Adolescent Brain Cognitive Development Study. *Biological Psychiatry*, *92*(3), 236–245. https://doi.org/10.1016/j.biopsych.2021.11.026

Lee, S. S., Humphreys, K. L., Flory, K., Liu, R., & Glass, K. (2011). Prospective association of childhood attention-deficit/hyperactivity disorder (ADHD) and substance use and abuse/dependence: A meta-analytic review. *Clinical Psychology Review*, *31*(3), 328–341. https://doi.org/10.1016/j.cpr.2011.01.006

Lev-Ran, S., Imtiaz, S., Rehm, J., & Le Foll, B. (2013). Exploring the association between lifetime prevalence of mental illness and transition from substance use to substance use disorders: Results from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC). *American Journal on Addictions*, *22*(2), 93–98. https://doi.org/10.1111/j.1521-0391.2013.00304.x

Levey, D. F., Polimanti, R., Cheng, Z., Zhou, H., Nuñez, Y. Z., Jain, S., He, F., Sun, X., Ursano, R. J.,

Kessler, R. C., Smoller, J. W., Stein, M. B., Kranzler, H. R., & Gelernter, J. (2019). Genetic associations with suicide attempt severity and genetic overlap with major depression. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-018-0340-2

Levin, E. D., Hawkey, A. B., Hall, B. J., Cauley, M., Slade, S., Yazdani, E., Kenou, B., White, H., Wells, C., Rezvani, A. H., & Murphy, S. K. (2019). Paternal THC exposure in rats causes long-lasting neurobehavioral effects in the offspring. *Neurotoxicology and Teratology*, *74*, 106806. https://doi.org/10.1016/j.ntt.2019.04.003

Levola, J., Laine, R., & Pitkänen, T. (2022). In-patient psychiatric care and non-substance-related psychiatric diagnoses among individuals seeking treatment for alcohol and substance use disorders: associations with all-cause mortality and suicide. *British Journal of Psychiatry*, *221*(1), 386–393. https://doi.org/10.1192/bjp.2022.20

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Medicine*, *12*(1), 1–11. https://doi.org/10.1186/s13073-020-00742-5

Li, D., Zhao, H., Kranzler, H. R., Li, M. D., Jensen, K. P., Zayats, T., Farrer, L. A., & Gelernter, J. (2015). Genome-wide association study of copy number variations (CNVs) with opioid dependence. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *40*(4), 1016–1026. https://doi.org/10.1038/npp.2014.290

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, *95*(3), 221–227. https://doi.org/10.1038/sj.hdy.6800717

Lim, K. X., Rijsdijk, F., Hagenaars, S. P., Socrates, A., Choi, S. W., Coleman, J. R. I., Glanville, K. P., Lewis, C. M., & Pingault, J. B. (2020). Studying individual risk factors for self-harm in the UK Biobank: A polygenic scoring and Mendelian randomisation study. *PLoS Medicine*, *17*(6). https://doi.org/10.1371/journal.pmed.1003137

Lipsky, S., Krupski, A., Roy-Byrne, P., Lucenko, B., Mancuso, D., & Huber, A. (2010). Effect of co-occurring disorders and intimate partner violence on substance abuse treatment outcomes. *Journal of Substance Abuse Treatment*, *38*(3), 231–244. https://doi.org/10.1016/j.jsat.2009.12.005

Liu, H. (2019). Genetic architecture of socioeconomic outcomes: Educational attainment, occupational status, and wealth. *Social Science Research*, *82*, 137–147. https://doi.org/10.1016/j.ssresearch.2019.04.008

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Agee, M., Alipanahi, B., Auton, A., Bell, R. K., Bryc, K., Elson, S. L., Fontanillas, P., Furlotte, N. A., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Livne, O., Shmulewitz, D., Stohl, M., Mannes, Z., Aharonovich, E., & Hasin, D. (2021). Agreement between DSM-5 and DSM-IV measures of substance use disorders in a sample of adult substance users. *Drug and Alcohol Dependence*, *227*, 108958. https://doi.org/10.1016/j.drugalcdep.2021.108958

Lochner, C., Fineberg, N. A., Zohar, J., Van Ameringen, M., Juven-Wetzler, A., Altamura, A. C., Cuzen, N. L., Hollander, E., Denys, D., Nicolini, H., Dell'Osso, B., Pallanti, S., & Stein, D. J. (2014). Comorbidity in obsessive-compulsive disorder (OCD): A report from the International

College of Obsessive-Compulsive Spectrum Disorders (ICOCS). *Comprehensive Psychiatry*, *55*(7), 1513–1519. https://doi.org/10.1016/j.comppsych.2014.05.020

Lopez-Leon, S., González-Giraldo, Y., Wegman-Ostrosky, T., & Forero, D. A. (2021). Molecular genetics of substance use disorders: An umbrella review. *Neuroscience and Biobehavioral Reviews*, *124*(January), 358–369. https://doi.org/10.1016/j.neubiorev.2021.01.019

Loughnan, R. J., Palmer, C. E., Makowski, C., Thompson, W. K., Barch, D. M., Jernigan, T. L., Dale, A. M., & Fan, C. C. (2022). Unique prediction of developmental psychopathology from genetic and familial risk. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *63*(12), 1631–1643. https://doi.org/10.1111/jcpp.13649

Lu, H., Qiao, J., Shao, Z., Wang, T., Huang, S., & Zeng, P. (2021). A comprehensive gene-centric pleiotropic association analysis for 14 psychiatric disorders with GWAS summary statistics. *BMC Medicine*, *19*(1), 1–17. https://doi.org/10.1186/s12916-021-02186-z

Lynch, F. L., Peterson, E. L., Lu, C. Y., Hu, Y., Rossom, R. C., Waitzfelder, B. E., Owen-Smith, A. A., Hubley, S., Prabhakar, D., Keoki Williams, L., Beck, A., Simon, G. E., & Ahmedani, B. K. (2020). Substance use disorders and risk of suicide in a general US population: A case control study. *Addiction Science and Clinical Practice*, *15*(1), 14. https://doi.org/10.1186/s13722-020-0181-1

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Marees, A. T., Smit, D. J. A., Ong, J. S., Macgregor, S., An, J., Denys, D., Vorspan, F., Van Den Brink, W., & Derks, E. M. (2020). Potential influence of socioeconomic status on genetic correlations between alcohol consumption measures and mental health. *Psychological Medicine*, *50*(3), 484–498. https://doi.org/10.1017/S0033291719000357

Martins, S. S., Kim, J. H., Chen, L. Y., Levin, D., Keyes, K. M., Cerdá, M., & Storr, C. L. (2015). Nonmedical prescription drug use among US young adults by educational attainment. *Social Psychiatry and Psychiatric Epidemiology*, *50*(5), 713–724. https://doi.org/10.1007/s00127-014-0980-3

Martz, M. E., Heitzeg, M. M., Lisdahl, K. M., Cloak, C. C., Ewing, S. W. F., Gonzalez, R., Haist, F., LeBlanc, K. H., Madden, P. A., Ross, J. M., Sher, K. J., Tapert, S. F., Thompson, W. K., & Wade, N. E. (2022). Individual-, peer-, and parent-level substance use-related factors among 9- and 10-year-olds from the ABCD Study: Prevalence rates and sociodemographic differences. *Drug and Alcohol Dependence Reports*, *3*, 100037. https://doi.org/10.1016/j.dadr.2022.100037

Mayer, W. (1983). Alcohol abuse and alcoholism: The psychologist's role in prevention, research, and treatment. *American Psychologist*, *38*(10), 1116–1121. https://doi.org/10.1037/0003-066X.38.10.1116

Mazzeo, F., & Meccariello, R. (2023). Cannabis and Paternal Epigenetic Inheritance. *International Journal of Environmental Research and Public Health*, *20*(9). https://doi.org/10.3390/ijerph20095663

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen,

S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. https://doi.org/10.1038/ng.3643

McGue, M., Iacono, W. G., Legrand, L. N., Malone, S., & Elkins, I. (2001). Origins and consequences of age at first drink. I. Associations with substance-use disorders, disinhibitory behavior and psychopathology, and P3 amplitude. *Alcoholism: Clinical and Experimental Research*, *25*(8), 1156–1165. https://doi.org/10.1111/j.1530-0277.2001.tb02330.x

McHugh, R. K., Votaw, V. R., Sugarman, D. E., & Greenfield, S. F. (2018). Sex and gender differences in substance use disorders. *Clinical Psychology Review*, *66*, 12–23. https://doi.org/10.1016/j.cpr.2017.10.012

McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness implications for treatment, insurance, and outcomes evaluation. *Jama*, *284*(13), 1689–1695. https://doi.org/10.1001/jama.284.13.1689

Melchior, M., Chollet, A., Elidemir, G., Galéra, C., & Younès, N. (2015). Unemployment and substance use in young adults: Does educational attainment modify the association? *European Addiction Research*, *21*(3), 115–123. https://doi.org/10.1159/000365887

Melis, M., Camarini, R., Ungless, M. A., & Bonci, A. (2002). Long-lasting potentiation of GABAergic synapses in dopamine neurons after a single in vivo ethanol exposure. *Journal of Neuroscience*, *22*(6), 2074–2082. https://doi.org/10.1523/jneurosci.22-06-02074.2002

Mericle, A. A., Slaymaker, V., Gliske, K., Ngo, Q., & Subbaraman, M. S. (2022). The role of recovery housing during outpatient substance use treatment. *Journal of Substance Abuse Treatment*, *133*, 108638. https://doi.org/10.1016/j.jsat.2021.108638

Merikangas, K. R., Stolar, M., Stevens, D. E., Goulet, J., Preisig, M. A., Fenton, B., Zhang, H., O'Malley, S. S., & Rounsaville, B. J. (1998). Familial transmission of substance use disorders. *Archives of General Psychiatry*, *55*(11), 973–979. https://doi.org/10.1001/archpsyc.55.11.973

Mertens, J. R., Kline-Simon, A. H., Delucchi, K. L., Moore, C., & Weisner, C. M. (2012). Ten-year stability of remission in private alcohol and drug outpatient treatment: Non-problem users versus abstainers. *Drug and Alcohol Dependence*, *125*(1–2), 67–74. https://doi.org/10.1016/j.drugalcdep.2012.03.020

Meyers, J. L., Cerdá, M., Galea, S., Keyes, K. M., Aiello, A. E., Uddin, M., Wildman, D. E., & Koenen, K. C. (2013). Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Translational Psychiatry*, 3. https://doi.org/10.1038/tp.2013.63

Meyers, J. L., Salvatore, J. E., Aliev, F., Johnson, E. C., McCutcheon, V. V., Su, J., Kuo, S. I. C., Lai, D., Wetherill, L., Wang, J. C., Chan, G., Hesselbrock, V., Foroud, T., Bucholz, K. K., Edenberg, H. J., Dick, D. M., Porjesz, B., & Agrawal, A. (2019). Psychosocial moderation of polygenic risk for cannabis involvement: the role of trauma exposure and frequency of religious service attendance. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0598-z

Milaniak, I., Watson, B., & Jaffee, S. R. (2015). Gene-Environment Interplay and Substance Use: A Review of Recent Findings. *Current Addiction Reports*, *2*(4), 364–371. https://doi.org/10.1007/s40429-015-0069-4

Mills, K. L., Teesson, M., Ross, J., & Peters, L. (2006). Trauma, PTSD, and substance use disorders: Findings from the Australian National Survey of Mental Health and Well-Being. *American Journal of Psychiatry*, *163*(4), 652–658. https://doi.org/10.1176/ajp.2006.163.4.652

Minică, C. C., Verweij, K. J. H., van der Most, P. J., Mbarek, H., Bernard, M., van Eijk, K. R., Lind, P. A., Liu, M. Z., Maciejewski, D. F., Palviainen, T., Sánchez-Mora, C., Sherva, R., Taylor, M., Walters, R. K., Abdellaoui, A., Bigdeli, T. B., Branje, S. J. T., Brown, S. A., Casas, M., … Derks, E. M. (2018). Genome-wide association meta-analysis of age at first cannabis use. *Addiction*, *113*(11), 2073–2086. https://doi.org/10.1111/add.14368

Mitjans, M., Papiol, S., Barrot, C., Guardiola-Ripoll, M., Giménez-Palomo, A., Gavín, P., Oraki Kohshour, M., Acosta, M., Pomarol-Clotet, E., Castellà-García, J., Pérez-Pérez, R. M., Gutiérrez, B., Schulze, T. G., Fatjó-Vilas, M., Arias, B., & Benabarre, A. (2022). Completed suicide is associated with a higher polygenic burden for psychiatric disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *272*(3), 355–358. https://doi.org/10.1007/s00406-022-01398-5

Moehring, A., Rumpf, H. J., Hapke, U., Bischof, G., John, U., & Meyer, C. (2019). Diagnostic performance of the Alcohol Use Disorders Identification Test (AUDIT) in detecting DSM-5 alcohol use disorders in the General population. *Drug and Alcohol Dependence*, *204*, 107530. https://doi.org/10.1016/j.drugalcdep.2019.06.032

Molina, B. S. G., Howard, A. L., Swanson, J. M., Stehli, A., Mitchell, J. T., Kennedy, T. M., Epstein, J. N., Arnold, L. E., Hechtman, L., Vitiello, B., & Hoza, B. (2018). Substance use through adolescence into early adulthood after childhood-diagnosed ADHD: findings from the MTA longitudinal study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(6), 692–702. https://doi.org/10.1111/jcpp.12855

Molinero, K., & Hinckley, J. D. (2023). Adolescent Cannabis Use, Comorbid Attention-Deficit/Hyperactivity Disorder, and Other Internalizing and Externalizing Disorders. *Child and Adolescent Psychiatric Clinics of North America*, *32*(1), 57–68. https://doi.org/10.1016/j.chc.2022.07.003

Morley, K. I., Lynskey, M. T., Moran, P., Borschmann, R., & Winstock, A. R. (2015). Polysubstance use, mental health and high-risk behaviours: Results from the 2012 Global Drug Survey. *Drug and Alcohol Review*, *34*(4), 427–437. https://doi.org/10.1111/dar.12263

Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M., & He, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics*, *52*(7), 740–747. https://doi.org/10.1038/s41588-020-0631-4

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., Bryois, J., Charney, A. W., Drange, O. K., Gandal, M. J., Hagenaars, S. P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., … Andreassen, O. A. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*, *53*(6), 817–829. https://doi.org/10.1038/s41588-021-00857-4

Mullins, N., Perroud, N., Uher, R., Butler, A. W., Cohen-Woods, S., Rivera, M., Malki, K., Euesden, J., Power, R. A., Tansey, K. E., Jones, L., Jones, I., Craddock, N., Owen, M. J., Korszun, A., Gill, M., Mors, O., Preisig, M., Maier, W., … Lewis, C. M. (2014). Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: A genome-wide association and polygenic scoring study. *American Journal of Medical Genetics, Part B: Neuropsychiatric*

Genetics, *165*(5), 428–437. https://doi.org/10.1002/ajmg.b.32247

Munafò, M. R., Higgins, J. P. T., & Smith, G. D. (2021). Triangulating Evidence through the Inclusion of Genetically Informed Designs. *Cold Spring Harbor Perspectives in Medicine*, *11*(8). https://doi.org/10.1101/cshperspect.a040659

Murphy, S. K., Itchon-Ramos, N., Visco, Z., Huang, Z., Grenier, C., Schrott, R., Acharya, K., Boudreau, M. H., Price, T. M., Raburn, D. J., Corcoran, D. L., Lucas, J. E., Mitchell, J. T., McClernon, F. J., Cauley, M., Hall, B. J., Levin, E. D., & Kollins, S. H. (2018). Cannabinoid exposure and altered DNA methylation in rat and human sperm. *Epigenetics*, *13*(12), 1208–1221. https://doi.org/10.1080/15592294.2018.1554521

Murray, G. K., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., & Wray, N. R. (2021). Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. *JAMA Psychiatry*, *78*(2), 210–219. https://doi.org/10.1001/jamapsychiatry.2020.3042

Musci, R. J., Uhl, G., Maher, B., & Ialongo, N. S. (2015). Testing gene × environment moderation of tobacco and marijuana use trajectories in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology*, *83*(5), 866–874. https://doi.org/10.1037/a0039537

Musliner, K. L., Krebs, M. D., Albiñana, C., Vilhjalmsson, B., Agerbo, E., Zandi, P. P., Hougaard, D. M., Nordentoft, M., Børglum, A. D., Werge, T., Mortensen, P. B., & Østergaard, S. D. (2020). Polygenic risk and progression to bipolar or psychotic disorders among individuals diagnosed with unipolar depression in early life. *American Journal of Psychiatry*, *177*(10), 936–943. https://doi.org/10.1176/appi.ajp.2020.19111195

Musliner, K. L., Mortensen, P. B., McGrath, J. J., Suppli, N. P., Hougaard, D. M., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Andreassen, O., Pedersen, C. B., Pedersen, M. G., Mors, O., Nordentoft, M., Børglum, A. D., Werge, T., & Agerbo, E. (2019). Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia with Risk for Depression in the Danish Population. *JAMA Psychiatry*, *76*(5), 516–525. https://doi.org/10.1001/jamapsychiatry.2018.4166

Na, P. J., Stefanovics, E. A., Rhee, T. G., & Rosenheck, R. A. (2022). A population-wide perspective on the reach of substance use disorders: Parental exposure, diagnostic remission, and current disorders. *American Journal on Addictions*, *31*(1), 69–79. https://doi.org/10.1111/ajad.13246

Nathan, P. E., Conrad, M., & Skinstad, A. H. (2016). History of the Concept of Addiction. *Annual Review of Clinical Psychology*, *12*, 29–51. https://doi.org/10.1146/annurev-clinpsy-021815-093546

Nelson, E. C., Agrawal, A., Heath, A. C., Bogdan, R., Sherva, R., Zhang, B., Al-Hasani, R., Bruchas, M. R., Chou, Y. L., Demers, C. H., Carey, C. E., Conley, E. D., Fakira, A. K., Farrer, L. A., Goate, A., Gordon, S., Henders, A. K., Hesselbrock, V., Kapoor, M., … Montgomery, G. W. (2016). Evidence of CNIH3 involvement in opioid dependence. *Molecular Psychiatry*, *21*(5), 608–614. https://doi.org/10.1038/mp.2015.102

Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The Heterogeneity of Mental Health Assessment. *Frontiers in Psychiatry*, *11*, 76. https://doi.org/10.3389/fpsyt.2020.00076

Nielsen, D. A., & Kreek, M. J. (2012). Common and specific liability to addiction: Approaches to association studies of opioid addiction. *Drug and Alcohol Dependence*, *123*(SUPPL.1).

https://doi.org/10.1016/j.drugalcdep.2012.03.026

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective and Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Nievergelt, C. M., Maihofer, A. X., Klengel, T., Atkinson, E. G., Chen, C. Y., Choi, K. W., Coleman, J. R. I., Dalvie, S., Duncan, L. E., Gelernter, J., Levey, D. F., Logue, M. W., Polimanti, R., Provost, A. C., Ratanatharathorn, A., Stein, M. B., Torres, K., Aiello, A. E., Almli, L. M., … Koenen, K. C. (2019). International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12576-w

Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., & Matsui, S. (2018). Sample size for successful genome-wide association study of major depressive disorder. *Frontiers in Genetics*, *9*, 227. https://doi.org/10.3389/fgene.2018.00227

Nobile, B., Ramoz, N., Jaussent, I., Gorwood, P., Olié, E., Castroman, J. L., Guillaume, S., & Courtet, P. (2019). Polymorphism A118G of opioid receptor mu 1 (OPRM1) is associated with emergence of suicidal ideation at antidepressant onset in a large naturalistic cohort of depressed outpatients. *Scientific Reports*, *9*(1), 2569. https://doi.org/10.1038/s41598-019-39622-3

Nordsletten, A. E., Larsson, H., Crowley, J. J., Almqvist, C., Lichtenstein, P., & Mataix-Cols, D. (2016). Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry*, *73*(4), 354–361. https://doi.org/10.1001/jamapsychiatry.2015.3192

Nurnberger, J. I., Wang, Y., Zang, Y., Lai, D., Wetherill, L., Edenberg, H. J., Aliev, F., Plawecki, M. H., Chorlian, D., Chan, G., Bucholz, K., Bauer, L., Kamarajan, C., Salvatore, J. E., Kapoor, M., Hesselbrock, V., Dick, D., Bierut, L., McCutcheon, V., … Parsian, A. (2022). High Polygenic Risk Scores Are Associated With Early Age of Onset of Alcohol Use Disorder in Adolescents and Young Adults at Risk. *Biological Psychiatry Global Open Science*, *2*(4), 379–388. https://doi.org/10.1016/j.bpsgos.2021.10.007

Nyholt, D. R. (2004). A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. *American Journal of Human Genetics*, *74*(4), 765–769. https://doi.org/10.1086/383251

O'Brien, B. S., & Sher, L. (2013). Child sexual abuse and the pathophysiology of suicide in adolescents and adults. In *International Journal of Adolescent Medicine and Health* (Vol. 25, Issue 3, pp. 201–205). Int J Adolesc Med Health. https://doi.org/10.1515/ijamh-2013-0053

O'Brien, C. P. (2011). Chapter 24. Drug Addiction. *Goodman & Gilman's: The Pharmacological Basis of Therapeutics*, 1–29. http://mhmedical.com/content.aspx?aid=16666364

O'Connor, L. J., & Price, A. L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature Genetics*, *50*(12), 1728–1734. https://doi.org/10.1038/s41588-018-0255-0

O'Neil, K. A., Conner, B. T., & Kendall, P. C. (2011). Internalizing disorders and substance use disorders in youth: Comorbidity, risk, temporal order, and implications for intervention. *Clinical Psychology Review*, *31*(1), 104–112. https://doi.org/10.1016/j.cpr.2010.08.002

Oliva, F., Mangiapane, C., Nibbio, G., Berchialla, P., Colombi, N., & Vigna-Taglianti, F. D. (2021). Prevalence of cocaine use and cocaine use disorder among adult patients with attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Journal of Psychiatric Research*, *143*, 587–598. https://doi.org/10.1016/j.jpsychires.2020.11.021

Olsen, Y. (2022). What Is Addiction? History, Terminology, and Core Concepts. *Medical Clinics of North America*, *106*(1), 1–12. https://doi.org/10.1016/j.mcna.2021.08.001

Onaemo, V. N., Fawehinmi, T. O., & D'Arcy, C. (2022). Risk of suicide ideation in comorbid substance use disorder and major depression. *PLoS ONE*, *17*(12). https://doi.org/10.1371/journal.pone.0265287

Ornell, F., Halpern, S., & Diemen, L. (2021). Substance use disorders and psychiatric comorbidity. In A. L. M. Andrade, D. De Micheli, E. A. da Silva, F. M. Lopes, B. de O. Pinheiro, & R. A. Reichert (Eds.), *Psychology of Substance Abuse: Psychotherapy, Clinical Management and Social Intervention* (pp. 23–41). Springer International Publishing. https://doi.org/10.1007/978-3-030-62106-3_2

Ostacher, M. J. (2007). Comorbid Alcohol and Substance Abuse Dependence in Depression: Impact on the Outcome of Antidepressant Treatment. In *Psychiatric Clinics of North America* (Vol. 30, Issue 1, pp. 69–76). Elsevier. https://doi.org/10.1016/j.psc.2006.12.009

Østergaard, M. L. D., Nordentoft, M., & Hjorthøj, C. (2017). Associations between substance use disorders and suicide or suicide attempts in people with mental illness: a Danish nation-wide, prospective, register-based study of patients diagnosed with schizophrenia, bipolar disorder, unipolar depression or personal. *Addiction*, *112*(7), 1250–1259. https://doi.org/10.1111/add.13788

Pain, O., Glanville, K. P., Hagenaars, S. P., Selzam, S., Furtjes, A. E., Gaspar, H. A., Coleman, J. R. I., Rimfeld, K., Breen, G., Plomin, R., Folkersen, L., & Lewis, C. M. (2021). Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genetics*, *17*(5 May), 1–22. https://doi.org/10.1371/journal.pgen.1009021

Palk, A. C., Dalvie, S., De Vries, J., Martin, A. R., & Stein, D. J. (2019). Potential use of clinical polygenic risk scores in psychiatry - Ethical implications and communicating high polygenic risk. In *Philosophy, Ethics, and Humanities in Medicine* (Vol. 14, Issue 1). https://doi.org/10.1186/s13010-019-0073-8

Palmer, R. H. C., Brick, L., Nugent, N. R., Bidwell, L. C., Mcgeary, J. E., Knopik, V. S., & Keller, M. C. (2015). Examining the role of common genetic variants on alcohol, tobacco, cannabis and illicit drug dependence: Genetics of vulnerability to drug dependence. *Addiction*, *110*(3), 530–537. https://doi.org/10.1111/add.12815

Palmer, R. H. C., Button, T. M., Rhee, S. H., Corley, R. P., Young, S. E., Stallings, M. C., Hopfer, C. J., & Hewitt, J. K. (2012). Genetic etiology of the common liability to drug dependence: Evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug and Alcohol Dependence*, *123*(SUPPL.1), S24-32. https://doi.org/10.1016/j.drugalcdep.2011.12.015

Park, Y. M., Shekhtman, T., & Kelsoe, J. R. (2020). Interaction between adverse childhood experiences and polygenic risk in patients with bipolar disorder. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-01010-1

Pasman, J. A., Verweij, K. J. H., Abdellaoui, A., Hottenga, J. J., Fedko, I. O., Willemsen, G., Boomsma,

D. I., & Vink, J. M. (2020). Substance use: Interplay between polygenic risk and neighborhood environment. *Drug and Alcohol Dependence*, *209*. https://doi.org/10.1016/j.drugalcdep.2020.107948

Pasman, J. A., Verweij, K. J. H., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J. L., Abdellaoui, A., Nivard, M. G., Baselmans, B. M. L., Ong, J. S., Ip, H. F., van der Zee, M. D., Bartels, M., Day, F. R., Fontanillas, P., Elson, S. L., de Wit, H., Davis, L. K., MacKillop, J., … Vink, J. M. (2018). GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature Neuroscience*, *21*(9), 1161–1170. https://doi.org/10.1038/s41593-018-0206-1

Pasman, J. A., Verweij, K. J. H., & Vink, J. M. (2019). Systematic Review of Polygenic Gene–Environment Interaction in Tobacco, Alcohol, and Cannabis Use. In *Behavior Genetics* (Vol. 49, Issue 4, pp. 349–365). https://doi.org/10.1007/s10519-019-09958-7

Paul, S. E., Hatoum, A. S., Fine, J. D., Johnson, E. C., Hansen, I., Karcher, N. R., Moreau, A. L., Bondy, E., Qu, Y., Carter, E. B., Rogers, C. E., Agrawal, A., Barch, D. M., & Bogdan, R. (2021). Associations between Prenatal Cannabis Exposure and Childhood Outcomes: Results from the ABCD Study. *JAMA Psychiatry*, *78*(1), 64–76. https://doi.org/10.1001/jamapsychiatry.2020.2902

Pietrzak, R. H., Goldstein, M. B., Malley, J. C., Johnson, D. C., & Southwick, S. M. (2009). Subsyndromal posttraumatic stress disorder is associated with health and psychosocial difficulties in veterans of Operations Enduring Freedom and Iraqi Freedom. *Depression and Anxiety*, *26*(8), 739–744. https://doi.org/10.1002/da.20574

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Prevalence and Axis I comorbidity of full and partial posttraumatic stress disorder in the United States: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Anxiety Disorders*, *25*(3), 456–465. https://doi.org/10.1016/j.janxdis.2010.11.010

Pingault, J. B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*(9), 566–580. https://doi.org/10.1038/s41576-018-0020-3

Polimanti, R., Agrawal, A., & Gelernter, J. (2017). Schizophrenia and substance use comorbidity: A genome-wide perspective. *Genome Medicine*, *9*(1), 9–11. https://doi.org/10.1186/s13073-017-0423-3

Polimanti, R., Kaufman, J., Zhao, H., Kranzler, H. R., Ursano, R. J., Kessler, R. C., Stein, M. B., & Gelernter, J. (2018). Trauma exposure interacts with the genetic risk of bipolar disorder in alcohol misuse of US soldiers. *Acta Psychiatrica Scandinavica*, *137*(2), 148–156. https://doi.org/10.1111/acps.12843

Polimanti, R., Levey, D. F., Pathak, G. A., Wendt, F. R., Nunez, Y. Z., Ursano, R. J., Kessler, R. C., Kranzler, H. R., Stein, M. B., & Gelernter, J. (2021). Multi-environment gene interactions linked to the interplay between polysubstance dependence and suicidality. *Translational Psychiatry*, *11*(1), 1–11. https://doi.org/10.1038/s41398-020-01153-1

Polimanti, R., Peterson, R. E., Ong, J. S., MacGregor, S., Edwards, A. C., Clarke, T. K., Frank, J., Gerring, Z., Gillespie, N. A., Lind, P. A., Maes, H. H., Martin, N. G., Mbarek, H., Medland, S. E., Streit, F., Agrawal, A., Edenberg, H. J., Kendler, K. S., Lewis, C. M., … Derks, E. M. (2019). Evidence of causal effect of major depression on alcohol dependence: Findings from the psychiatric

genomics consortium. *Psychological Medicine*, *49*(7), 1218–1226. https://doi.org/10.1017/S0033291719000667

Polimanti, R., Walters, R. K., Johnson, E. C., McClintick, J. N., Adkins, A. E., Adkins, D. E., Bacanu, S. A., Bierut, L. J., Bigdeli, T. B., Brown, S., Bucholz, K. K., Copeland, W. E., Costello, E. J., Degenhardt, L., Farrer, L. A., Foroud, T. M., Fox, L., Goate, A. M., Grucza, R., … Gelernter, J. (2020). Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. *Molecular Psychiatry*, *25*(8), 1673–1687. https://doi.org/10.1038/s41380-020-0677-9

Poudel, A., & Gautam, S. (2017). Age of onset of substance use and psychosocial problems among individuals with substance use disorders. *BMC Psychiatry*, *17*(1). https://doi.org/10.1186/s12888-016-1191-0

Privé, F., Vilhjálmsson, B. J., Aschard, H., & Blum, M. G. B. (2019). Making the Most of Clumping and Thresholding for Polygenic Scores. *American Journal of Human Genetics*, *105*(6), 1213–1221. https://doi.org/10.1016/j.ajhg.2019.11.001

Prom-Wormley, E. C., Ebejer, J., Dick, D. M., & Bowers, M. S. (2017). The genetic epidemiology of substance use disorder: A review. In *Drug and Alcohol Dependence* (Vol. 180, pp. 241–259). NIH Public Access. https://doi.org/10.1016/j.drugalcdep.2017.06.040

Quach, B. C., Bray, M. J., Gaddis, N. C., Liu, M., Palviainen, T., Minica, C. C., Zellers, S., Sherva, R., Aliev, F., Nothnagel, M., Young, K. A., Marks, J. A., Young, H., Carnes, M. U., Guo, Y., Waldrop, A., Sey, N. Y. A., Landi, M. T., McNeil, D. W., … Hancock, D. B. (2020). Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19265-z

Quinn, P. D., Chang, Z., Hur, K., Gibbons, R. D., Lahey, B. B., Rickert, M. E., Sjölander, A., Lichtenstein, P., Larsson, H., & D'Onofrio, B. M. (2017). ADHD medication and substance-related problems. *American Journal of Psychiatry*, *174*(9), 877–885. https://doi.org/10.1176/appi.ajp.2017.16060686

Rajagopal, V. M., Watanabe, K., Mbatchou, J., Ayer, A., Quon, P., Sharma, D., Kessler, M. D., Praveen, K., Gelfman, S., Parikshak, N., Otto, J. M., Bao, S., Chim, S. M., Pavlopoulos, E., Avbersek, A., Kapoor, M., Chen, E., Jones, M. B., Leblanc, M., … Coppola, G. (2023). Rare coding variants in CHRNB2 reduce the likelihood of smoking. *Nature Genetics*, *55*(7), 1138–1148. https://doi.org/10.1038/s41588-023-01417-8

Ramos-Quiroga, J. A., Bosch, R., Richarte, V., Valero, S., Gómez-Barros, N., Nogueira, M., Palomar, G., Corrales, M., Sáez-Francàs, N., Corominas, M., Real, A., Vidal, R., Chalita, P. J., & Casas, M. (2012). Criterion and concurrent validity of Conners Adult ADHD Diagnostic Interview for DSM-IV (CAADID) Spanish version. *Revista de Psiquiatría y Salud Mental (English Edition)*, *5*(4), 229–235. https://doi.org/10.1016/j.rpsmen.2012.05.002

Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. P., & Foulkes, A. S. (2015). A guide to genome-wide association analysis and post-analytic interrogation. *Statistics in Medicine*, *34*(28), 3769–3792. https://doi.org/10.1002/sim.6605

Regnart, J., Truter, I., & Meyer, A. (2017). Critical exploration of co-occurring Attention-Deficit/Hyperactivity Disorder, mood disorder and Substance Use Disorder. *Expert Review of Pharmacoeconomics and Outcomes Research*, *17*(3), 275–282. https://doi.org/10.1080/14737167.2017.1351878

Rehm, J., & Shield, K. D. (2019). Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Current Psychiatry Reports*, *21*(2), 10. https://doi.org/10.1007/s11920-019-0997-0

Rial Boubeta, A., Golpe, S., Barreiro, C., Gómez, P., & Isorna, M. (2020). The age of onset for alcohol consumption among adolescents: Implications and related variables. *Adicciones*, *32*(1), 52–62. https://doi.org/10.20882/adicciones.1266

Richmond-Rakerd, L. S., Slutske, W. S., & Wood, P. K. (2017). Age of initiation and substance use progression: A multivariate latent growth analysis. *Psychology of Addictive Behaviors*, *31*(6), 664–675. https://doi.org/10.1037/adb0000304

Rioux, C., Castellanos-Ryan, N., Parent, S., Vitaro, F., Tremblay, R. E., & Séguin, J. R. (2018). Age of Cannabis Use Onset and Adult Drug Abuse Symptoms: A Prospective Study of Common Risk Factors and Indirect Effects. *Canadian Journal of Psychiatry*, *63*(7), 457–464. https://doi.org/10.1177/0706743718760289

Roehrs, T., Sibai, M., & Roth, T. (2021). Sleep and alertness disturbance and substance use disorders: A bi-directional relation. In *Pharmacology Biochemistry and Behavior* (Vol. 203). https://doi.org/10.1016/j.pbb.2021.173153

Rohner, H., Gaspar, N., Philipsen, A., & Schulze, M. (2023). Prevalence of Attention Deficit Hyperactivity Disorder (ADHD) among Substance Use Disorder (SUD) Populations: Meta-Analysis. *International Journal of Environmental Research and Public Health*, *20*(2). https://doi.org/10.3390/ijerph20021275

Romero, C., Werme, J., Jansen, P. R., Gelernter, J., Stein, M. B., Levey, D., Polimanti, R., de Leeuw, C., Posthuma, D., Nagel, M., & van der Sluis, S. (2022). Exploring the genetic overlap between twelve psychiatric disorders. *Nature Genetics*, *54*(12), 1795–1802. https://doi.org/10.1038/s41588-022-01245-2

Rosoff, D. B., Smith, G. D., & Lohoff, F. W. (2021). Prescription Opioid Use and Risk for Major Depressive Disorder and Anxiety and Stress-Related Disorders: A Multivariable Mendelian Randomization Analysis. *JAMA Psychiatry*, *78*(2), 151–160. https://doi.org/10.1001/jamapsychiatry.2020.3554

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, H. T., Abebe, M., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., Abraha, H. N., … Murray, C. J. L. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1736–1788. https://doi.org/10.1016/S0140-6736(18)32203-7

Ruan, Y., Lin, Y. F., Feng, Y. C. A., Chen, C. Y., Lam, M., Guo, Z., Ahn, Y. M., Akiyama, K., Arai, M., Baek, J. H., Chen, W. J., Chung, Y. C., Feng, G., Fujii, K., Glatt, S. J., Ha, K., Hattori, K., Higuchi, T., Hishimoto, A., … Ge, T. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*, *54*(5), 573–580. https://doi.org/10.1038/s41588-022-01054-7

Ruderfer, D. M., Walsh, C. G., Aguirre, M. W., Tanigawa, Y., Ribeiro, J. D., Franklin, J. C., & Rivas, M. A. (2020). Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Molecular Psychiatry*, *25*(10), 2422–2430. https://doi.org/10.1038/s41380-018-0326-8

Sallis, H. M., Davey Smith, G., & Munafò, M. R. (2019). Cigarette smoking and personality: interrogating causality using Mendelian randomisation. *Psychological Medicine*, *49*(13), 2197–2205. https://doi.org/10.1017/S0033291718003069

Salvatore, J. E., Aliev, F., Edwards, A. C., Evans, D. M., Macleod, J., Hickman, M., Lewis, G., Kendler, K. S., Loukola, A., Korhonen, T., Latvala, A., Rose, R. J., Kaprio, J., & Dick, D. M. (2014). Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes*, *5*(2), 330–346. https://doi.org/10.3390/genes5020330

Salvatore, J. E., Barr, P. B., Stephenson, M., Aliev, F., Kuo, S. I. C., Su, J., Agrawal, A., Almasy, L., Bierut, L., Bucholz, K., Chan, G., Edenberg, H. J., Johnson, E. C., McCutcheon, V. V., Meyers, J. L., Schuckit, M., Tischfield, J., Wetherill, L., & Dick, D. M. (2020). Sibling comparisons elucidate the associations between educational attainment polygenic scores and alcohol, nicotine and cannabis. *Addiction*, *115*(2), 337–346. https://doi.org/10.1111/add.14815

Sanchez-Roige, S., & Palmer, A. A. (2020). Emerging phenotyping strategies will advance our understanding of psychiatric genetics. *Nature Neuroscience*, *23*(4), 475–480. https://doi.org/10.1038/s41593-020-0609-7

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., Adams, M. J., Howard, D. M., Edenberg, H. J., Davies, G., Crist, R. C., Deary, I. J., McIntosh, A. M., Clarke, T. K., Elson, L., Fontanillas, P., Furlotte, N. A., Hinds, D. A., Huber, K. E., Kleinman, A., Litterman, N. K., … Wilson, C. H. (2019). Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry*, *176*(2), 107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., & Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 6. https://doi.org/10.1038/s43586-021-00092-5

Saunders, G. R. B., Wang, X., Chen, F., Jang, S. K., Liu, M., Wang, C., Gao, S., Jiang, Y., Khunsriraksakul, C., Otto, J. M., Addison, C., Akiyama, M., Albert, C. M., Aliev, F., Alonso, A., Arnett, D. K., Ashley-Koch, A. E., Ashrani, A. A., Barnes, K. C., … Vrieze, S. (2022). Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, *612*(7941), 720–724. https://doi.org/10.1038/s41586-022-05477-4

Schoeler, T., Baldwin, J., Allegrini, A., Barkhuizen, W., McQuillin, A., Pirastu, N., Kutalik, Z., & Pingault, J. B. (2022). Novel Biological Insights Into the Common Heritable Liability to Substance Involvement: A Multivariate Genome-wide Association Study. *Biological Psychiatry*, *93*(6), 524–535. https://doi.org/10.1016/j.biopsych.2022.07.027

Schoenfelder, E. N., Faraone, S. V., & Kollins, S. H. (2014). Stimulant treatment of ADHD and cigarette smoking: A meta-analysis. *Pediatrics*, *133*(6), 1070–1080. https://doi.org/10.1542/peds.2014-0179

Schwantes-An, T. H., Zhang, J., Chen, L. S., Hartz, S. M., Culverhouse, R. C., Chen, X., Coon, H., Frank, J., Kamens, H. M., Konte, B., Kovanen, L., Latvala, A., Legrand, L. N., Maher, B. S., Melroy, W. E., Nelson, E. C., Reid, M. W., Robinson, J. D., Shen, P. H., … Saccone, N. L. (2016). Association of the OPRM1 Variant rs1799971 (A118G) with Non-Specific Liability to Substance Dependence in a Collaborative de novo Meta-Analysis of European-Ancestry Cohorts. *Behavior Genetics*, *46*(2), 151–169. https://doi.org/10.1007/s10519-015-9737-3

Sheidow, A. J., McCart, M., Zajac, K., & Davis, M. (2012). Prevalence and impact of substance use

among emerging adults with serious mental health conditions. *Psychiatric Rehabilitation Journal*, *35*(3), 235–243. https://doi.org/10.2975/35.3.2012.235.243

Sherva, R., Wang, Q., Kranzler, H., Zhao, H., Koesterer, R., Herman, A., Farrer, L. A., & Gelernter, J. (2016). Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry*, *73*(5), 472–480. https://doi.org/10.1001/jamapsychiatry.2016.0036

Silveira, P. P., Pokhvisneva, I., Howard, D. M., & Meaney, M. J. (2023). A sex-specific genome-wide association study of depression phenotypes in UK Biobank. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-023-01960-0

Slade, E. P., Stuart, E. A., Salkever, D. S., Karakus, M., Green, K. M., & Ialongo, N. (2008). Impacts of age of onset of substance use disorders on risk of adult incarceration among disadvantaged urban youth: A propensity score matching approach. *Drug and Alcohol Dependence*, *95*(1–2), 1–13. https://doi.org/10.1016/j.drugalcdep.2007.11.019

Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), 477–485. https://doi.org/10.1038/nrg2361

Smith, A., Kaufman, F., Sandy, M. S., & Cardenas, A. (2020). Cannabis Exposure During Critical Windows of Development: Epigenetic and Molecular Pathways Implicated in Neuropsychiatric Disease. *Current Environmental Health Reports*, *7*(3), 325–342. https://doi.org/10.1007/s40572-020-00275-4

Soler Artigas, M., Sánchez-Mora, C., Rovira, P., Richarte, V., Garcia-Martínez, I., Pagerols, M., Demontis, D., Stringer, S., Vink, J. M., Børglum, A. D., Neale, B. M., Franke, B., Faraone, S. V., Casas, M., Ramos-Quiroga, J. A., & Ribasés, M. (2020). Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. *Molecular Psychiatry*, *25*(10), 2493–2503. https://doi.org/10.1038/s41380-018-0339-3

Steinhausen, H. C., & Bisgaard, C. (2014). Substance use disorders in association with attention-deficit/hyperactivity disorder, co-morbid mental disorders, and medication in a nationwide sample. *European Neuropsychopharmacology*, *24*(2), 232–241. https://doi.org/10.1016/j.euroneuro.2013.11.003

Stevens, E., Jason, L. A., Ram, D., & Light, J. (2015). Investigating Social Support and Network Relationships in Substance Use Disorder Recovery. *Substance Abuse*, *36*(4), 396–399. https://doi.org/10.1080/08897077.2014.965870

Su, J., Kuo, S. I. C., Aliev, F., Chan, G., Edenberg, H. J., Kamarajan, C., McCutcheon, V. V., Meyers, J. L., Schuckit, M., Tischfield, J., & Dick, D. M. (2021). The Associations Between Polygenic Risk, Sensation Seeking, Social Support, and Alcohol Use in Adulthood. *Journal of Abnormal Psychology*, *130*(5), 525–536. https://doi.org/10.1037/abn0000568

Sulovari, A., Liu, Z., Zhu, Z., & Li, D. (2018). Genome-wide meta-analysis of copy number variations with alcohol dependence. *The Pharmacogenomics Journal*, *18*(3), 398–405. https://doi.org/10.1038/tpj.2017.35

Sun, J., Kranzler, H. R., Gelernter, J., & Bi, J. (2020). A genome-wide association study of cocaine use disorder accounting for phenotypic heterogeneity and gene–environment interaction. *Journal of Psychiatry and Neuroscience*, *45*(1), 34–44. https://doi.org/10.1503/jpn.180098

Swann, G., Byck, G. R., Dick, D. M., Aliev, F., Latendresse, S. J., Riley, B., Kertes, D., Sun, C., Salvatore, J. E., Bolland, J., & Mustanski, B. (2014). Effect of OPRM1 and stressful life events on symptoms of major depression in African American adolescents. *Journal of Affective Disorders*, *162*, 12–19. https://doi.org/10.1016/j.jad.2014.03.020

Syan, S. K., Minhas, M., Oshri, A., Costello, J., Sousa, S., Samokhvalov, A. V., Rush, B., & MacKillop, J. (2020). Predictors of premature treatment termination in a large residential addiction medicine program. *Journal of Substance Abuse Treatment*, *117*. https://doi.org/10.1016/j.jsat.2020.108077

Taqi, M. M., Faisal, M., & Zaman, H. (2019). OPRM1 A118G polymorphisms and its role in opioid addiction: Implication on severity and treatment approaches. *Pharmacogenomics and Personalized Medicine*, *12*, 361–368. https://doi.org/10.2147/PGPM.S198654

Tarter, R. E., Kirisci, L., Mezzich, A., Cornelius, J. R., Pajer, K., Vanyukov, M., Gardner, W., Blackson, T., & Clark, D. (2003). Neurobehavioral disinhibition in childhood predicts early age at onset of substance use disorder. *American Journal of Psychiatry*, *160*(6), 1078–1085. https://doi.org/10.1176/appi.ajp.160.6.1078

Tiego, J., Martin, E. A., DeYoung, C. G., Hagan, K., Cooper, S. E., Pasion, R., Satchell, L., Shackman, A. J., Bellgrove, M. A., Fornito, A., Abend, R., Goulter, N., Eaton, N. R., Kaczkurkin, A. N., & Nusslock, R. (2023). Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nature Mental Health*, *1*(5), 304–315. https://doi.org/10.1038/s44220-023-00057-5

Toftdahl, N. G., Nordentoft, M., & Hjorthøj, C. (2016). Prevalence of substance use disorders in psychiatric patients: a nationwide Danish population-based study. *Social Psychiatry and Psychiatric Epidemiology*, *51*(1), 129–140. https://doi.org/10.1007/s00127-015-1104-4

Toikumo, S., Jennings, M. V, Pham, B., Lee, H., Mallard, T. T., Bianchi, S. B., Meredith, J. J., Vilar-Ribó, L., Xu, H., Hatoum, A. S., Johnson, E. C., Pazdernik, V., Jinwala, Z., Leger, B. S., Niarchou, M., Ehinmowo, M., BioBank, P. M., Veteran, M., Jenkins, G. D., … Sanchez-Roige, S. (2023). Multi-ancestry meta-analysis of tobacco use disorders based on electronic health record data prioritizes novel candidate risk genes and reveals associations with numerous health outcomes. *MedRxiv*. https://doi.org/10.1101/2023.03.27.23287713

Torok, M., Darke, S., & Kaye, S. (2012). Attention deficit hyperactivity disorder and severity of substance use: The role of comorbid psychopathology. *Psychology of Addictive Behaviors*, *26*(4), 974–979. https://doi.org/10.1037/a0027846

Treur, J. L., Demontis, D., Smith, G. D., Sallis, H., Richardson, T. G., Wiers, R. W., Børglum, A. D., Verweij, K. J. H., & Munafò, M. R. (2021). Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. *Addiction Biology*, *26*(1), e12849. https://doi.org/10.1111/adb.12849

Treur, J. L., Willemsen, G., Bartels, M., Geels, L. M., Van Beek, J. H. D. A., Huppertz, C., Van Beijsterveldt, C. E. M., Boomsma, D. I., & Vink, J. M. (2015). Smoking during adolescence as a risk factor for attention problems. *Biological Psychiatry*, *78*(9), 656–663. https://doi.org/10.1016/j.biopsych.2014.06.019

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C.-Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., … O'Donovan, M. C. (2022).

Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, *604*(7906), 502–508. https://doi.org/10.1038/s41586-022-04434-5

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., … van Os, J. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, *604*(7906), 502. https://doi.org/10.1038/S41586-022-04434-5

Tsuang, M. T., Bar, J. L., Harley, R. M., & Lyons, M. J. (2001). The Harvard Twin Study of Substance Abuse: What we have learned. *Harvard Review of Psychiatry*, *9*(6), 267–279. https://doi.org/10.1080/10673220127912

Tsuang, M. T., Lyons, M. J., Meyer, J. M., Doyle, T., Eisen, S. A., Goldberg, J., True, W., Lin, N., Toomey, R., & Eaves, L. (1998). Co-occurrence of abuse of different drugs in men: The role of drug-specific and shared vulnerabilities. *Archives of General Psychiatry*, *55*(11), 967–972. https://doi.org/10.1001/archpsyc.55.11.967

Tuithof, M., ten Have, M., van den Brink, W., Vollebergh, W., & de Graaf, R. (2013). Predicting persistency of DSM-5 alcohol use disorder and examining drinking patterns of recently remitted individuals: A prospective general population study. *Addiction*, *108*(12), 2091–2099. https://doi.org/10.1111/add.12309

Turner, J. K., Athamneh, L. N., Basso, J. C., & Bickel, W. K. (2021). The phenotype of recovery V: Does delay discounting predict the perceived risk of relapse among individuals in recovery from alcohol and drug use disorders. *Alcoholism: Clinical and Experimental Research*, *45*(5), 1100–1108. https://doi.org/10.1111/acer.14600

Tyrer, P. (2014). A comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment*, *20*(4), 280–285. https://doi.org/DOI: 10.1192/apt.bp.113.011296

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Uher, R., & Zwicker, A. (2017). Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry*, *16*(2), 121–129. https://doi.org/10.1002/wps.20436

United Nations Office on Drugs and Crime (UNODC). (2023). World drug report. *United Nations Publication*.

van Amsterdam, J., van der Velde, B., Schulte, M., & van den Brink, W. (2018). Causal Factors of Increased Smoking in ADHD: A Systematic Review. *Substance Use and Misuse*, *53*(3), 432–445. https://doi.org/10.1080/10826084.2017.1334066

van de Glind, G., Konstenius, M., Koeter, M. W. J., van Emmerik-van Oortmerssen, K., Carpentier, P. J., Kaye, S., Degenhardt, L., Skutle, A., Franck, J., Bu, E. T., Moggi, F., Dom, G., Verspreet, S., Demetrovics, Z., Kapitány-Fövény, M., Fatséas, M., Auriacombe, M., Schillinger, A., Møller, M., … Young, J. (2014). Variability in the prevalence of adult ADHD in treatment seeking substance use disorder patients: Results from an international multi-center study exploring DSM-IV and DSM-5 criteria. *Drug and Alcohol Dependence*, *134*(1), 158–166. https://doi.org/10.1016/j.drugalcdep.2013.09.026

van de Weijer, M. P., Pelt, D. H. M., van Beijsterveldt, C. E. M., Willemsen, G., & Bartels, M. (2022). Genetic factors explain a significant part of associations between adolescent well-being and the social environment. *European Child and Adolescent Psychiatry*, *31*(10), 1611–1622. https://doi.org/10.1007/s00787-021-01798-3

Van Den Bree, M. B. M., & Pickworth, W. B. (2005). Risk factors predicting changes in marijuana involvement in teenagers. *Archives of General Psychiatry*, *62*(3), 311–319. https://doi.org/10.1001/archpsyc.62.3.311

Van Emmerik-van Oortmerssen, K., van de Glind, G., Koeter, M. W. J., Allsop, S., Auriacombe, M., Barta, C., Bu, E. T. H., Burren, Y., Carpentier, P. J., Carruthers, S., Casas, M., Demetrovics, Z., Dom, G., Faraone, S. V., Fatseas, M., Franck, J., Johnson, B., Kapitány-Fövény, M., Kaye, S., … Young, J. (2014). Psychiatric comorbidity in treatment-seeking substance use disorder patients with and without attention deficit hyperactivity disorder: Results of the IASP study. *Addiction*, *109*(2), 262–272. https://doi.org/10.1111/add.12370

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. In *Nature Reviews Genetics* (Vol. 20, Issue 10, pp. 567–581). https://doi.org/10.1038/s41576-019-0137-z

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., O'Reilly, P., Curtis, C., Kolliakou, A., Patel, H., Newhouse, S., Traylor, M., Ajnakina, O., Mondelli, V., Marques, T. R., Gardner-Sood, P., Aitchison, K. J., Powell, J., Atakan, Z., … Breen, G. (2017). An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biological Psychiatry*, *81*(6), 470–477. https://doi.org/10.1016/j.biopsych.2016.06.028

Vaucher, J., Keating, B. J., Lasserre, A. M., Gan, W., Lyall, D. M., Ward, J., Smith, D. J., Pell, J. P., Sattar, N., Paré, G., & Holmes, M. V. (2018). Cannabis use and risk of schizophrenia: A Mendelian randomization study. *Molecular Psychiatry*, *23*(5), 1287–1292. https://doi.org/10.1038/mp.2016.252

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: A meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*(5), 1061–1072. https://doi.org/10.1017/S0033291714002165

Verlouw, J. A. M., Clemens, E., de Vries, J. H., Zolk, O., Verkerk, A. J. M. H., am Zehnhoff-Dinnesen, A., Medina-Gomez, C., Lanvers-Kaminsky, C., Rivadeneira, F., Langer, T., van Meurs, J. B. J., van den Heuvel-Eibrink, M. M., Uitterlinden, A. G., & Broer, L. (2021). A comparison of genotyping arrays. *European Journal of Human Genetics*, *29*(11), 1611–1624. https://doi.org/10.1038/s41431-021-00917-7

Vermeulen, J. M., Wootton, R. E., Treur, J. L., Sallis, H. M., Jones, H. J., Zammit, S., van den Brink, W., Goodwin, G. M., de Haan, L., & Munafò, M. R. (2021). Smoking and the risk for bipolar disorder: evidence from a bidirectional Mendelian randomisation study. *The British Journal of Psychiatry : The Journal of Mental Science*, *218*(2), 88–94. https://doi.org/10.1192/bjp.2019.202

Vink, J. M. (2016). Genetics of addiction: Future focus on gene × environment interaction? *Journal of Studies on Alcohol and Drugs*, *77*(5), 684–687. https://doi.org/10.15288/jsad.2016.77.684

Vink, J. M., Treur, J. L., Pasman, J. A., & Schellekens, A. (2021). Investigating genetic correlation and causality between nicotine dependence and ADHD in a broader psychiatric context. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *186*(7), 423–429.

https://doi.org/10.1002/ajmg.b.32822

Volkow, N. D., Fowler, J. S., Wang, G. J., Telang, F., Logan, J., Jayne, M., Ma, Y., Pradhan, K., Wong, C., & Swanson, J. M. (2010). Cognitive control of drug craving inhibits brain reward regions in cocaine abusers. *NeuroImage*, *49*(3), 2536–2543. https://doi.org/10.1016/j.neuroimage.2009.10.088

Volkow, N. D., Michaelides, M., & Baler, R. (2019). The neuroscience of drug reward and addiction. *Physiological Reviews*, *99*(4), 2115–2140. https://doi.org/10.1152/physrev.00014.2018

Volkow, N. D., Wang, G. J., Telang, F., Fowler, J. S., Logan, J., Jayne, M., Ma, Y., Pradhan, K., & Wong, C. (2007). Profound decreases in dopamine release in striatum in detoxified alcoholics: Possible orbitofrontal involvement. *Journal of Neuroscience*, *27*(46), 12700–12706. https://doi.org/10.1523/JNEUROSCI.3371-07.2007

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S. A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L. S., Clarke, T. K., Chou, Y. L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., … Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Wang, Y., Tsuo, K., Kanai, M., Neale, B. M., & Martin, A. R. (2022). Challenges and Opportunities for Developing More Generalizable Polygenic Risk Scores. *Annual Review of Biomedical Data Science*, *5*(1), 293–320. https://doi.org/10.1146/annurev-biodatasci-111721-074830

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (Sf-36): I. conceptual framework and item selection. *Medical Care*, *30*(6), 473–483. https://doi.org/10.1097/00005650-199206000-00002

Waszczuk, M. A., Eaton, N. R., Krueger, R. F., Shackman, A. J., Waldman, I. D., Zald, D. H., Lahey, B. B., Patrick, C. J., Conway, C. C., Ormel, J., Hyman, S. E., Fried, E. I., Forbes, M. K., Docherty, A. R., Althoff, R. R., Bach, B., Chmielewski, M., DeYoung, C. G., Forbush, K. T., … Kotov, R. (2019). Redefining Phenotypes to Advance Psychiatric Genetics: Implications From Hierarchical Taxonomy of Psychopathology. *Journal of Abnormal Psychology*, *129*(2), 143–161. https://doi.org/10.1037/abn0000486

Weerts, E. M., Wand, G. S., Maher, B., Xu, X., Stephens, M. A., Yang, X., & McCaul, M. E. (2017). Independent and Interactive Effects of OPRM1 and DAT1 Polymorphisms on Alcohol Consumption and Subjective Responses in Social Drinkers. *Alcoholism: Clinical and Experimental Research*, *41*(6), 1093–1104. https://doi.org/10.1111/acer.13384

Wendt, F. R., Pathak, G. A., Levey, D. F., Nuñez, Y. Z., Overstreet, C., Tyrrell, C., Adhikari, K., De Angelis, F., Tylee, D. S., Goswami, A., Krystal, J. H., Abdallah, C. G., Stein, M. B., Kranzler, H. R., Gelernter, J., & Polimanti, R. (2021). Sex-stratified gene-by-environment genome-wide interaction study of trauma, posttraumatic-stress, and suicidality. *Neurobiology of Stress*, *14*, 100309. https://doi.org/10.1016/j.ynstr.2021.100309

Wertz, J., Caspi, A., Belsky, D. W., Beckley, A. L., Arseneault, L., Barnes, J. C., Corcoran, D. L., Hogan, S., Houts, R. M., Morgan, N., Odgers, C. L., Prinz, J. A., Sugden, K., Williams, B. S., Poulton, R., & Moffitt, T. E. (2018). Genetics and Crime: Integrating New Genomic Discoveries Into Psychological Research About Antisocial Behavior. *Psychological Science*, *29*(5), 791–803. https://doi.org/10.1177/0956797617744542

Whitfield, T. W., Schlosburg, J. E., Wee, S., Gould, A., George, O., Grant, Y., Zamora-Martinez, E. R., Edwards, S., Crawford, E., Vendruscolo, L. F., & Koob, G. F. (2015). κ Opioid Receptors in the Nucleus Accumbens Shell Mediate Escalation of Methamphetamine Intake. *Journal of Neuroscience*, *35*(10), 4296–4305. https://doi.org/10.1523/JNEUROSCI.1978-13.2015

Wilens, T. E. (2007). The nature of the relationship between attention-deficit/hyperactivity disorder and substance use. *Journal of Clinical Psychiatry*, *68*(SUPPL. 11), 4–8.

Wimberley, T., Agerbo, E., Horsdal, H. T., Ottosen, C., Brikell, I., Als, T. D., Demontis, D., Børglum, A. D., Nordentoft, M., Mors, O., Werge, T., Hougaard, D., Bybjerg-Grauholm, J., Hansen, M. B., Mortensen, P. B., Thapar, A., Riglin, L., Langley, K., & Dalsgaard, S. (2020). Genetic liability to ADHD and substance use disorders in individuals with ADHD. *Addiction*, *115*(7), 1368–1377. https://doi.org/10.1111/add.14910

Wootton, R. E., Richmond, R. C., Stuijfzand, B. G., Lawn, R. B., Sallis, H. M., Taylor, G. M. J., Hemani, G., Jones, H. J., Zammit, S., Davey Smith, G., & Munafò, M. R. (2020). Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. *Psychological Medicine*, *50*(14), 2435–2443. https://doi.org/10.1017/S0033291719002678

World Health Organization. (2004). *ICD-10 : international statistical classification of diseases and related health problems : tenth revision* (2nd ed). World Health Organization. https://apps.who.int/iris/handle/10665/42980

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*(10), 1520–1528. https://doi.org/10.1101/gr.6665407

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(10), 1068–1087. https://doi.org/10.1111/jcpp.12295

Wray, N. R., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., Murray, G. K., & Visscher, P. M. (2021). From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry*, *78*(1), 101–109. https://doi.org/10.1001/jamapsychiatry.2020.3049

Xu, K., Li, B., McGinnis, K. A., Vickers-Smith, R., Dao, C., Sun, N., Kember, R. L., Zhou, H., Becker, W. C., Gelernter, J., Kranzler, H. R., Zhao, H., & Justice, A. C. (2020). Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nature Communications 2020 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-18489-3

Xue, A., Jiang, L., Zhu, Z., Wray, N. R., Visscher, P. M., Zeng, J., & Yang, J. (2021). Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. *Nature Communications*, *12*(1), 20211. https://doi.org/10.1038/s41467-020-20237-6

Yang, B. Z., Zhou, H., Cheng, Z., Kranzler, H. R., & Gelernter, J. (2019). Genomewide Gene-by-Sex Interaction Scans Identify ADGRV1 for Sex Differences in Opioid Dependent African Americans. *Scientific Reports*, *9*(1), 18070. https://doi.org/10.1038/s41598-019-53560-0

Yang, L. H., Wong, L. Y., Grivel, M. M., & Hasin, D. S. (2017). Stigma and substance use disorders: An international phenomenon. *Current Opinion in Psychiatry*, *30*(5), 378–388.

https://doi.org/10.1097/YCO.0000000000000351

Young, G. (2016). DSM-5: Basics and Critics. In G. Young (Ed.), *Unifying Causality and Psychology* (pp. 565–590). Springer International Publishing. https://doi.org/10.1007/978-3-319-24094-7_22

Young, J. G. (1967). Meeting the social needs of medicare patients. *Gerontologist*, *7*(4), 261–265. https://doi.org/10.1093/geront/7.4.261

Young, S., Abbasian, C., Al-Attar, Z., Branney, P., Colley, B., Cortese, S., Cubbin, S., Deeley, Q., Gudjonsson, G. H., Hill, P., Hollingdale, J., Jenden, S., Johnson, J., Judge, D., Lewis, A., Mason, P., Mukherjee, R., Nutt, D., Roberts, J., … Cocallis, K. (2023). Identification and treatment of individuals with attention-deficit/hyperactivity disorder and substance use disorder: An expert consensus statement. *World Journal of Psychiatry*, *13*(3), 84–112. https://doi.org/10.5498/wjp.v13.i3.84

Yuodelis-Flores, C., & Ries, R. K. (2015). Addiction and suicide: A review. *American Journal on Addictions*, *24*(2), 98–104. https://doi.org/10.1111/AJAD.12185

Zeggini, E., & Ioannidis, J. P. A. (2009). Meta-analysis in genome-wide association studies. *Pharmacogenomics*, *10*(2), 191–201. https://doi.org/10.2217/14622416.10.2.191

Zhang, X., Zhang, X., Liu, M., Zhu, L., & He, Z. (2022). Global, regional, and national burden of cirrhosis and other chronic liver diseases due to alcohol use, 1990–2019: a systematic analysis for the Global Burden of Disease study 2019. *BMC Gastroenterology*, *22*(1), 484. https://doi.org/10.1186/s12876-022-02518-0

Zheng, J., Richardson, T. G., Millard, L. A. C., Hemani, G., Elsworth, B. L., Raistrick, C. A., Vilhjalmsson, B., Neale, B. M., Haycock, P. C., Smith, G. D., & Gaunt, T. R. (2018). PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. *GigaScience*, *7*(8), 1–10. https://doi.org/10.1093/gigascience/giy090

Zheutlin, A. B., Dennis, J., Linnér, R. K., Moscati, A., Restrepo, N., Straub, P., Ruderfer, D., Castro, V. M., Chen, C. Y., Ge, T., Huckins, L. M., Charney, A., Kirchner, H. L., Stahl, E. A., Chabris, C. F., Davis, L. K., & Smoller, J. W. (2019). Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. *American Journal of Psychiatry*, *176*(10), 846–855. https://doi.org/10.1176/appi.ajp.2019.18091085

Zhou, H., Rentsch, C. T., Cheng, Z., Kember, R. L., Nunez, Y. Z., Sherva, R. M., Tate, J. P., Dao, C., Xu, K., Polimanti, R., Farrer, L. A., Justice, A. C., Kranzler, H. R., & Gelernter, J. (2020). Association of OPRM1 Functional Coding Variant with Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry*, *77*(10), 1072–1080. https://doi.org/10.1001/jamapsychiatry.2020.1206

Zhou, H., Sealock, J. M., Sanchez-Roige, S., Clarke, T. K., Levey, D. F., Cheng, Z., Li, B., Polimanti, R., Kember, R. L., Smith, R. V., Thygesen, J. H., Morgan, M. Y., Atkinson, S. R., Thursz, M. R., Nyegaard, M., Mattheisen, M., Børglum, A. D., Johnson, E. C., Justice, A. C., … Gelernter, J. (2020). Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nature Neuroscience*, *23*(7), 809–818. https://doi.org/10.1038/s41593-020-0643-5

Zilberman, N., Yadid, G., Efrati, Y., Neumark, Y., & Rassovsky, Y. (2018). Personality profiles of substance and behavioral addictions. *Addictive Behaviors*, *82*, 174–181.

https://doi.org/10.1016/j.addbeh.2018.03.007

Zucker, R. A., Heitzeg, M. M., & Nigg, J. T. (2011). Parsing the undercontrol-disinhibition pathway to substance use disorders: A multilevel developmental problem. *Child Development Perspectives*, *5*(4), 248–255. https://doi.org/10.1111/j.1750-8606.2011.00172.x

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A Comparison of Three Structural Models for Personality: The Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology*, *65*(4), 757–768. https://doi.org/10.1037/0022-3514.65.4.757

APPENDIX A

PUBLICATIONS INDIRECTLY RELATED TO THE THESIS

2023

Cabana-Domínguez, J., Llonga, N., Arribas, L., Alemany, S., **Vilar-Ribó, L.**, Demontis, D., Fadeuilhe, C., Corrales, M., Richarte, V., Børglum, A. D., Ramos-Quiroga, J. A., Soler Artigas, M., & Ribasés, M. (2023). Transcriptomic risk scores for attention deficit/hyperactivity disorder. Molecular psychiatry. https://doi.org/10.1038/s41380-023-02200-1.

Alemany, S., Soler-Artigas, M., Cabana-Domínguez, J., Fakhreddine, D., Llonga, N., **Vilar-Ribó, L.,** Rodríguez-Urrutia, A., Palacio, J., González-Castro, A. M., Lobo, B., Alonso-Cotoner, C., Simrén, M., Santos, J., Ramos-Quiroga, J. A., & Ribasés, M. (2023). Genome-wide multi-trait analysis of irritable bowel syndrome and related mental conditions identifies 38 new independent variants. Journal of translational medicine, 21(1), 272. https://doi.org/10.1186/s12967-023-04107-5.

Vilar-Ribó, L., Cabana-Domínguez, J., Martorell, L., Ramos-Quiroga, J. A., Sanchez-Roige, S., Palmer, A. A., Vilella, E., Ribasés, M., Muntané, G., & Soler Artigas, M. (2023). Shared genetic architecture between attention-deficit/hyperactivity disorder and lifespan. Neuropsychopharmacology. 48(7), 981–990. https://doi.org/10.1038/s41386-023-01555-x.

Soler Artigas, M., Sánchez-Mora, C., Rovira, P., **Vilar-Ribó, L.**, Ramos-Quiroga, J. A., & Ribasés, M. (2023). Mendelian randomization analysis for attention deficit/hyperactivity disorder: studying a broad range of exposures and outcomes. International journal of epidemiology, 52(2), 386–402. https://doi.org/10.1093/ije/dyac128.

2022

Cabana-Domínguez, J., Soler Artigas, M., Arribas, L., Alemany, S., **Vilar-Ribó, L.**, Llonga, N., Fadeuilhe, C., Corrales, M., Richarte, V., Ramos-Quiroga, J. A., & Ribasés, M. (2022). Comprehensive analysis of omics data identifies relevant gene networks for Attention-Deficit/Hyperactivity Disorder (ADHD). Translational psychiatry, 12(1), 409. https://doi.org/10.1038/s41398-022-02182-8.

Rajagopal, V. M., Duan, J., **Vilar-Ribó, L.**, Grove, J., Zayats, T., Ramos-Quiroga, J. A., Satterstrom, F. K., Artigas, M. S., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Als, T. D., Rosengren, A., Daly, M. J., Neale, B. M., Nordentoft, M., Werge, T., Mors, O., Hougaard, D. M., Mortensen, P. B., Ribasés, M., … Demontis, D. (2022). Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder. Nature genetics, 54(8), 1117–1124. https://doi.org/10.1038/s41588-022-01143-7.

Mullins, N., Kang, J., Campos, A. I., Coleman, J. R. I., Edwards, A. C., Galfalvy, H., Levey, D. F., Lori, A., Shabalin, A., Starnawska, A., Su, M. H., Watson, H. J., Adams, M., Awasthi, S., Gandal, M., Hafferty, J. D., Hishimoto, A., Kim, M., Okazaki, S., Otsuka, I., […] **Vilar-Ribó L.**, […] Ruderfer, D. M. (2022). Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. Biological psychiatry, 91(3), 313–327. https://doi.org/10.1016/j.biopsych.2021.05.029

2021

Richarte, V., Sánchez-Mora, C., Corrales, M., Fadeuilhe, C., **Vilar-Ribó, L.,** Arribas, L., Garcia, E., Rosales-Ortiz, S. K., Arias-Vasquez, A., Soler-Artigas, M., Ribasés, M., & Ramos-Quiroga, J. A. (2021). Gut microbiota signature in treatment-naïve attention-deficit/hyperactivity disorder. Translational psychiatry, 11(1), 382. https://doi.org/10.1038/s41398-021-01504-6.

2020

Mortimer, N., Sánchez-Mora, C., Rovira, P., **Vilar-Ribó, L.**, Richarte, V., Corrales, M., Fadeuilhe, C., Rivero, O., Lesch, K. P., Casas, M., Ramos-Quiroga, J. A., Artigas, M. S., & Ribasés, M. (2020). Transcriptome profiling in adult attention-deficit hyperactivity disorder. European neuropsychopharmacology, 41, 160–166. https://doi.org/10.1016/j.euroneuro.2020.11.005.

Rovira, P., Sánchez-Mora, C., Pagerols, M., Richarte, V., Corrales, M., Fadeuilhe, C., **Vilar-Ribó, L.**, Arribas, L., Shireby, G., Hannon, E., Mill, J., Casas, M., Ramos-Quiroga, J. A., Soler Artigas, M., & Ribasés, M. (2020). Epigenome-wide association study of attention-deficit/hyperactivity disorder in adults. Translational psychiatry, 10(1), 199. https://doi.org/10.1038/s41398-020-0860-4.

Rovira, P., Demontis, D., Sánchez-Mora, C., Zayats, T., Klein, M., Mota, N. R., Weber, H., Garcia-Martínez, I., Pagerols, M., **Vilar-Ribó, L.,** Arribas, L., Richarte, V., Corrales, M., Fadeuilhe, C., Bosch, R., Martin, G. E., Almos, P., Doyle, A. E., Grevet, E. H., Grimm, O., … Ribasés, M. (2020). Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology, 45(10), 1617–1626. https://doi.org/10.1038/s41386-020-0664-5.

2019

Gisbert, L., **Vilar, L.**, Rovira, P., Sánchez-Mora, C., Pagerols, M., Garcia-Martínez, I., Richarte, V., Corrales, M., Casas, M., Ramos-Quiroga, J. A., Soler Artigas, M., & Ribasés, M. (2019). Genome-wide analysis of emotional lability in adult attention deficit hyperactivity disorder (ADHD). European neuropsychopharmacology, 29(6), 795–802. https://doi.org/10.1016/j.euroneuro.2019.04.004.

ARTICLE OPEN

Check for updates

Transcriptomic risk scores for attention deficit/hyperactivity disorder

Judit Cabana-Domínguez[1,2,3,4,10]✉, Natalia Llonga[1,2,10], Lorena Arribas[1,2], Silvia Alemany[1,2], Laura Vilar-Ribó[1,2,3], Ditte Demontis[5,6,7,8], Christian Fadeuilhe[1,2,3,9], Montse Corrales[1,2,3,9], Vanesa Richarte[1,2,3,9], Anders D. Børglum[5,6,7], Josep Antoni Ramos-Quiroga[1,2,3,9], María Soler Artigas[1,2,3,4,11]✉ and Marta Ribasés[1,2,3,4,11]✉

Attention deficit/hyperactivity disorder (ADHD) is a highly heritable neurodevelopmental disorder. We performed a transcriptome-wide association study (TWAS) using the latest genome-wide association study (GWAS) meta-analysis, in 38,691 individuals with ADHD and 186,843 controls, and 14 gene-expression reference panels across multiple brain tissues and whole blood. Based on TWAS results, we selected subsets of genes and constructed transcriptomic risk scores (TRSs) for the disorder in peripheral blood mononuclear cells of individuals with ADHD and controls. We found evidence of association between ADHD and TRSs constructed using expression profiles from multiple brain areas, with individuals with ADHD carrying a higher burden of TRSs than controls. TRSs were uncorrelated with the polygenic risk score (PRS) for ADHD and, in combination with PRS, improved significantly the proportion of variance explained over the PRS-only model. These results support the complementary predictive potential of genetic and transcriptomic profiles in blood and underscore the potential utility of gene expression for risk prediction and deeper insight in molecular mechanisms underlying ADHD.

Molecular Psychiatry; https://doi.org/10.1038/s41380-023-02200-1

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by inappropriate levels of inattentiveness, hyperactivity, or impulsivity that affects around 2.6% of persistent adult ADHD and 6.8% of symptomatic adult ADHD [1]. ADHD increases the risk of health problems, psychiatric comorbidities, psychological dysfunction, social disability, academic and occupational failure, and risk behaviours throughout the individual's life [2].

Twin and family studies show a strong genetic component underlying the disorder, with a heritability of 76% [3, 4]. Recently, the largest genome-wide association study meta-analysis (GWAS-MA) on ADHD so far in 38,691 individuals with ADHD and 186,691 controls identified 27 hits for the disorder [5]. In addition, to date more than 40 relevant studies on polygenic risk scores (PRS) for ADHD have been published and show evidence of association between ADHD-PRS and a wide range of traits and disorders, including ADHD-related traits, reduced brain volume, lower education attainment, externalizing behaviours, impaired working memory, higher body mass index or lower socioeconomic status, among others [6].

The SNP-based heritability for ADHD estimated so far is 0.14 [5] and the PRS for the disorder explains 5.5% of phenotypic variance in individuals of European ancestry [7]. A large proportion of the heritability still needs to be explained and gene expression, which results from the interplay between genetic and environmental factors, may help to elucidate additional phenotypic variance. To date, eight studies on transcriptome profiling in ADHD have been performed and highlighted genes involved in several neuronal functions and in the immune system [8–16]. However, this approach is limited by the inaccessibility of brain samples and has mainly focused on blood. Alternatively, integrative approaches have been developed, including transcriptome-wide association studies (TWAS), which are a powerful method to integrate GWAS data and multi-tissue expression quantitative trait loci (eQTL) to correlate genetically predicted gene expression levels with complex traits. To date, four TWAS on ADHD have been performed: three using summary statistics from the first GWAS-MA on ADHD by Demontis et al. [7, 17–19] and one using data from the latest GWAS-MA on ADHD [5]. Briefly, Fahira et al. conducted multiple TWAS approaches to identify 47 putative causal genes and the glutamate receptor signalling pathway underlying ADHD [17]. Liao et al. performed TWAS on 11 brain tissues and identified novel genes and several pathways relevant for ADHD, including the dopaminergic neuron differentiation and norepinephrine neurotransmitter release cycle [18]. Qi et al.

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. [5]Department of Biomedicine/Human Genetics, Aarhus University, Aarhus, Denmark. [6]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [7]Center for Genomics and Personalized Medicine, Aarhus, Denmark. [8]The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [9]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [10]These authors contributed equally: Judit Cabana-Domínguez, Natalia Llonga. [11]These authors jointly supervised this work: María Soler Artigas, Marta Ribasés. ✉email: judit.cabana@vhir.org; maria.soler@vhir.org; marta.ribases@vhir.org

Received: 7 February 2023 Revised: 17 July 2023 Accepted: 21 July 2023
Published online: 03 August 2023

considered Chinese and European ancestry cohorts and did not identify transcriptome-wide associated genes with the disorder either in brain or blood [19]. Finally, Demontis et al. identified 23 distinct genes with differential predicted gene expression in the dorsolateral prefrontal cortex (DLPFC) in ADHD using the largest GWAS-MA on ADHD to date and highlighted *PPP1R16A* and *B4GALT2* as top genes [5].

Given that a substantial proportion of GWAS association signals demonstrate gene regulation effects [20], risk scores built on eQTL variants, known as transcriptomic risk scores (TRSs), are promising gene-based approaches that use gene expression information to identify trait-associated genes from GWAS. TRSs are significantly associated with a range of outcomes, including Amyotrophic Lateral Sclerosis [21], Alzheimer's disease [22], and Crohn's disease [23] based on observed gene expression data, as well as with ADHD symptoms [24], schizophrenia [25, 26], and major depressive disorder [24, 27] constructed with predicted gene expression. In addition, the combination of TRS with PRS improves risk prediction of several traits, including rheumatoid arthritis, height, body mass index or intelligence [24].

In the present study, we ran a multi-tissue TWAS on the latest GWAS-MA on ADHD performed so far [5], and for the first time used TWAS results to select a subset of signature genes per tissue and construct microarray-based TRSs in peripheral blood mononuclear cells (PBMCs), tested their association with ADHD and assessed whether the combination of PRS and TRS increases significantly the proportion of variance explained of ADHD over PRS alone, in subjects with ADHD and controls.

MATERIALS AND METHODS

Multi-tissue transcriptome-wide association study (TWAS)

TWAS was performed with S-PrediXcan (https://github.com/hakyimlab/MetaXcan) [28] using summary statistics from the largest GWAS-MA on ADHD to date in 38,691 individuals with ADHD and 186,843 controls [5], and SNP-weights of gene expression precomputed with the joint-tissue imputation (JTI) approach [29]. We used genetic variants with minor allele frequency (MAF) ≥ 0.01 and INFO score ≥ 0.80, and gene expression reference panels from GTEx v8 in 14 tissues, including whole blood, amygdala, anterior cingulate cortex, caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia, spinal cord cervical C1 and substantia nigra [30]. According to the GTEx webpage (https://gtexportal.org/home/samplingSitePage) both cortex and frontal cortex correspond to the same brain area, right cerebral frontal pole cortex, sampled and collected using different techniques. We considered default settings in S-Predixcan and linkage disequilibrium (LD) estimates from the European subset of the 1000 Genomes Phase 3 reference sample with the precalculated covariances. As TWAS results from different brain areas were highly correlated ($r^2 > 0.96$ when considering genes nominally associated with ADHD), we applied Bonferroni correction considering the number of genes tested within each of the 14 expression reference panels separately to account for multiple testing.

Summary statistics from TWAS in DLPFC described in Demontis et al. 2022 were also used in the TRS analysis [5]. In brief, the reference panel was constructed using EpiXcan and expression data on DLPF of 924 samples with European ancestry from the PsychENCODE Consortium [31], and the S-PrediXcan method was used to integrate the ADHD GWAS meta-analysis summary statistics [5].

Enrichment analyses on gene-sets from the Molecular Signatures Database (MSigDB v6.2), including Gene ontology (GO), KEGG, Reactome, miRNA targets and GWAS Catalog, were performed on genes nominally associated with ADHD in each TWAS using a hypergeometric test with the GENE2FUNC module of FUMA and considering all genes from the TWAS as background [32]. Enrichment analyses results were corrected for multiple comparisons in each tissue considering each category separately using 5% False Discovery Rate (FDR).

Gene locus-level colocalization analysis. Gene locus-level colocalization probability (GLCP) for significant genes identified in TWAS was performed using fastENLOC and only genes with a GLCP ≥ 0.5 were considered further [33, 34]. First, we selected the genetic variants within 1 Mb upstream and 500 kb downstream from each of the 56 significant genes identified in TWAS with a $P < 0.05$ in the GWAS-MA of Demontis et al. [5]. These variants were fine-mapped to generate 95% credible sets, assuming one causal variant per locus, using the CAUSALdb pipeline (https://github.com/mulinlab/CAUSALdb-finemapping-pip#4; [35]) which includes three different fine-mapping tools, FINEMAP 1.3.1 [36], PAINTOR v3.0 [37] and CAVIARBF v.0.2.1 [38]. We used the recommended parameters of each tool and only variants selected by all three methods were considered. For these variants, Z-scores from the GWAS-MA on ADHD [5] were then converted to posterior inclusion probabilities using the *torus* software [39]. Finally, these data were colocalized with fastENLOC for the 14 GTEx v8 tissues included in the study [33]. Colocalization was performed using pre-computed GTEx multi-tissue annotations obtained from https://github.com/xqwen/fastenloc.

Transcriptomic and polygenic risk scores

Participants and clinical assessment. TRSs and PRS were constructed in an in-house sample of 222 medication-naïve adult ADHD cases (59.45% male, mean age=34.03 years, s.d = 11.62) and 269 controls (57.25% male, mean age=36.6 years, s.d = 10.06). All subjects were from European ancestry, which was confirmed through principal component analysis (PCA) using genetic data. Clinical assessment was conducted by structured interviews and self-reported questionnaires as previously described [14], based in two steps: (i) assessment of ADHD diagnosis based on symptomatology using the Conner's Adult ADHD Diagnostic Interview for DSM-IV (CAADID) and (ii) assessment of the severity of ADHD symptoms, the levels of impairment and the presence of comorbid disorders to increase the diagnostic accuracy with the Conners' ADHD Rating Scale (CAARS), the ADHD Rating Scale (ADHD-RS), the Clinical Global Impression (CGI), the Wender Utah Rating Scale (WURS), the Sheehan Disability Inventory (SDS), and the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II). Exclusion criteria were IQ < 70; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight ≤1.5 kg; and other neurological or systemic disorders that might explain ADHD symptoms. All cases were evaluated and recruited prospectively from a restricted geographic area in a specialized out-patient program for adult ADHD at the Hospital Universitari Vall d'Hebron of Barcelona (Spain).

The control sample consisted of unrelated blood donors matched by sex with the clinical group. Individuals with ADHD symptomatology were excluded retrospectively from the control sample under the following criteria: (1) diagnosed with ADHD previously and (2) answering positively to the life-time presence of the following ADHD symptoms: (a) often has trouble in keeping attention on tasks, (b) usually loses things needed for tasks, (c) often fidgets with hands or feet or squirms in seat, and (d) often gets up from seat when remaining in seat is expected. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, methods were performed in accordance with the relevant guidelines and regulations and written informed consent was obtained from all subjects before inclusion in the study.

Transcriptomic risk scores. TRSs were constructed from transcriptomic profiles in PBMCs separated by a Ficoll density gradient method immediately after blood extraction. Total RNA was isolated using Qiazol Lysis reagent and the RNAeasy Midi Kit (QIAgen, Hilden, Germany). RNA integrity and concentration were assayed by 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA). RNA was retrotranscribed using the Ambion WT Expression Kit (Life Technologies, Carlsbad, CA, USA). The cDNA was subsequently fragmented, labelled, and hybridized with the GeneChip WT Terminal Labelling and Hybridization Kit (Affymetrix, Santa Clara, CA, USA). Samples were hybridized to the GeneChip Human Gene 1.1 ST 96-Array plate (Affymetrix), covering a total of 36,079 transcripts that correspond to 21,014 genes. The array processing and data generation were assessed using the Gene Titan Affymetrix microarray platform. Raw data were pre-processed as previously described [40]. In brief, data was processed with the Robust Multichip Analysis (RMA) algorithm from *OligoR* [41], sample outliers were removed using the *arrayQualityMetrics* [42] and transcript probes were filtered ending up with 19,004 probes corresponding to 18,055 unique genes. Microarray batch effects and non-biological experimental variation (RNA integrity number (RIN), age and gender) were adjusted for using the *empiricalBayesLM* algorithm included in WGCNA R package [43]. Raw data from this article is not publicly available because of limitations in ethical approvals and the summary data will be available

upon request.

TRSs were calculated as the sum of the standardized expression of each gene weighted by its signed Z-score value from TWAS results on the different expression reference panels. TRSs per tissue were constructed by selecting genes under several TWAS *P*-value thresholds (Bonferroni, 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 1) and tested for association with ADHD using a logistic regression model in R, with sex, age, GWAS wave and the 10 first principal components based on GWAS data as covariates. For the best *P*-value threshold in each tissue, the empirical *P*-value was calculated by permuting the target phenotype 10,000 times and repeating the TRS analysis on each set of permuted phenotypes [44]. P*seudo*-R^2 were calculated using the Lee's formula [45] and considering an ADHD population prevalence of 5%. The effective number of independent tests was assessed with the Galwey method [46] considering Pearson correlation among TRSs from the best *P*-value threshold at each tissue, which resulted in 11 independent tissues out of 14. To account for multiple testing, we used the Sidák correction (*P*-value < 4.6e-03) for 11 independent tests. To discard an artificial inflation of the results due to the inclusion of different genes at the same genomic loci under the control of the same eQTL in the TRS construction, a sensitivity analysis was performed by calculating TRSs considering a single gene per locus: the one showing the lowest *P*-value in the TWAS at each genomic loci (defined by genes < 500 kb apart). Colocalization analyses were conducted using the same strategy described in the TWAS section, selecting genetic variants within a genomic window of 1 Mb upstream and 500 kb downstream from each of the genes in the best *P*-value threshold of TRSs associated with ADHD after multiple comparison corrections and sensitivity analyses.

Polygenic risk score. DNA samples were genotyped in two genotyping waves using Omni2.5 ($n = 163$) and Infinium™ Global Screening Array-24 v2.0 ($n = 328$) Illumina arrays. Polygenic scoring was conducted using the summary statistics from the largest GWAS-MA on ADHD in 38,691 individuals with ADHD and 186,843 controls [5], the PRS-CS software to generate posterior SNP effect sizes under continuous shrinkage (CS) priors

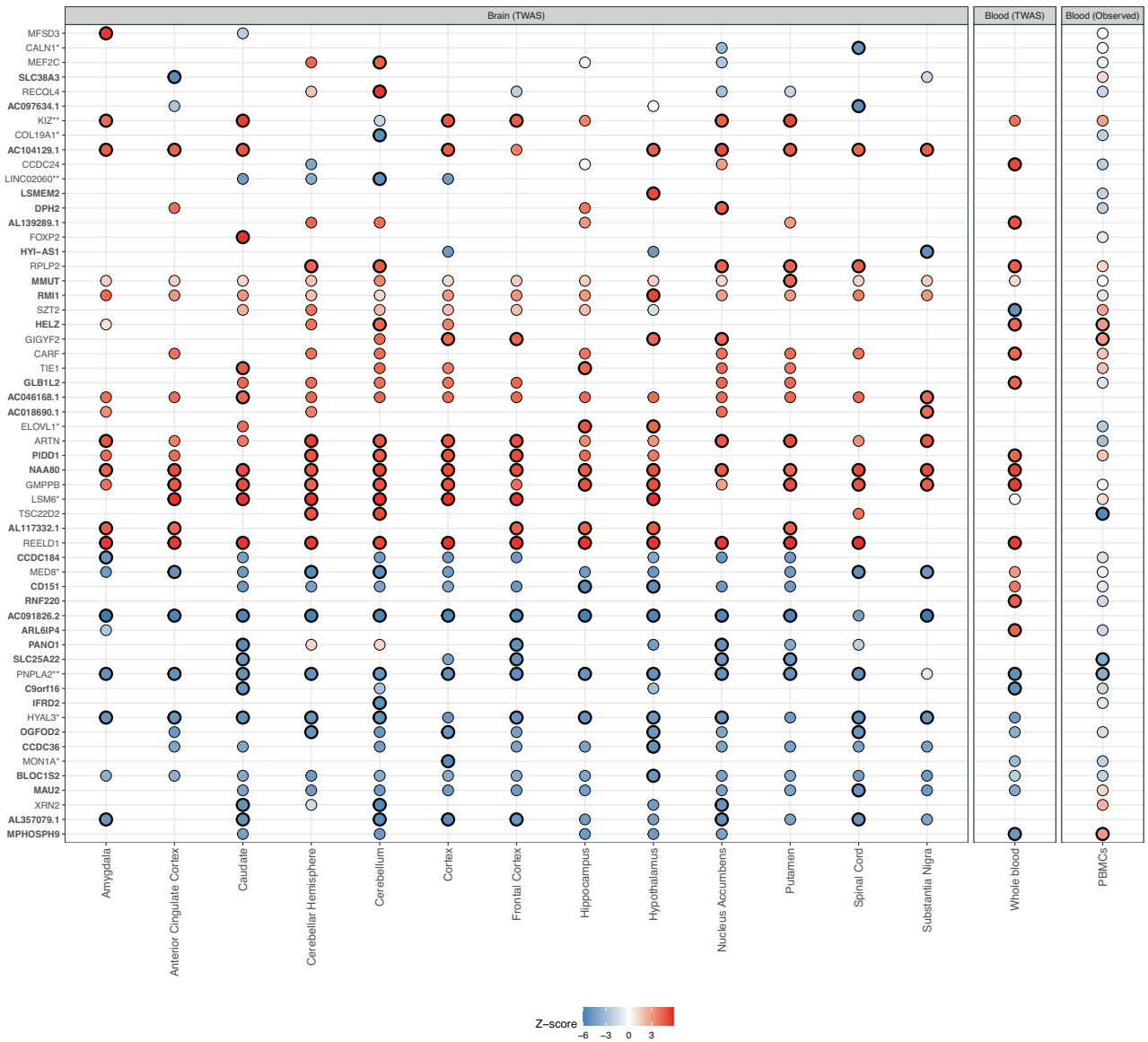


Fig. 1 Observed and predicted (multi-tissue TWAS) differential expression for ADHD. Z-scores are plotted for the 56 genes significantly associated with ADHD in at least one of the studied tissues. Significantly associated genes are outlined in black (TWAS results with $P < 8e\text{-}06$, and observed expression in blood FDR < 0.05). The dots are color-coded based on the Z-scores of the genes, with white indicating a Z-score of 0, blue indicating a negative Z-score, and red indicating a positive Z-score. In the y axis novel genes that were not previously identified in the original ADHD GWAS or TWAS on the dorsolateral prefrontal cortex by Demontis et al., 2022 are highlighted in bold. * Genes previously reported in the GWAS and TWAS on ADHD by Demontis et al., 2022. ** Genes previously reported in the TWAS on ADHD by Demontis et al., 2022.

Table 1. Models considering Transcriptomic Risk Scores (TRSs) or TRSs in combination with Polygenic Risk Scores (PRS).

	N*	TRS					PRS+TRS					
							PRS**		TRS			Likelihood ratio test *P*-value
		Best *P*-value threshold[a]	Estimate	*P*-value	Empirical *P*-value	*Pseudo*-R^2	Estimate	*P*-value	Estimate	*P*-value	*Pseudo*-R^2	
Amygdala[b]	68	0.001	0.28	4.3E-03	3.5E-03	0.013	0.34	6.9E-04	0.29	3.2E-03	0.034	2.8E-03
Anterior Cingulate Cortex	71	0.001	0.25	8.9E-03	9.7E-03	0.011	0.36	3.7E-04	0.28	3.4E-03	0.033	3.0E-03
Caudate[b]	12	BF	0.30	1.9E-03	1.5E-03	0.016	0.34	7.0E-04	0.31	1.5E-03	0.036	1.2E-03
Cerebellum	19	BF	0.23	0.019	0.017	0.009	0.33	8.0E-04	0.23	0.016	0.029	0.015
Cortex[b]	10	BF	0.44	1.5E-05	1.0E-04	0.032	0.34	8.1E-04	0.44	1.3E-05	0.052	7.1E-06
Dorsolateral Prefrontal Cortex[bc]	21	BF	0.40	6.6E-05	9.9E-05	0.028	0.32	1.5E-03	0.39	1.1E-04	0.046	6.9E-05
Frontal Cortex[b]	87	0.001	0.36	2.4E-04	4.0E-04	0.023	0.33	9.1E-04	0.37	2.3E-04	0.042	1.7E-04
Hippocampus	6	BF	0.22	0.020	0.021	0.009	0.34	6.4E-04	0.24	0.013	0.029	0.012
Nucleus accumbens	14	BF	0.27	6.5E-03	6.9E-03	0.012	0.33	1.1E-03	0.26	7.9E-03	0.031	7.2E-03
Putamen[bd]	10	BF	0.36	3.2E-04	5.0E-04	0.022	0.30	2.7E-03	0.34	9.1E-04	0.038	7.1E-04
Spinal cord	7	0.001	0.26	7.9E-03	8.4E-03	0.012	0.34	6.1E-04	0.27	5.2E-03	0.032	4.7E-03
Substantia nigra	1968	0.4	-0.21	0.028	0.033	0.008	0.33	1.0E-03	-0.21	0.031	0.027	0.030

**PRS-only model results: Estimate= 0.3295; *p*-value = 9.41E-04, Pseudo-R^2 = 0.0189. *Number of genes included in the TRS.

[a]BF: Bonferroni significance threshold.

[b]TRSs significantly associated with ADHD after the multiple testing correction of Sidák.

[c]TWAS from Demontis et al. 2022.

[d]Not significant after sensitivity analysis.

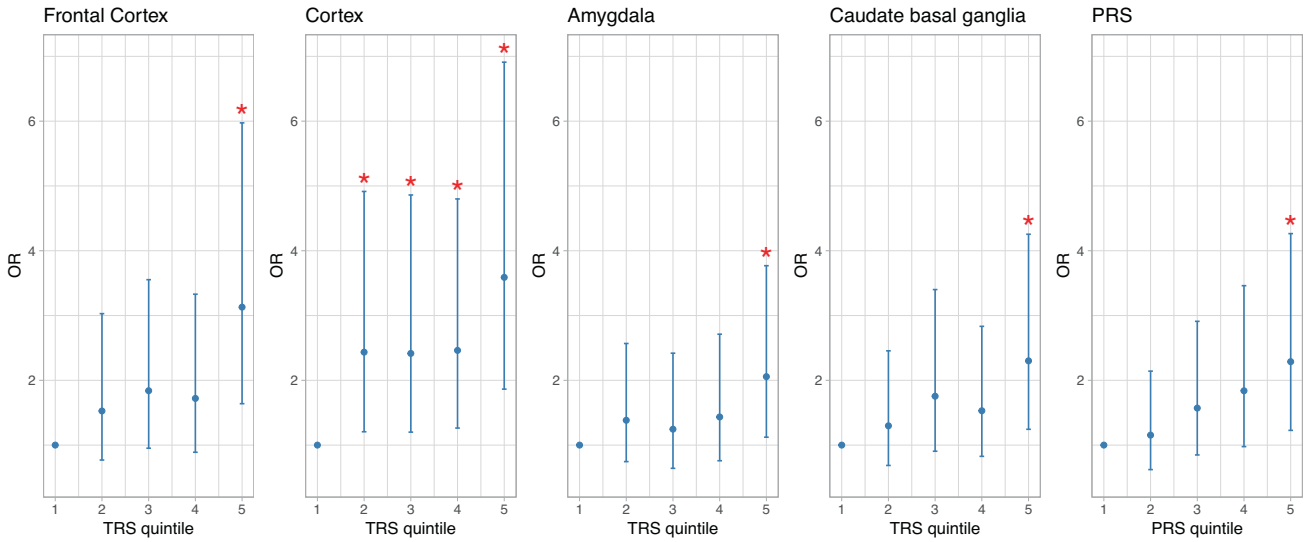


Fig. 2 Quintile plot of odds ratios for PRS and TRS. Odds ratios (OR) with 95% confidence intervals are shown for PRS and TRSs that overcome multiple comparison corrections and sensitivity analysis using the first quintile as baseline.

to model LD between genetic variants (https://github.com/getian107/PRScs) [47]. The European subset of the 1000 Genomes Phase 3 reference was used to estimate LD and a global shrinkage parameter of phi = 1e-02 was considered. The PRS was generated using PLINK 1.09 software [48] and it was tested for association with ADHD using a logistic regression model, with sex, age, GWAS wave and the 10 first principal components based on GWAS data as covariates. The increment in *pseudo*-R2 was calculated using the Lee's formula [45] and considering an ADHD population prevalence of 5%. Correlation between significant TRSs and PRS were calculated using the Pearson correlation coefficient. A likelihood ratio test with the *lmtest* R-package was used to compare the goodness of fit of the model that includes the PRS and covariates with the model that also includes the TRS.

RESULTS

Transcriptome-wide association study

We performed a TWAS in ADHD using multiple brain tissues and whole blood expression reference panels and summary-level data from the largest GWAS-MA on ADHD so far in 38,691 cases and 186,843 controls [5, 30] (Supplementary Fig. S1). Overall, we tested 20,225 predicted genes across expression reference panels, ranging from 6213 to 11,473 depending on the tissue under study, representing at least 95% of the genes included in each expression reference panel (Supplementary Table S1). We identified a total of 4134 unique genes showing nominal association ($P < 0.05$) with ADHD in at least one tissue, including 2234 that were significant in more than one and 94 in all of them. These genes were enriched for genes previously associated with social interaction (e.g. regular attendance at a religious group, regular attendance at a gym or sports club or social communication problems), psychiatric disorders (e.g. autism spectrum disorder, schizophrenia or bipolar disorder) and body fat distribution, among others (Supplementary Table S2). Besides, analysis on miRNA target genes revealed significant enrichment of targets of miRNA-34b/c and miR-449 among genes differentially expressed in the cerebellum and of 14 mature miRNAs in cortex (Supplementary Table S3). No association with other categories from the MSigDB was found.

After Bonferroni correction, 56 unique genes in 28 independent loci (defined by genes > 500 kb apart) showed transcriptome-wide significance, of which 28 were significant in more than one tissue, all of them showing consistent direction of the effect (Fig. 1 and Supplementary Table S4). Of them, 8 genes were identified both in blood and at least one brain tissue, and 26 in at least two brain areas, being *NAA80* the only gene differentially expressed in all the studied tissues (Fig. 1 and Supplementary Table S4). From the genes identified in the TWAS, 31 were novel and 25 were previously associated with ADHD either by TWAS or GWAS in the study by Demontis et al. 2022 (Fig. 1, Supplementary Table S5 and Supplementary Fig. S2).

When comparing the predicted differential expression from TWAS with observed differential expression in PBMCs in our in-house sample, we found that 41 out of 56 genes identified in TWAS were available in our microarray analysis and from those, six were significantly differentially expressed. Out of the five genes differentially expressed in PBMCs and in at least one brain tissue, four showed consistent direction of effect (*HELZ, GIGYF2, SLC25A22* and *PNPLA2*), with *PNPLA2* and *HELZ* also differentially expressed in the whole blood TWAS and with consistent direction of effect (Fig. 1 and Supplementary Table S4). *TSC22D2* had discordant direction of effect between PBMCs and cerebellar hemisphere/cerebellum and *MPHOSPH9* between PBMCs and whole blood (Fig. 1 and Supplementary Table S4). Finally, colocalization analyses of the 56 genes identified in TWAS revealed 14 genes with a GLCP ≥ 0.5 in at least one of the studied tissues, four of them differentially expressed also in PBMCs with consistent direction of effect (*GIGYF2*, *HELZ*, *PNPLA2* and *SLC25A22*; Supplementary Table S6). *PNPLA2* was the most ubiquitous gene found colocalized in 9 tissues (GCLP range: 0.643– 0.896), followed by *REELD1* in 8 tissues (GCLP range: 0.615–0.818), and *LSM6* in 7 tissues (GCLP range: 0.724–0.844; Supplementary Table S6).

Transcriptomic risk scores

TRSs based on multi-tissue TWAS results were constructed at different significance thresholds using expression data from PBMCs in an in-house sample of 222 subjects with ADHD and 269 controls (Supplementary Fig. S1). We found strong evidence of association in brain, with TRSs based on TWAS from 11 out of 13 brain tissues significantly associated with ADHD status after computing the empirical *P*-values (free from inflation due to overfitting) being cortex the most significant one ($P_{empirical}$ = 1e-04; Table 1 and Supplementary Fig. S3).

Although significant associations with ADHD were observed across the different TWAS *P*-value thresholds in most of the brain areas, there was clear evidence of increased proportion of variance explained by TRSs as lower *P*-value thresholds were used (Supplementary Fig. S3). After correction for multiple comparisons, TRSs remained significantly associated with ADHD when constructed on TWAS from five brain tissues,

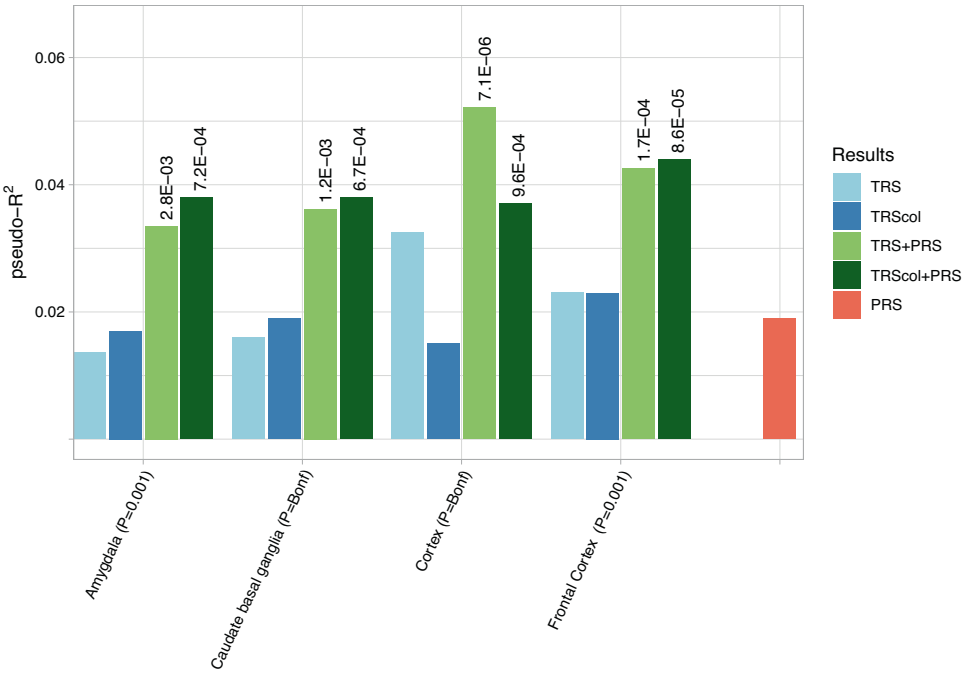


Fig. 3 Proportion of variance explained by TRSs significantly associated with ADHD and by PRS. Pseudo-R^2 (in the y-axis) is presented for each TRS, TRS restricted to colocalized genes (TRS_{col}), the PRS and the model that combines both scores (TRS + PRS or TRS_{col} + PRS). Values in brackets indicate the best *P*-value threshold for a given tissue (P). Likelihood ratio test *P*-values for TRS and PRS vs. PRS-only model comparisons are given above the bars. Further statistical details can be found Supplementary Table S9.

including cortex ($P_{empirical}$ = 1.0e-04, *pseudo*-R^2 = 0.032), frontal cortex ($P_{empirical}$ = 4.0e-04, *pseudo*-R^2 = 0.023), putamen ($P_{empirical}$ = 5.0e-04, *pseudo*-R^2 = 0.023), caudate basal ganglia ($P_{empirical}$ = 1.5e-03, *pseudo*-R^2 = 0.016) and amygdala ($P_{empirical}$ = 3.5e-03, *pseudo*-R^2 = 0.014), with subjects with ADHD-having a significantly higher ADHD-TRS than controls in all of them (Table 1 and Supplementary Fig. S4). Associations remained significant in the sensitivity analyses considering only the most significant gene per locus in the TRS construction, with the exception of the TRS based on TWAS results in putamen (Supplementary Table S7 and Fig. S5). The quintiles of the remaining TRSs showed the expected trend of higher ADHD odds for individuals in higher quintiles (Fig. 2) and positive correlations were found between the four TRSs (corrected $P < 7.1e-04$ and $0.22 \leq r \leq 0.62$) (Supplementary Fig. S6). Out of the 112 genes included in at least one of these TRSs, three were used in all four: *GMPPB, PLK1S1* and *PNPLA2* (Supplementary Table S8). Despite the proportion of variance explained for the TRSs being in line with that of PRS (Estimate = 0.3295, P = 9.4e-04, *pseudo*-R^2 = 0.019, Fig. 3), both scores were not correlated in any of the tissues with significant results after the sensitivity analyses ($r \leq -0.02$; Supplementary Fig. S6) and combining TRSs and PRS improved the fit of the model over PRS alone ($P < 0.03$), with TRSs from cortex showing the best results and reaching a *pseudo*-R^2 of 0.052 in the combined model (P = 7.1e-06, Table 1 and Fig. 3).

We also constructed TRSs restricted to colocalized genes (TRS_{col}) for the TRSs significantly associated with ADHD after multiple comparison corrections and sensitivity analyses. We found that, despite reducing the number of genes included, the association signal remained in all four tissues and that the predictive performance improved for TRS_{col} in three of them, amygdala, caudate basal ganglia and frontal cortex (Supplementary Table S9). Interestingly, out of the 24 genes included in at least one of these TRS_{col}, three genes were included in three out of the four analyses: *LSM6*, *PIDD1* and *PNPLA2*, with consistent direction effects across tissues (Supplementary Table S8). In line with the results from TRSs calculated with all genes, the combination of TRS_{col} with PRS improved the fit of the model over PRS alone for all four tissues ($P < 9.59e-04$; Fig. 3 and Supplementary Table S9).

Finally, to assess the robustness of our results further, we used TWAS results from DLPFC [5] on a larger reference panel from the PsychENCODE Consortium [31]. TRS from DLPFC was also significantly associated with ADHD ($P_{empirical}$=9.9e-05, *pseudo*-R^2 = 0.028), remained significant in the sensitivity analysis considering only the most significant gene per locus (Supplementary Table S7), and combined with PRS improved the fit of the model over PRS alone (P = 6.9e-05), reaching a *pseudo*-R^2 of 0.046 in the combined model (Table 1).

DISCUSSION

To our knowledge, this is the first study to construct TRSs for ADHD based on observed expression data. We undertook TWAS on ADHD using the latest ADHD GWAS-MA summary statistics and 14 expression reference panels across a range of brain tissues and whole blood to prioritize genes and construct transcriptome-based risk scores for the disorder [5, 30]. Given that a substantial proportion of GWAS hits demonstrate gene regulation effects [20], risk scores based on eQTL variants integrate biological information for disease prediction, link genetic associations to biological disease mechanisms and provide an additional layer of biological interpretability.

We found 56 genes showing transcriptome-wide significant association with ADHD, of which 31 did not overlap with previously described GWAS loci or TWAS results by Demontis et al. [5]. The variability observed between studies could be mainly due to differences in the tissues and methods used to construct the expression reference panel, as Demontis et al. used a different eQTL reference panel in DLPFC from the PsychENCODE Consortium [31], and we used GTEx v8 data on 14 tissues based on JTI methodology, to exploit the power of multi-tissue transcriptomes to improve prediction accuracy. Among the new genes identified, *NAA80*, associated with ADHD in all expression reference panels, encodes an actin-specific N-acetyltransferase that may play a role

in excitatory synapses, which is consistent with alterations in the reorganization of synaptic actin described in neurodevelopmental disorders [49]. *PNPLA2* was transcriptome-wide significant in all the expression reference panels but substantia nigra and differentially expressed in the PBMCs with consistent direction of effect. It encodes a lipase related with obesity, highly comorbid with ADHD [50], and was recently pointed as one of the most high-confidence causal genes for ADHD [17]. Other interesting transcriptome-wide significant signals included several long non-coding RNA, a group of regulatory RNA involved in neural differentiation and synaptic plasticity that have been related with psychiatric disorders [51, 52], or target genes for miRNA-34, previously associated with ADHD [53]. This miRNA family participates in neuronal differentiation and synaptogenesis [54] and is among the most upregulated miRNAs during dopaminergic differentiation [55].

We selected a subset of relevant genes from TWAS results and constructed TRSs using microarray expression data in PBMCs from 222 individuals with ADHD and 269 controls. TRSs based on TWAS results from most of the brain tissues were associated with ADHD, with individuals with ADHD carrying a higher burden of TRS than controls. In contrast, no association was found when the TRS was constructed based on TWAS results in whole blood, which suggests that the performance of the TRS is optimized when selecting genes from expression reference panels in relevant tissues for the disorder. This is likely due to the eQTL tissue specificity previously described [56] and is in line with our findings where the TRSs that surpassed multiple comparison corrections and sensitivity analyses were constructed from expression reference panels in four brain areas associated with ADHD, namely cortex, frontal cortex, caudate basal ganglia and amygdala [57–59].

Genes included in the best-performing TRSs provide additional information to prioritize candidates for further investigation of biological mechanisms underlying ADHD. For example, all TRSs associated with ADHD include three genes, *PNPLA2, PLK1S1* and *GMPPB*, previously associated with ADHD and/or other neurodevelopmental disorders [17, 60–62]. Of them, *PNPLA2*, already discussed as one of the top hits in the multiple-tissue TWAS, is the only gene with a high colocalization score in three out of the four tissues studied, and seems to play an important role in the TRS_{col} of amygdala, caudate basal ganglia and frontal cortex, which points it as one of the most promising candidate genes. Besides, we also highlight other genes with high colocalization scores in different tissues: the *GIGYF2* gene, significantly associated with ADHD across the lifespan [63], which contributes to the TRS_{col} from both cortex and frontal cortex, the *SLC25A22* gene, which encodes a glutamate transporter with strong expression in the developing brain, that adds important weight to the TRS_{col} from caudate basal ganglia and frontal cortex, and *CKS2*, a cyclin-dependent kinase involved in the control neuronal differentiation [64], which contributes to the TRS_{col} from the amygdala. Interestingly, according to the GWAS catalog genetic variants in these genes and others included in TRS_{col} (i.e. *CTNNB1, COPA, CCDC71* and *BLOC1S2*) have been associated with psychiatric disorders (e.g. schizophrenia, externalizing behavior, smoking initiation, autism spectrum disorder, anorexia nervosa, depression and anxiety disorder), cognitive function (e.g. intelligence, educational attainment and mathematical ability) or ADHD comorbid somatic traits like obesity or extreme body mass index, suggesting a potential importance of these genes in the context of ADHD and its comorbid conditions.

For most of the brain tissues, the TRSs constructed under stricter TWAS *P*-value thresholds showed clear evidence of better performance and stronger associations with ADHD, a pattern similar to the one observed for TRS in amyotrophic lateral sclerosis based on observed expression data [21]. This contrasts with the pattern of association found for PRSs or imputed gene expression-based risk scores, where the variance explained tends to increase as more relaxed *P*-value thresholds are used [24, 26]. These different patterns could result from methodological limitations in TWAS that hamper the statistical power of TRSs from observed gene expression, especially when more genes with weaker association signals are included in the analysis. These could include noisy beta estimates in TWAS due to the limited sample size of both GWAS-MA on ADHD and GTEx v8 reference panels [5, 30] , or false positive associations in the TWAS due to pleiotropy or linkage disequilibrium. Also, TRSs-based on observed expression data may reflect a dynamic layer of biological regulation that could explain the difference found. While using predicted expression data provides an accurate estimate of the genetic risk conferred via cis-regulated gene expression, TRSs constructed on observed expression datasets may be also attributable to other influences including trans-acting genetic effects or environmental effects and may provide a closer connection to the disorder than standard PRSs or TRSs calculated on imputed gene expression levels. This is consistent with findings showing that a substantial proportion of gene expression heritability may not result from common cis-eQTL SNPs, but rather stem from trans-variants which may act predominantly in a tissue-specific manner, and points to the need for further studies on the trans-regulatory landscape [65].

In agreement with a previous study in depression [66], TRSs were uncorrelated with genome-wide PRS. This lack of correlation may highlight that TRSs based on observed gene expression data capture more information than cis-eQTL genetic risk variants, such as trans-eQTL, environment factors or epigenetics, as well as interaction effects between genes and environment, among others. In addition, compared with PRS-only models, models combining PRS and TRSs provided substantial improvement in model fit for ADHD, which supports that gene expression explains additional phenotypic variance for the disorder than PRSs and is consistent with the complementary predictive potential of genetic and transcriptomic signatures [24].

Apart from TWAS, other methods have been designed to prioritize likely causal genes by combining genomic, transcriptomic, and other regulatory and functional information including colocalization methods, that use a Bayesian framework to infer whether a regulatory SNP is also responsible for the association with a trait of interest, or summary-based Mendelian randomization (SMR), that combines GWAS and eQTL data to prioritize target genes with evidence for causal or pleiotropic effects. In order to narrow down the number of genes identified by TWAS and included in the TRS analyses, we assessed colocalization and found that the signal for 14 out of the 56 genes identified in the TWAS was supported by the colocalization analyses. This low convergence between TWAS and colocalization signals is consistent with other studies [34] and may result from several factors including failure to identify either the phenotype-SNP association or the expression-SNP association, given the relatively limited sample size of both GWAS-MA on ADHD and GTEx v8 reference panels [5, 30], especially for brain areas. Also, colocalization signals may arise from direct genetic effects, while TWAS signals may result from complex interactions between multiple genes and genetic variants [33]. When restricting best-performing TRSs to the colocalized genes, despite a reduction of at least the 70% in the number of genes included, the association signal remained and even became stronger for amygdala, caudate basal ganglia and frontal cortex. These results are in line with previous studies [21, 23] and point to the high specificity of the colocalization approach [33].

The results of the present study, however, should be interpreted in the context of several strengths and limitations: (i) Due to linkage disequilibrium, a single genetic variant might point to several TWAS associations in the same locus. For that reason, sensitivity analysis using only the most significant gene in each

locus was performed to discard artefactual inflation in the TRS analysis. However, considering that genes located in the same region are not necessarily involved in the same biological processes and given the difficulty to determine which ones really contribute to the phenotype, enrichment analysis were performed including all significant genes from TWAS, which could have potentially biased these results. (ii) In this study, the PRS failed to approach the performance of the two best-performing TRSs (from cortex and frontal cortex), which suggests that TRSs may potentially outperform PRSs and provide a closer physiological picture of the disorder; (iii) While TRSs differences may reflect distinct molecular pathways captured by each of the tissues considered, the variability in sample size between the expression reference panels may limit our ability to compare TRSs results across tissues. Besides, in the present study we used multiple-tissue TWAS. Although this method shows improved prediction over single tissue approaches and it underscores specific genes overlapping between tissues [29], additional approaches are required to identify tissue-specific expression profiles; (iv) We found significant correlation between the TRSs associated with ADHD, probably, in part, because the different brain areas from which they were constructed are both functionally and structurally connected. However, selecting genes for the construction of TRSs based on multiple-tissue TWAS results, where information is borrowed across transcriptomes of different tissues, may also contributed to artificially inflate these correlations; (v) The positive results obtained for TRSs capturing expression in brain areas implicated in ADHD but not in whole blood suggests that the relevance of the tissue to the outcome may also influence the predictive performance of the TRS; (vi) Although TRS constructed on real expression datasets may provide a closer connection to the disorder and may capture gene expression within a range of contexts, they may be influenced by confounding factors such as gender, age, comorbid disorders or medication. We frequency sex-matched ADHD cases and controls and restricted the clinical sample to ADHD medication-naïve adult subjects, which is a major strength of our study design that may allow us to identify transcriptomic signatures that might be neglected by broader study designs. We cannot discard residual confounding by other factors not available. In the same line, observed differential expression associated with ADHD may reflect both a gene's causal role in the disorder or be consequence of the disorder itself. However, given that genetically-inferred differential expression from TWAS may not be susceptible to reverse causation, we think that most genes included in our TRSs are more prone to it because of the disorder rather than consequence; (vii) Further studies considering low frequency and rare variants and using more unbiased profiling methods, such as RNA sequencing techniques, may allow the inclusion of novel and low abundance transcripts and relevant genes to improve the predictive power of TRS approaches. In addition, as resources used for eQTL mapping expand in sample size and integrate additional regulatory and epigenetic data, we expect TRS performance to improve. (viii) Finally, longitudinal studies will be required to disentangle the performance of TRSs across the lifespan and their role on the remittent and/or persistent form of the disorder.

In conclusion, we found association between ADHD and TRSs in PBMCs constructed using TWAS results from multiple brain areas implicated in the disorder, showing that individuals with ADHD carry a higher burden of TRSs than controls. TRSs combined with PRS increased significantly the proportion of variance explained of ADHD over genome-wide PRS alone, which points to the complementary predictive potential of genetic and transcriptomic signatures and support that integrating biological information may benefit standard PRS prediction approaches. Through this approach that leverages GWAS summary statistics, multi-tissue cis-eQTL reference panels and target sample gene expression data we underscore the potential of utilizing transcriptomic information to improve risk prediction and provide deeper insight into the molecular mechanisms underlying ADHD.

REFERENCES

1. Song P, Zha M, Yang Q, Zhang Y, Li X, Rudan I. The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis. J Glob Health. 2021;11:1–9.
2. Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Primers. 2015;1:15020.
3. Brikell I, Larsson H, Lu Y, Pettersson E, Chen Q, Kuja-Halkola R, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. Mol Psych. 2018;25:1809–21.
4. Larsson H, Chang Z, D'Onofrio BM, Lichtenstein P. The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan. Psychological Med. 2014;44:2223–9.
5. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nat Genet. 2023;55:198–208.
6. Ronald A, de Bode N, Polderman TJC. Systematic Review: How the Attention-Deficit/Hyperactivity Disorder Polygenic Risk Score Adds to Our Understanding of ADHD and Associated Traits. J Am Acad Child Adolesc Psych. 2021;60:1234–77.
7. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.
8. Pujol-Gualdo N, Sánchez-Mora C, Ramos-Quiroga JA, Ribasés M, Soler Artigas M. Integrating genomics and transcriptomics: Towards deciphering ADHD. Eur Neuropsychopharmacol. 2021;44:1–13.
9. Garcia-Martínez I, Sánchez-Mora C, Soler Artigas M, Rovira P, Pagerols M, Corrales M, et al. Gene-wide Association Study Reveals RNF122 Ubiquitin Ligase as a Novel Susceptibility Gene for Attention Deficit Hyperactivity Disorder. Sci Rep. 2017;7:1–13.
10. De Jong S, Newhouse SJ, Patel H, Lee S, Dempster D, Curtis C, et al. Immune signatures and disorder-specific patterns in a cross-disorder gene expression analysis. Br J Psych. 2016;209:202–8.
11. Lorenzo G, Braun J, Muñoz G, Casarejos MJ, Bazán E, Jimenez-Escrig A. RNA-Seq blood transcriptome profiling in familial attention deficit and hyperactivity disorder (ADHD). Psych Res. 2018;270:544–6.
12. Marín-Méndez JJ, Patiño-García A, Segura V, Ortuño F, Gálvez MD, Soutullo CA. Differential expression of prostaglandin D2 synthase (PTGDS) in patients with attention deficit-hyperactivity disorder and bipolar disorder. J Affect Disord. 2012;138:479–84.
13. McCaffrey TA, St. Laurent G, Shtokalo D, Antonets D, Vyatkin Y, Jones D, et al. Biomarker discovery in attention deficit hyperactivity disorder: RNA sequencing of whole blood in discordant twin and case-controlled cohorts. BMC Med Genomics. 2020;13:160.
14. Mortimer N, Sánchez-Mora C, Rovira P, Vilar-Ribó L, Richarte V, Corrales M, et al. Transcriptome profiling in adult attention-deficit hyperactivity disorder. Eur Neuropsychopharmacol. 2020;41:160–6.
15. Nuzziello N, Craig F, Simone M, Consiglio A, Licciulli F, Margari L, et al. Integrated Analysis of microRNA and mRNA Expression Profiles: An Attempt to Disentangle the Complex Interaction Network in Attention Deficit Hyperactivity Disorder. Brain Sci. 2019;9:288.
16. Sánchez-Mora C, Soler Artigas M, Garcia-Martínez I, Pagerols M, Rovira P, Richarte V, et al. Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. Neuropsychopharmacol. 2019;44:890.
17. Fahira A, Li Z, Liu N, Shi Y. Prediction of causal genes and gene expression analysis of attention-deficit hyperactivity disorder in the different brain region, a comprehensive integrative analysis of ADHD. Behavioural Brain Res. 2019;364:183–92.
18. Liao C, Laporte AD, Spiegelman D, Akçimen F, Joober R, Dion PA, et al. Transcriptome-wide association study of attention deficit hyperactivity disorder identifies associated genes and phenotypes. Nat Commun. 2019;10:1–7.
19. Qi X, Wang S, Zhang L, Liu L, Wen Y, Ma M, et al. An integrative analysis of transcriptome-wide association study and mRNA expression profile identified candidate genes for attention-deficit/hyperactivity disorder. Psych Res. 2019;282:112639.
20. Porcu E, Rüeger S, Lepik K, Agbessi M, Ahsan H, Alves I, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun. 2019;10:1–12.
21. Pain O, Jones A, Khleifat AA, Agarwal D, Hramyka D, Karoui H. Harnessing Transcriptomic Signals for Amyotrophic Lateral Sclerosis to Identify Novel Drugs and Enhance Risk Prediction. MedRxiv Preprint. 2023. 2023. https://doi.org/10.1101/2023.01.18.23284589.

22. Park YH, Hodges A, Simmons A, Lovestone S, Weiner MW, Kim SY, et al. Association of blood-based transcriptional risk scores with biomarkers for Alzheimer disease. Neurology Genetics. 2020;6:e517.
23. Marigorta UM, Denson LA, Hyams JS, Mondal K, Prince J, Walters TD, et al. Transcriptional Risk Scores link GWAS to eQTL and Predict Complications in Crohn's Disease. Nat Genet. 2017;49:1517.
24. Pain O, Glanville KP, Hagenaars S, Selzam S, Fürtjes A, Coleman JRI, et al. Imputed gene expression risk scores: a functionally informed component of polygenic risk. Hum Mol Genet. 2021;30:727.
25. Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet. 2018;50:538.
26. Rodriguez-López J, Arrojo M, Paz E, Páramo M, Costas J. Identification of relevant hub genes for early intervention at gene coexpression modules with altered predicted expression in schizophrenia. Prog Neuro-Psychopharmacol Biol Psych. 2020;98:109815.
27. Miles AE, Dos Santos FC, Byrne EM, Renteria ME, McIntosh AM, Adams MJ, et al. Novel polygenic risk score links depression-related cortical transcriptomic changes to brain morphology and depressive symptoms in men. Neuropsychopharmacol. 2021;46:2304–11.
28. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9:1825.
29. Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. Nat Genet. 2020;52:1239–46.
30. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020;369:1318–30.
31. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. Science. 2018;362:eaat8464.
32. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
33. Hukku A, Sampson MG, Luca F, Pique-Regi R, Wen X. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. Am J Hum Genet. 2022;109:825–37.
34. Wang X, Selvaggi C, Weng L-C, Jurgens SJ, Choi SH, Jha A, et al. Clinical and Therapeutic Applications of Individual-level Tissue-Specific Imputed Transcriptomes. MedRxiv. 2022. https://doi.org/10.1101/2022.11.23.22282644.
35. Wang J, Huang D, Zhou Y, Yao H, Liu H, Zhai S, et al. CAUSALdb: A database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucl Acids Res. 2020;48:D807–16.
36. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: Efficient variable selection using summary data from genome-wide association studies. Bioinformatics 2016;32:1493–501.
37. Greenbaum J, Deng HW. A Statistical Approach to Fine Mapping for the Identification of Potential Causal Variants Related to Bone Mineral Density. J Bone Miner Res. 2017;32:1651–8.
38. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, et al. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. Genetics. 2015;200:719–36.
39. Wen X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. Ann Appl Stat. 2016;10:1619–38.
40. Cabana-Domínguez J, Soler Artigas M, Arribas L, Alemany S, Vilar-Ribó L, Llonga N, et al. Comprehensive analysis of omics data identifies relevant gene networks for Attention-Deficit/Hyperactivity Disorder (ADHD). Transl Psych. 2022;12:1–9.
41. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010;26:2363–7.
42. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics - A bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25:415–6.
43. Schäfer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics. 2005;21:754–64.
44. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019;8:1–6.
45. Lee SH, Goddard ME, Wray NR, Visscher PM. A Better Coefficient of Determination for Genetic Profile Analysis. Genet Epidemiol. 2012;36:214–24.
46. Galwey NW. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. Genet Epidemiol. 2009;33:559–68.
47. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10:1–10.
48. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.
49. Yan Z, Kim E, Datta D, Lewis DA, Soderling SH. Synaptic Actin Dysregulation, a Convergent Mechanism of Mental Disorders? J Neurosci. 2016;36:11411.
50. Cortese S. The association between ADHD and obesity: Intriguing, progressively more investigated, but still puzzling. Brain Sci. 2019;9:256.
51. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22:96–118.
52. Jovčevska I, Videtič Paska A. Neuroepigenetics of psychiatric disorders: Focus on lncRNA. Neurochem Int. 2021;149:105140.
53. Garcia-Martínez I, Sánchez-Mora C, Pagerols M, Richarte V, Corrales M, Fadeuilhe C, et al. Preliminary evidence for association of genetic variants in pri-miR-34b/c and abnormal miR-34c expression with attention deficit and hyperactivity disorder. Transl Psych. 2016;6:e879.
54. Jauhari A, Yadav S. MiR-34 and MiR-200: Regulator of Cell Fate Plasticity and Neural Development. NeuroMolecular Med. 2019;21:97–109.
55. De Gregorio R, Pulcrano S, De Sanctis C, Volpicelli F, Guatteo E, von Oerthel L, et al. miR-34b/c Regulates Wnt1 and Enhances Mesencephalic Dopaminergic Neuron Differentiation. Stem Cell Rep. 2018;10:1237–50.
56. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: The muTHER study. PLoS Genet. 2011;7:e1002003.
57. Hess JL, Akutagava-Martins GC, Patak JD, Glatt SJ, Faraone SV. Why is there selective subcortical vulnerability in ADHD? Clues from postmortem brain gene expression data. Mol Psych. 2018;23:1787–93.
58. Postema MC, Hoogman M, Ambrosino S, Asherson P, Banaschewski T, Bandeira CE, et al. Analysis of structural brain asymmetries in attention-deficit/hyperactivity disorder in 39 datasets. J Child Psychol Psych Allied Discip. 2021;62:1202–19.
59. Hoogman M, Muetzel R, Guimaraes JP, Shumskaya E, Mennes M, Zwiers MP, et al. Brain imaging of the cortex in ADHD: A coordinated analysis of large-scale clinical and population-based samples. Am J Psych. 2019;176:531–42.
60. Alonso-Gonzalez A, Calaza M, Rodriguez-Fontenla C, Carracedo A. Novel gene-based analysis of ASD GWAS: Insight into the biological role of associated genes. Front Genet. 2019;10:733.
61. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
62. Panicucci C, Fiorillo C, Moro F, Astrea G, Brisca G, Trucco F, et al. Mutations in GMPPB presenting with pseudometabolic myopathy. JIMD Rep. 2018;38:23–31.
63. Rovira P, Demontis D, Sánchez-Mora C, Zayats T, Klein M, Mota NR, et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacol. 2020;45:1617–26.
64. Frontini M, Kukalev A, Leo E, Ng YM, Cervantes M, Cheng CW, et al. The CDK Subunit CKS2 Counteracts CKS1 to Control Cyclin A/CDK2 Activity in Maintaining Replicative Fidelity and Neurodevelopment. Dev Cell. 2012;23:356–70.
65. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012;44:1084–9.
66. Miles AE, Dos Santos FC, Byrne EM, Renteria ME, McIntosh AM, Adams MJ, et al. Transcriptome-based polygenic score links depression-related corticolimbic gene expression changes to sex-specific brain morphology and depression risk. Neuropsychopharmacol. 2021;46:2304–11.

ACKNOWLEDGEMENTS

The authors are grateful to patients and controls who kindly participated in this research. We are thankful to deCODE (G. Bragi Walters, Hreinn Stefansson and Kari Stefansson) for giving us access to the recent GWAS meta-analysis summary statistics of ADHD which include deCODE data.

AUTHOR CONTRIBUTIONS

JCD, MR and MSA conceived the project. CF, MC, VR, JARQ participated in the clinical assessment and in the recruitment of patients. JCD, LA and LVR participated in the RNA isolation and preparation of samples. JCD and NL undertook the statistical analyses. JCD, NL, LVR, SA, MR, and MSA participated in the study design and the discussion of results. JCD, LN, MR, and MSA participated in the manuscript preparation. All authors contributed to the interpretation of the findings and revised and approved the final version of the manuscript.

FUNDING

This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-1461, 2021SGR-00840), the Instituto de Salud Carlos III and co-funded by the European Union Found: Fondo Social Europeo Plus (FSE+) (P19/01224, PI20/00041, PI22/00464 and FI18/00285 to LVR, CP22/00128 to MSA and CP22/00026 to SA), the Ministry of Science, Innovation and Universities (IJC2018-035346-I

to MSA and RYC2021-031324-I to JCD), the Network Center for Biomedical Research (CIBER) to JCD, the European Regional Development Fund (ERDF), the European Union H2020 Programme (H2020/2014-2020) under grant agreements no. 728018 (Eat2beNICE), no. 848228 (DISCOvERIE) and no. 2020604 (TIMESPAN), the ECNP Network 'ADHD across the Lifespan' and the "la Marató de TV3" (202228-30 and 202228-31). ADB was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003) and NIH/NIMH (1R01MH124851-01). DD was supported by the Novo Nordisk Foundation (NNF20OC0065561) and the Lundbeck Foundation (R344-2020-1060).

COMPETING INTERESTS

JARQ was on the speakers' bureau and/or acted as consultant for Biogen, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Rubió, Uriach, Technofarma and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious, and Rubió. CF and VR have received fees to give talks for Shire/Takeda and Rubió. All other authors declare no biomedical financial interests or conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-023-02200-1.

Correspondence and requests for materials should be addressed to Judit Cabana-Domínguez, María Soler Artigas or Marta Ribasés.

Journal of Translational Medicine

RESEARCH

Open Access

Genome-wide multi-trait analysis of irritable bowel syndrome and related mental conditions identifies 38 new independent variants

Silvia Alemany[1,2,3*], María Soler-Artigas[1,2,3,4], Judit Cabana-Domínguez[1,2,3], Dana Fakhreddine[1], Natalia Llonga[1,2], Laura Vilar-Ribó[1,2,3], Amanda Rodríguez-Urrutia[1,2,3,5], Judit Palacio[2], Ana María González-Castro[6], Beatriz Lobo[6,7,8], Carmen Alonso-Cotoner[6,7,8,9], Magnus Simrén[10,11], Javier Santos[6,7,8,9], Josep Antoni Ramos-Quiroga[1,2,3,5] and Marta Ribasés[1,2,3,4*]

Abstract

Background Irritable bowel syndrome (IBS) is a chronic disorder of gut-brain interaction frequently accompanied by mental conditions, including depression and anxiety. Despite showing substantial heritability and being partly determined by a genetic component, the genetic underpinnings explaining the high rates of comorbidity remain largely unclear and there are no conclusive data on the temporal relationship between them. Exploring the overlapping genetic architecture between IBS and mental conditions may help to identify novel genetic loci and biological mechanisms underlying IBS and causal relationships between them.

Methods We quantified the genetic overlap between IBS, neuroticism, depression and anxiety, conducted a multi-trait genome-wide association study (GWAS) considering these traits and investigated causal relationships between them by using the largest GWAS to date.

Results IBS showed to be a highly polygenic disorder with extensive genetic sharing with mental conditions. Multi-trait analysis of IBS and neuroticism, depression and anxiety identified 42 genome-wide significant variants for IBS, of which 38 are novel. Fine-mapping risk loci highlighted 289 genes enriched in genes upregulated during early embryonic brain development and gene-sets related with psychiatric, digestive and autoimmune disorders. IBS-associated genes were enriched for target genes of anti-inflammatory and antirheumatic drugs, anesthetics and opioid dependence pharmacological treatment. Mendelian-randomization analysis accounting for correlated pleiotropy identified bidirectional causal effects between IBS and neuroticism and depression and causal effects of the genetic liability of IBS on anxiety.

Conclusions These findings provide evidence of the polygenic architecture of IBS, identify novel genome-wide significant variants for IBS and extend previous knowledge on the genetic overlap and relationship between gastrointestinal and mental disorders.

Keywords Irritable bowel syndrome (IBS), Neuroticism, Depression, Anxiety, Multi-trait genome-wide association study (MTAG)

*Correspondence:
Silvia Alemany
silvia.alemany@vhir.org
Marta Ribasés
marta.ribases@vhir.org
Full list of author information is available at the end of the article

Introduction

Irritable bowel syndrome (IBS) is one of the most prevalent disorders of gut-brain interaction with a population lifetime risk of 11% [1] and a point prevalence of 4.1% according to the strict Rome IV criteria [2]. IBS research is extremely challenging due to the multifactorial etiology of the disease and the heterogeneity of patients, who present high comorbidity rates for mental disorders, particularly, anxiety and depression, which impacts negatively on the patients' quality of life [1, 3, 4].

A recent systematic review revealed that the prevalence of anxiety and depression symptoms among IBS patients is 39.1% and 28.8%, respectively [5]. In addition, IBS has been associated with more severe depressive symptoms compared to healthy controls and, when co-existing with psychiatric disorders, gastrointestinal symptoms are more severe and disabling [6–11]. This close association between IBS, anxiety and depression is also supported by neuroimaging studies and might be related to the bi-directional communication between the brain and the digestive system, termed the brain-gut-axis, which occurs through microbiota, neural, neuroimmune and neuroendocrine pathways [12–14]. This idea agrees with evidence indicating that psychiatric interventions, including antidepressants or cognitive-behavioral therapy, improve IBS patients functioning and suggests that common pathophysiological mechanisms may be underlying these conditions [15].

IBS, anxiety and depression are partly determined by a genetic component and show substantial heritability ranging from 6% for IBS to 30%-50% for anxiety and depression [16–18]. The largest genome-wide association study (GWAS) on IBS conducted to date included 53,400 cases and 433,201 controls and identified six genome-wide significant single nucleotide polymorphisms (SNPs) [18] which represents an improvement over the previous study, identifying four independent genome-wide significant SNPs [19]. Interestingly, among 173 traits, three mental conditions (neuroticism, depression and anxiety) were the most genetically correlated traits with IBS [18]. Despite these strong genetic correlations, the genetic underpinnings explaining the high rates of comorbidity between IBS and mental conditions remain largely unclear and there are no conclusive data on the temporal and causal relationship between them [18, 19].

In the present study we investigated the shared genetic architecture and the nature of the relationship between IBS and three highly genetically correlated conditions (neuroticism, depression and anxiety) using summary statistics of the largest GWAS datasets available so far by (i) estimating the genetic correlation and overlap between them, (ii) conducting a Multi-Trait Analysis of GWAS (MTAG) to identify novel genetic loci for IBS and (iii) performing downstream analyses to explore the overlaping genetic basis with other disorders and traits as well as causal relationships between them.

Materials and methods

Samples

We used publicly available SNP-level GWAS summary statistics for IBS [18], neuroticism [20], depression [21] and anxiety (Table 1). For further details see Additional file 1: Note 1.

SNP-based heritability, genetic correlation and overlap

SNP heritability (h^2_{SNP}) and pair-wise genetic correlation between IBS and each mental condition was calculated using linkage disequilibrium score regression (LDSC) analysis [22]. Conversion of h^2_{SNP} estimates from observed to liability scale was done using a population prevalence of 11%, 25%, 30% and 14% for IBS, neuroticism, depression and anxiety, respectively. Polygenic overlap between IBS and each mental condition was quantified using MiXeR [23]. MiXeR caclulates the

Table 1 Summary of the GWAS datasets used in the current study

Phenotype	N cases	N controls	N total	N effective [a]	GWAS genome-wide significant SNPs[b]	References
IBS	53,400	433,201	486,601	190,159	6	[18]
Neuroticism	–	–	390,278[c]	390,278	136[d]	[20]
Depression	170,756	329,443	500,199[c]	449,856	102[d]	[21]
Anxiety nerves or GAD	16,730	101,021	117,751	57,412	1	UKBB phenotype code: 20544_15

GAD generalized anxiety disorder; *UKBB* UK Biobank

[a] N effective sample sizes were calculated following the equation: Neff = 4/(1/Ncases + 1/Ncontrols)

[b] Number of genome-wide significant independent SNPs

[c] Sample size excluding the 23andMe cohort

[d] Genome-wide significant SNPs including the 23andMe cohort

number of trait-influencing SNPs for each trait (univariate model) and for both traits (bivariate model) and the proportion of variants with concordant direction of effects for both traits. The proportion of SNPs shared by two traits is indicated by the Dice coefficient. Model fit was assessed using the Akaike Information Criterion (AIC). For further details see Additional file 1: Note 2.

Multi-trait analysis of GWAS (MTAG)

To identify new loci for IBS, SNP-level GWAS for IBS, neuroticism, depression and anxiety were meta-analyzed using MTAG [24]. MTAG estimates trait-specific effects from GWAS summary statistics of several traits genetically correlated while accounting for sample overlap across the discovery samples [24]. To discard inflation in our results we calculated the max-false discovery rate (max-FDR) using default settings as previously described [24, 25]. The LDSC intercept was used to quantify inflation resulting from confounding bias [22].

Independent SNPs from MTAG-IBS results (P-value < 5E-08) were identified through clumping ($r^2 = 0.05$, kb = 5000) using the 1000 Genomes Project Phase 3 European reference panel (http://www.internationalgenome.org/) and PLINK1.09 as described by Eijsbouts et al. [18]. We defined loci as a 1Mb region centered around the most significant variant (lead variant) and we carried out conditional analyses to confirm independence between lead and any other variant identified in the clumping step (secondary variants) within each locus (i.e. within 1Mb and $r^2 < 0.05$) using COJO implemented in Genome-wide Complex Trait Analysis (GCTA) [26]. For further details on conditional analysis see Additional file 1: Note 3.

Credible variants and functional annotation

Sets of credible variants (credible-sets) were identified by fine-mapping the independent lead SNPs of MTAG-IBS using three different tools, FINEMAP 1.3.1 [27], PAINTOR v3.0 [28] and CAVIARBF v.0.2.1 [29] following the pipeline available elsewhere [30]. Variants located in a region of 1Mb around the lead SNPs were included in the analysis and we assumed that there was only one causal variant per locus. We used the recommended parameters of each tool and only variants identified by all three methods were considered. Functional annotation of the credible variants was conducted using FUMA [31]. For further details see Additional file 1: Note 4.

Gene-based and gene-set analyses of MTAG-IBS results

Gene-based and gene-set analyses of MTAG-IBS associated SNPs were performed using MAGMA v1.08 [32] implemented in FUMA [31]. Tissue specific gene expression was explored using MAGMA gene-property analysis of expression data from GTEx v8 and BrainSpan available in FUMA (databases detailed in Additional file 1: Note 5). All gene sets were obtained from the Molecular Signatures Database (MSigDB v6.2) and included GO, KEGG, BIOCARTA and Reactome representing a total of 11,960 gene sets. The Bonferroni-corrected significance threshold for gene-based analysis was 0.05/18135 genes = 2.7571E-06 and for gene-set analysis was 0.05/11960 gene sets = 4.18E-06.

Drug target identification

To explore whether finemapped genes related with IBS were enriched for target genes of drugs (druggable genes) we performed enrichment analysis based on information from the PharmGKB using WebGestAlt [33]. Identified drugs were classified according to available information from the Anatomical Therapeutic Chemical (ATC) classification system.

Partitioned heritability and genetic correlations

We partitioned h^2_{SNP} of MTAG-IBS results by functional annotation categories using stratified LDSC [34]. We calculated whether any of the 28 specific genomic categories included in the analysis was enriched for variants that contribute to h^2_{SNP}. Annotations for these functional genomic categories (e.g. coding or regulatory regions) were obtained from LDSC website (https://github.com/bulik/ldsc/wiki/Partitioned-Heritability) and included coding; intron; promoter; 3′5′ untranslated region; digital genomic footprint; transcription factor binding site; chromHMM and Segway annotations for six cell lines; DNase I hypersensitivity sites; H3K4me1, H3K4me3 and H3K9ac marks; two sets of H3K27ac marks; super-enhancers; conserved regions in mammals; and FANTOM5 enhancers (further details in Additional file 1: Note 6). We focused on categories extended by 500 bp in either direction. Enrichment/depletion of heritability in each category is calculated as the proportion of heritability attributable to SNPs in the specified category divided by the proportion of total SNPs annotated to that category. The Bonferroni-corrected significance threshold was 0.05/28 annotations = 0.0018.

We explored genetic correlations between our MTAG-IBS results and gastrointestinal, immunological and psychiatric disorders using LDSC analysis [22]. We selected all GWAS summary statistics of gastrointestinal/abdominal, immunological/systemic (UK Biobank: 21 phenotypes) and psychiatric disorders (PGC: 7 phenotypes) available in the MR-Base database (Additional file 3: Table S14) [35]. We used GWAS summary statistics including both males and females of European ancestry. If several GWAS were available for the same disorder, we chose the study with the largest effective sample size

(N effective = 4/(1/Ncases + 1/Ncontrols)). The Bonferroni-corrected significance threshold used was 0.05/28 traits = 0.0018.

Causal analysis using summary effect estimates (CAUSE)

Causal relationships between IBS and correlated traits were assessed considering independent variants ($r^2 = 0.05$; kb = 5000) associated with the exposure with $P < 1.0E-03$ using CAUSE [36]. Bidirectional relationships were tested considering IBS as exposure and depression, anxiety or neuroticism as outcomes and vice-versa. Given that standard errors, required by CAUSE, were not available from the largest study on neuroticism to date [37], we used the GWAS dataset on neuroticism by Luciano et al. in 329,821 subjects as an alternative [38]. The strengths of CAUSE involve accounting for correlated horizontal pleiotropic effects (i.e. when a variant affects the outcome and the mediator through shared heritable factors) and using a less stringent significance threshold ($P < 1.0E-3$) allowing the incorporation of more variants to the analyses. CAUSE compares two nested models, a sharing and a causal model. Both models allow for horizontal pleiotropy (correlated pleiotropy (eta)) but only the casual model includes a causal effect parameter (gamma). The sharing and the causal model are compared against a null model and against each other. Model comparisons are carried out using the expected log pointwise posterior density (ELPD), a Bayesian model comparison approach that estimates how well the posterior distributions of a particular model are expected to predict a new set data. When $P < 0.05$ the second model fits the data better than the first model. There is evidence of causal effects when the causal model represents a significant improvement in the model fit of the sharing model.

For further details see Additional file 1: Note 7.

Results

SNP-based heritability, genetic correlation and overlap

The latest GWAS on IBS [18], neuroticism [20], depression [21] and anxiety used herein are summarized in Table 1 and Additional file 1: Note 1. The estimated SNP heritability (h^2_{SNP}) was 6.9% (SE = 0.004) for IBS, 14.6% (SE = 0.005) for neuroticism, 9.9% (SE = 0.004) for depression and 8.3% (SE = 0.011) for anxiety (Table 2). We found evidence of strong genetic correlation between IBS and all three mental conditions, ranging from 53 to 68% (Table 2). Univariate MiXeR analysis revealed that IBS and neuroticism were highly polygenic, with around twelve thousand variants explaining 90% of SNP heritability (12,438 variants for IBS and 12,308 for neuroticism; Additional file 3:Table S1a). Bivariate MiXeR analysis showed that the majority of the variants influencing IBS were shared with neuroticism (10,793 (SE = 1094) out of 12,438 (SE = 1305) variants, Dice coefficient = 0.87), with a high proportion of variants being concordant (71%) (Additional file 3: Table S1a and Additional file 2: Figure S1). Unfortunately, MiXeR was unable to accurately quantify the genetic overlap between IBS and depression or anxiety according to the Akaike Information Criterion (AIC; Additional file 3: Table S1b).

Multi-trait analysis of GWAS (MTAG)

To identify novel loci for IBS, we combined the summary statistics from the GWAS on IBS, neuroticism, depression and anxiety using MTAG, increasing the estimated effective sample size from 486,601 in the original IBS dataset to 887,490. The max-FDR of MTAG-IBS analysis was low (0.020) suggesting no inflation, consistent with the similar mean chi-square values for the different GWAS, ranging from 1.08 for anxiety to 1.69 for neuroticism. There was no evidence of residual stratification or confounding leading to an inflation of test statistics (LD Score regression intercept = 0.857, SE = 0.009, See Additional file 2: Figure S2).

After MTAG analysis, the number of genome-wide significant SNPs for IBS increased from six in the original GWAS to 42 independent SNPs in 37 loci ($r^2 < 0.05$ between variants within each locus defined as regions of 1Mb) in the current study (Fig. 1, Additional file 2: Figure S3, Additional file 3: Table S2, S3). Five loci in chromosomes 5, 6, 11 and 18 (there were 2 loci in chromosome 18) had one or more secondary variants (i.e. in each locus there were more than one independent genome-wide significant variants). After conditional analysis to confirm

Table 2 Genetic correlation estimates for IBS and neuroticism, depression and anxiety using Linkage Disequilibrium Score Regression (LDSC)

Trait 1	Trait 2	Genetic Correlation	SE	Z	*P*-value	Intercept (SE)	Trait 1 h^2 (SE)	Trait 2 h^2 (SE)
IBS	Neuroticism	0.526	0.027	19.298	5.54E-83	1.013 (0.013)	0.069 (0.004)	0.146 (0.005)
IBS	Depression	0.587	0.026	22.714	3.23E-114	0.992 (0.01)	0.069 (0.004)	0.099 (0.004)
IBS	Anxiety	0.677	0.065	10.360	3.75E-25	0.999 (0.74)	0.068 (0.004)	0.083 (0.011)

SE, standard error; h^2 heritability

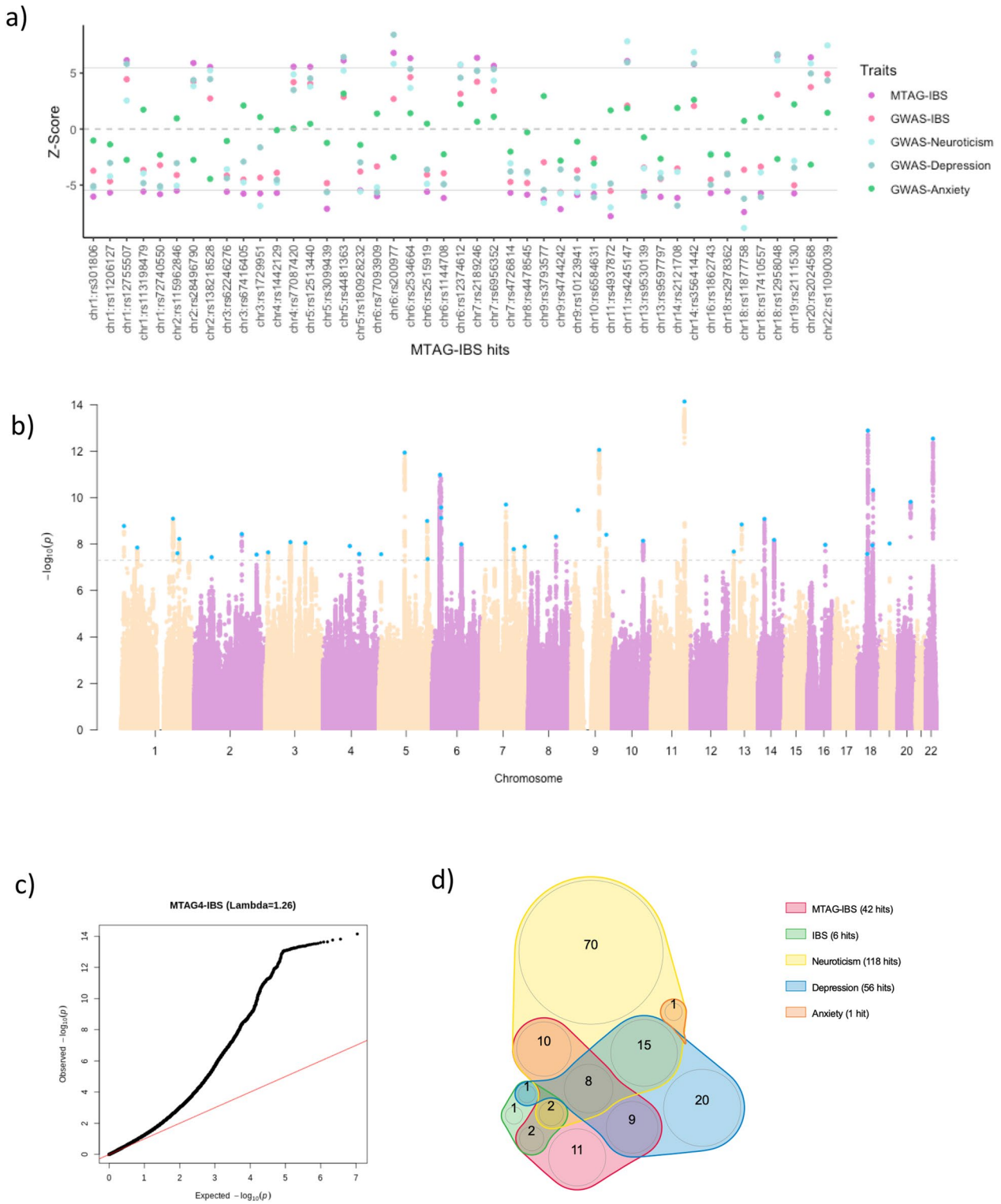


Fig. 1 MTAG results of IBS and overlap with previous GWAS on IBS, neuroticism, depression and anxiety. **a** Z-scores of MTAG-IBS and original GWAS on IBS, neuroticism, depression and anxiety for each of the independent lead SNPs (n = 42) found in MTAG-IBS results. Dotted grey line indicates 0 Z-score and solid grey lines indicate statistical significance at P < 5-E08. **b** Manhattan plot of the MTAG-IBS results. Dotted grey line indicates statistical significance at P < 5-E08. **c** QQ plot of the MTAG-IBS results. **d** Venn diagram depicting overlap among MTAG-IBS independent lead SNPs and genome-wide significant SNPs in the original GWAS

independence, secondary variants remained significant in the loci in chromosome 6 (4 secondary variants), 5 (1 secondary variant) and 18 (1 secondary lead variant). The secondary variant in chromosome 11 was no longer significant after conditional analysis leaving only the lead variant in this locus (Table 3 and Additional file 3: Table S2).

Comparing these results with the ones originally described for IBS [18], 38 out of the 42 SNPs identified herein were novel for IBS and all of them showed consistent direction of the association (Fig. 1a and Additional file 3: Table S3). Of them, 11 were not previously associated with neuroticism, depression or anxiety (Fig. 1d). The remaining signals, 27 in total, were novel associated SNPs for IBS but previously reported for neuroticism and/or depression (Table 3, Fig. 1d) and overall showed consistent direction of association with that reported in the original studies (Fig. 1a). Of the six SNPs previously identified in IBS [18], four of them, on chromosome 3, 6, 9 and 11, were among the significant SNPs for IBS in the current study and the two additional ones, in chromosome 13, showed suggestive evidence of association ($P < 5E-07$; Table 3). Among top findings, we found lead SNPs nearby genes involved in transcriptional regulation, including non-coding RNAs (*RP11-629G13.1* and *MSH5-SAPCD1*), RNA splicing (*CELF4*), chromatin remodeling (*EP300* and *HIST1H3J*), mRNA transport (*FAM120A*) or nucleic acid binding (*TCF4* and *ELAVL2*), as well as in brain development (*TMEM161B*) or presynaptic activity (*PCLO*).

Credible variants and functional annotation

We identified a total of 1,818 Bayesian credible variants in the 37 independent loci for IBS (Additional file 3: Table S4). Their functional annotation revealed over-presentation of SNPs in introns (64.6%), intergenic regions (21.7%) or located in non-coding RNA (9.4%) (Fig. 2 and Additional file 3: Table S5). A total of 75% of the variants within credible sets were located in open chromatin regions (minimum chromatin state ≤ 7), 3% were likely to affect the binding of transcription factors (RegulomeDB scores from 1b to 2c) and 0.05% may be deleterious (Combined Annotation Dependent Depletion (CADD) score > 12.37) (Fig. 2 and Additional file 3: Table S5). Forty-eight variants were previously related by GWAS ($P < 5E-07$) to digestive-related phenotypes (e.g. inflammatory bowel disease, gastroesophageal reflux or gut microbiota relative abundance), lifestyle factors (e.g. alcohol consumption, lifetime smoking, coffee consumption or moderate to vigorous physical activity levels) and brain and neuropsychiatric phenotypes (e.g. neuroticism, depression, anxiety, cognition or brain morphology) (Additional file 3: Table S6). In addition, we found that more that half of the credible variants (n = 953; 52%) were expression quantitative trait loci (eQTL) for at least one gene in one brain area (n = 895; 49%) and/or digestive tissue (n = 690; 38%; Additional file 3: Table S7).

Credible variants were mapped to 289 unique genes (Additional file 3: Table S8 and Additional file 2: Figure S4) that were significantly enriched in genes upregulated during early embryonic brain development (8th post conceptual week; Additional file 2: Figure S5) and in several gene-sets (Additional file 3: Table S9). Among the most significant ones, we found psychiatric disorders (GWAS catalog: autism spectrum disorder or schizophrenia, *P*-adjusted = 5.0E-193), digestive disorders (GWAS catalog: ulcerative colitis, *P*-adjusted = 1.1E-57 and inflammatory bowel disease, *P*-adjusted = 7.1E-40), autoimmune disease (KEGG: Systemic lupus erythematosus, *P*-adjusted = 7.9E-61) and histone deacetylases (Reactome: HDACS deacetylate histones, *P*-adjusted = 3.1E-46) (Additional file 3: Table S9).

Gene-based and gene-set analyses of MTAG-IBS risk loci

The gene-based analysis identified 76 significant genes, which were associated with expression changes in the cerebellum ($P = 5.2E-09$), frontal cortex ($P = 9.8E-07$), anterior cingulate cortex ($P = 1.8E-05$), basal ganglia nuclei (nucleus accumbens: $P = 6.9E-05$; caudate: $P = 9.7E-04$) and hypothalamus ($P = 4.3E-04$) (Additional file 3: Table S10, Additional file 2: Figure S6–S7) as well as with gene expression during the 21st post conceptual week ($P = 8.5E-04$) (Additional file 2: Figure S7). Among top findings, we found genes with a role in brain development and synaptic function, including *CADM2* and *NCAM1*, previously identified in the latest GWAS on IBS, and also genes involved in transcriptional regulation through mRNA transport or chromatin structure, including *FAM120A*, *PHF2* and different histone coding genes. When we conducted the gene-set analysis we found the *branching morphogenesis of a nerve* pathway significantly associated with IBS (gene-set size = 10 genes; $P = 1.7E-06$) (Additional file 3: Table S11).

Drug target identification

The enrichment analysis on druggable genes showed enrichment of MTAG-IBS-finemapped credible genes in druggable genes for 21 drugs (Additional file 3: Table S12), being l-lysine ($P < 2.2E-16$), belinostat ($P = 8.6E-10$), s-adenosylmethionine ($P = 7.0E-09$) and allopurinol ($P = 1.5E-07$), the top ones (Additional file 3: Table S12). They also included drugs related to musculoskeletal system, such as anti-inflammatory and antirheumatic drugs, or related to the nervous system, such as

Table 3 Results for the 42 independent lead SNPs identified in the MTAG-IBS analysis

Locus	Lead SNP	CHR	A1/A2	BP	Cross-trait analysis			FRQ	Nearest Gene	Functional category	Overlap with original GWAS IBS	Overlap with previous GWAs on psychiatric traits	Overlap with previous GWAs	CADD	RDB
					Beta	SE	P								
1	rs301806	1	T/C	8482078	− 0.009	0.002	1.67E-09	0.58	*RERE*	Intronic	*NO*	*Neuroticism*	*Known*	0.117	4
2	rs11206127	1	A/G	53713549	− 0.009	0.002	1.42E-08	0.43	*LRP8*	Intronic	*NO*	*No*	*Novel*	0.128	6
3	rs12755507	1	T/C	176164865	0.01	0.002	8.03E-10	0.625	*RFWD2*	Intronic	*NO*	*Depression*	*Known*	6.038	4
4	rs113198479	1	A/G	191347803	− 0.02	0.004	2.48E-08	0.953	*RP11-309H21.2*	Intergenic	*NO*	*No*	*Novel*	1.241	6
5	rs72740550	1	A/G	197342380	− 0.011	0.002	6.02E-09	0.219	*CRB1*	Intronic	*NO*	*Neuroticism & depression*	*Known*	5.063	7
6	rs115962846	2	A/G	58967058	− 0.015	0.003	3.68E-08	0.912	*LINC01122*	ncRNA_intronic	*NO*	*Neuroticism*	*Known*	2.103	7
7	rs28496790	2	A/C	161950047	0.01	0.002	3.70E-09	0.708	*AC009313.1*	Intergenic	*NO*	*No*	*Novel*	6.027	5
8	rs138218528	2	T/C	212676884	0.009	0.002	2.84E-08	0.667	*ERBB4*	Intronic	*NO*	*Neuroticism & depression*	*Known*	8.481	6
9	rs62246276	3	T/G	9445173	− 0.011	0.002	2.28E-08	0.179	*SETD5*	Intronic	*NO*	*No*	*Novel*	1.944	5
10	rs67416405	3	T/C	85539234	− 0.009	0.002	8.27E-09	0.353	*CADM2*	Intronic	*YES*	*No*	*Known*	3.769	6
11	rs1729951	3	T/G	136500733	− 0.009	0.002	9.01E-09	0.389	*RP11-102M11.2*	Intergenic	*NO*	*Neuroticism*	*Known*	0.078	NA
12	rs1442129	4	A/G	90849446	− 0.009	0.002	1.22E-08	0.453	*MMRN1*	Intronic	*NO*	*No*	*Novel*	5.378	NA
13	rs77087420	4	A/G	123122856	0.018	0.003	2.64E-08	0.945	*KIAA1109*	Intronic	*NO*	*No*	*Novel*	4.579	7
14	rs12513440	5	A/G	7259853	0.01	0.002	2.73E-08	0.243	*RP11-404K5.3*	Intergenic	*NO*	*No*	*Novel*	0.327	5
15	rs3099439	5	T/C	87545318	− 0.011	0.002	1.14E-12	0.539	*TMEM161B*	Intronic	*NO*	*Depression*	*Known*	1.562	NA
16	rs4481363	5	A/C	164474719	0.009	0.001	1.01E-09	0.524	*CTC-340A15.2*	ncRNA_intronic	*NO*	*Neuroticism & depression*	*Known*	6.522	6
16	rs180928232	5	A/G	166185949	− 0.012	0.002	4.46E-08	0.149	*CTB-7E3.1*	Intergenic	*NO*	*Neuroticism*	*Known*	2.692	6
17	rs200977	6	T/C	27854301	0.015	0.002	1.04E-11	0.873	*HIST1H3J*	Intergenic	*NO*	*Neuroticism & depression*	*Known*	1.251	NA
17	rs2534664	6	A/G	31469591	0.01	0.002	2.63E-10	0.456	*MICB*	Intronic	*NO*	*Depression*	*Known*	3.484	NA
17	rs1144708	6	T/C	31710020	− 0.01	0.002	7.49E-10	0.357	*MSH5:MSH5-SAPCD1*	Intronic	*YES*	*No*	*Known*	0.372	6
18	rs12374612	6	T/C	100955752	0.009	0.001	1.02E-08	0.478	*ASCC3*	Downstream	*NO*	*Neuroticism*	*Known*	0.29	6
19	rs2189246	7	A/G	82444372	0.01	0.001	1.98E-10	0.523	*PCLO*	Intronic	*NO*	*Depression*	*Known*	1.139	7
20	rs6956352	7	A/G	109131367	0.009	0.002	1.64E-08	0.458	*AC073071.1*	Intergenic	*NO*	*Depression*	*Known*	9.195	7
21	rs4726814	7	T/C	146691924	− 0.01	0.002	1.30E-08	0.275	*CNTNAP2*	Intronic	*NO*	*No*	*Novel*	1.37	7
22	rs4478545	8	A/G	94672542	− 0.01	0.002	4.77E-09	0.285	*LINC00535*	ncRNA_intronic	*NO*	*No*	*Novel*	1.326	6
23	rs3793577	9	A/G	23737627	− 0.01	0.002	3.46E-10	0.463	*ELAVL2*	Intronic	*NO*	*Neuroticism*	*Known*	19.76	5
24	rs4744242	9	T/G	96236711	− 0.011	0.002	8.68E-13	0.336	*FAM120A*	Intronic	*YES*	*Neuroticism*	*Known*	2.858	6
25	rs10123941	9	T/C	120518162	− 0.01	0.002	3.96E-09	0.727	*snoZ13_snr52*	Intergenic	*NO*	*Neuroticism*	*Known*	1.108	6
26	rs6584631	10	T/C	106656137	− 0.01	0.002	7.23E-09	0.244	*SORCS3*	Intronic	*NO*	*Depression*	*Known*	0.167	4
27	rs4937872	11	A/G	112827715	− 0.012	0.002	7.15E-15	0.589	*RP11-629G13.1*	Intergenic	*YES*	*Neuroticism*	*Known*	0.044	6

Table 3 (continued)

Locus	Lead SNP	CHR	A1/A2	BP	Cross-trait analysis			FRQ	Nearest Gene	Functional category	Overlap with original GWAS IBS	Overlap with previous GWAs on psychiatric traits	Overlap with previous GWAs	CADD	RDB
					Beta	SE	P								
28	rs9530139	13	T/C	31847324	− 0.011	0.002	2.11E-08	0.194	*B3GALTL*	Intronic	*NO*	*Depression*	*Known*	0.529	6
29	rs9597797	13	T/G	59183795	− 0.01	0.002	1.42E-09	0.251	*CTAGE16P*	Intergenic	*NO*	*Neuroticism*	*Known*	0.278	7
30	rs2121708	14	A/G	42146572	− 0.009	0.001	8.26E-10	0.517	*LRFN5*	Intronic	*NO*	*Depression*	*Known*	0.043	NA
31	rs35641442	14	A/G	75207263	0.009	0.002	6.65E-09	0.459	*FCF1*	Intergenic	*NO*	*Neuroticism & depression*	*Known*	11.4	7
32	rs1862743	16	A/C	60743834	− 0.009	0.001	1.08E-08	0.492	*GNPATP*	Intergenic	*NO*	*No*	*Novel*	1.06	6
33	rs11877758	18	T/G	35138110	− .,012	0.002	1.28E-13	0,692	*CELF4*	Intronic	*NO*	*Neuroticism & depression*	*Known*	2.718	7
33	rs2978362	18	T/C	32959397	− 0.008	0.001	2.65E-08	0.527	*ZNF396*	Intergenic	*NO*	*Depression*	*Known*	1.024	NA
34	rs12958048	18	A/G	53101598	0.01	0.002	4.76E-11	0,333	*TCF4*	Intronic	*NO*	*Neuroticism*	*Known*	2,08	5
34	rs17410557	18	T/C	50776391	− .,009	0.002	1.13E-08	0,606	*DCC*	Intronic	*NO*	*Neuroticism & depression*	*Known*	4.502	7
35	rs2111530	19	A/G	31891006	− 0.009	0.002	9.47E-09	0.602	*AC007796.1*	ncRNA_intronic	*NO*	*No*	*Novel*	17.04	7
36	rs2024568	20	T/C	44732089	0.011	0.002	1.52E-10	0.246	*RPL13P2*	Intergenic	*NO*	*Neuroticism & depression*	*Known*	0.149	6
37	rs11090039	22	A/G	41496800	0.012	0.002	2.87E-13	0.284	*EP300*	Intronic	*NO*	*Neuroticism*	*Known*	9.707	5

Overlap with previous GWAS was examined by identifying genome-wide significant SNPs within ± 5000 kb in the MTAG genome-wide significant for IBS and original GWAS genome-wide significant SNPs for each trait (i.e. neuroticism, depression and anxiety). If there were overlapping SNPs within this distance, they were considered independent signal if $r^2 > 0.2$. The independent signals identified (indicated as novel) were further confirmed using conditonal analysis

CHR chromosome; *A1* effect allele with respect to the Beta; *A2* alternate allele; *BP* base pair position Genome Reference Consortium Human Build 37 (GRCh37); *SE* standard error; *FRQ* frequency of the A1; *CADD* Combined Annotation Dependent Depletion score; *RDB* RegulomeDB score

anesthetics and drugs used in opioid dependence (Additional file 3: Table S12).

Partitioned heritability and genetic correlations

When we partitioned the h^2_{SNP} of IBS, we observed significant heritability enrichment in ten functional categories (Fig. 2 and Additional file 3: Table S13), with the strongest enrichment of variants in conserved regions (enrichment = 2.01; P = 4.0E-09), DNase I hypersensitive sites (DHSs) regions (enrichment = 1.66; P = 9.1E-08) and histone H3 lysine 9 acetylation (H3K9ac) peaks (enrichment = 6.88; P = 1.1E-07).

We found significant genetic correlations between IBS and 13 gastrointestinal, immunological or psychiatric disorders using GWAS summary statistics available in the MR-Base database [35], including gastric reflux (rg = 0.51; P = 2.6E-36), the cross-disorder GWAS from the PGC involving schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorders and attention-deficit/hyperactivity disorder (ADHD) (rg = 0.44, P = 9.7E-46), diverticulitis (rg = 0.44, P = 7.4E-22), hiatus hernia (rg = 0.43; P = 4.7E-20) and chronic fatigue syndrome (rg = 0.39, P = 2.0E-04), among others (Fig. 2 and Additional file 3: Table S14).

Causal analysis using summary effect estimates (CAUSE)

CAUSE [36] showed consistent evidence for a causal effect of the genetic liability of IBS on neuroticism (ΔELPD = -3.6, SE = 1.9, P = 0.031), depression (ΔELPD = -5.9, SE = 1.8, P = 5.4E-03) and anxiety (ΔELPD = -2.9, SE = 1.7, P = 0.049). We also found evidence for reverse causality with a causal effect of the genetic liability of neuroticism and depression on IBS (ΔELPD = -7.3, SE = 1.4, P = 1.5E-07 and ΔELPD = -6.3, SE = 1.4, P = 1.8E-06 respectively) but there was no evidence for a causal relationship when anxiety was considered as exposure and IBS as outcome (Fig. 2, Additional file 3: Table S15a, b and Additional file 2: Figure S8).

Discussion

In the present study we found extensive genetic sharing between IBS, neuroticism, depression and anxiety, and identified 42 genome-wide significant SNPs for IBS, of which 38 are novel. Our findings confirm the polygenic architecture of the disorder, with more than 12,000 variants explaining 90% of the $h^2_{SNP,}$ and represent a great advance over the previously reported six genome-wide associated SNPs [18]. Significant signal enrichment was found in genes showing heightened expression in the brain during early embryonic development and playing prominent roles in mental and digestive disorders, autoimmune diseases and transcription regulation.

Our results confirm a role on IBS of genes involved in brain development and synaptic function as well as genes previously associated with psychiatric conditions [18]. We detected 27 SNPs for IBS also associated with at least one of the three mental conditions under study, and found evidence supporting that IBS and neuroticism, which is genetically correlated with many psychiatric disorders [39], share a considerable proportion of their genetic background. The widespread common genetic risk sharing with mental conditions was further supported by the positive genetic correlation found between IBS and many psychiatric disorders (i.e. schizophrenia, ADHD, autism or depression) and by the IBS associated variants being located within genes significantly expressed in the brain. These results are in agreement with the higher burden of mental disorders often co-existing in IBS patients, add further evidence of substantial pleiotropy of contributing loci and underscore that genetic influences on IBS may transcend diagnostic boundaries.

Among top findings we identified genes associated with IBS in previous GWAS, such as *CADM2* and *NCAM1,* members of the synaptic cell adhesion molecules that play a role in synapse organization and plasticity [40, 41]. Interestingly, NCAM peptide mimetics have been proven to have both antidepressant and anti-inflammatory effects [42, 43], pointing them as a potential therapeutic target for IBS. Novel loci for IBS include interesting genes previously associated with

(See figure on next page.)

Fig. 2 Follow-up analysis of MTAG-IBS results and causal analysis. **a** Functional annotation of the credible variants associated with MTAG-IBS. **b** RegulomeDB scores of the credible variants associated with MTAG-IBS. Low scores indicate increasing likelihood of having regulatory function. **c** Distribution of the credible variants associated with MTAG-IBS across 15 categories of minimum chromatin state. Lower state indicating higher accessibility and states from 1 to 7 refer to open chromatin states. **d** Genetic correlations (rg) between MTAG-IBS results and 17 phenotypes involving digestive, immunological and psychiatric disorders. Only significant correlations after Bonferroni correction are displayed. **e** Bar graphs depicting the size of the genomic locus (left), number of candidate SNPs in the locus (center) and number of mapped genes in the genomic locus (right). Genomic loci are displayed by "chromosome: start position-end position". **f** Partitioning of the SNP heritability of the MTAG-IBS results using LD Score regression. Enrichment was calculated by dividing the partial heritability of a category by the proportion of SNPs in that category (proportion indicated by color). Only significant enrichments are displayed. **g** Causal relationships between IBS and neuroticism, depression and anxiety assessed using Causal Analysis Using Summary Effect estimates (CAUSE). Only associations with evidence of causal relationship are displayed

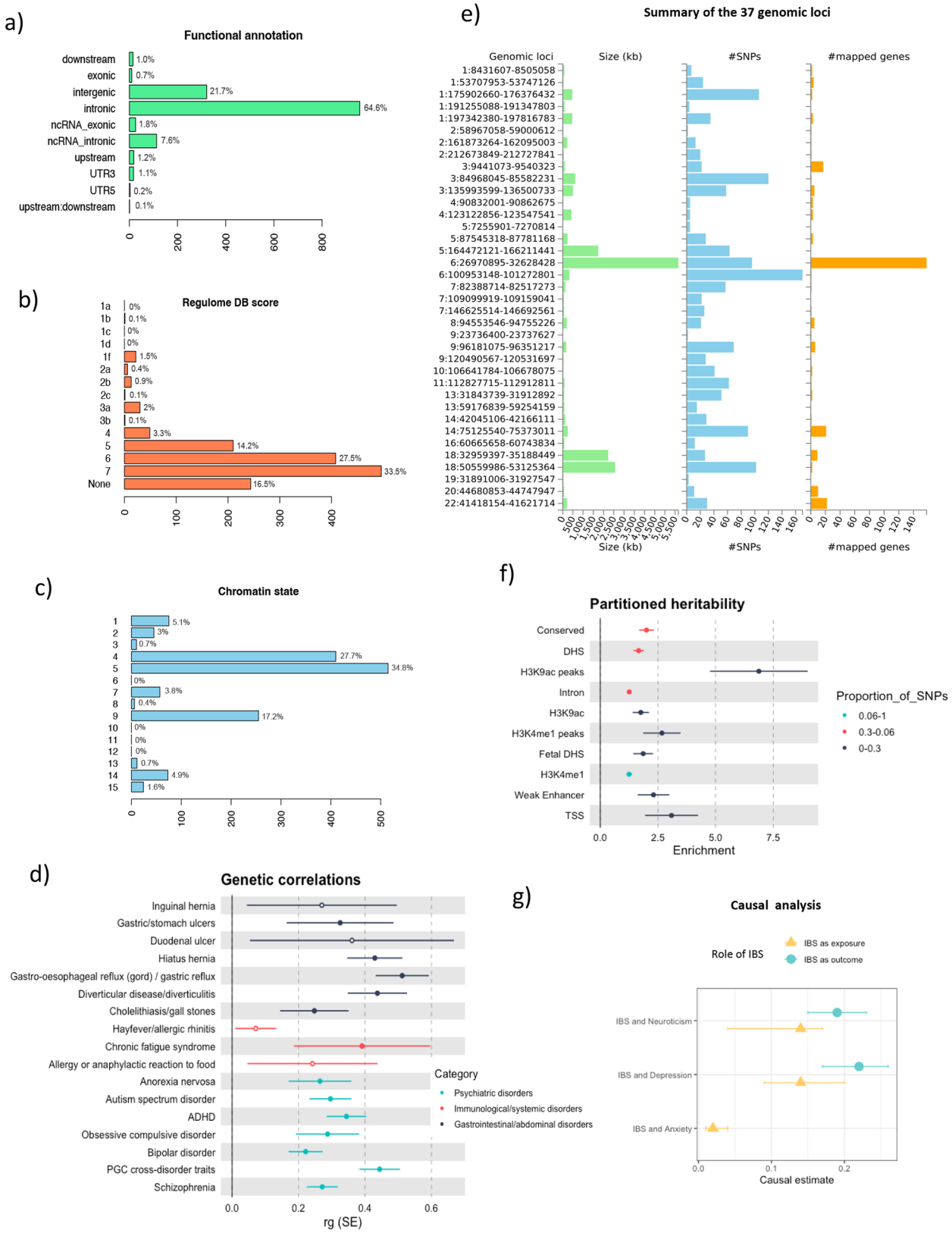


Fig. 2 (See legend on previous page.)

depression and other mental disorders, such as *RERE*, that regulates retinoic acid signaling during development [44–46], *PCLO,* involved in synaptic vesicle trafficking, *TMEM161B* [47], a brain-expressed transmembrane protein [48], *RBFOX1,* a splicing regulator mainly expressed in neurons, that is one of the most pleiotropic genes among psychiatric disorders [49] or *DRD2,* encoding the dopamine receptor D2R and one of the strongest candidates for psychiatric disorders and traits [50]. Interestingly, several studies in animal models suggested an important role for dopamine signaling both in the development and progression of inflammatory bowel disease [51] and treatment with D2R agonists decreased the severity of ulcerative colitis in mice and rats [52].

Interestingly, three of the identified genome-wide significant SNPs had been tested for association with psychiatric and neurological phenotypes, which contributes to clarify their potential functional role. One of these SNPs is the rs301806 (MTAG-IBS P-value = 1.7E-09) located in chromosome 1 in the *RERE* gene and previously associated with neuroticism. A neuroimaging study of drug-naïve individuals with MDD found that reductions in cortical thickness among patients (n = 47) compared to controls (n = 42) were significantly larger among those with the T/T genotype of this SNP compared to C carriers [53]. Another SNP, the rs4481363 (MTAG-IBS P-value = 1.0E-09) located in chromosome 5 in the *CTC-340A15.2* gene, previously associated with neuroticism and depression, has been examined in a study testing associations between genetic variants associated with subjective well-being and depressive symptoms and these, and metabolic phenotypes in a Chinese elderly sample (n = 1788). However, this SNP did not show association with any of the phenotypes studied [54]. The third SNP is the rs2024568 (MTAG-IBS P-value = 1.5E-10) in chromosome 20 (nearest gene was the *RPL13P2*) previously associated with neuroticism and depression. This variant was identified as likely affecting DNA methylation patterns in multiple sclerosis (MS) in a gene-regulatory network integrating GWAS summary statistics and DNA methylation profiles from 140 cases of MS and 139 controls [55].

We also provide new insights underlying IBS, showing strong evidence of transcriptional regulation mechanisms playing a role in the disorder, including non-coding RNAs and histone modification. The over-representation of credible variants in non-coding regions is a common finding when investigating the genetic basis of complex traits [56]. Although the role of non-coding variants is still unclear, it has been suggested that non-coding variants may impact the phenotype by alteration of regulatory elements such as enhancers, transcription factor binding sites or chromatin state [56]. Indeed, we found 75% of the variants within credible sets were located in open chromatin regions (minimum chromatin state ≤ 7), 3% were likely to affect the binding of transcription factors (RegulomeDB scores from 1b to 2c) and 0.05% may be deleterious (CADD score > 12.37). These results point towards a potential role for IBS associated non-coding variants in gene regulation. More specifically, we found genes encoding histones and histone modifying enzymes among top findings, and enrichment of IBS associations in histone acetylation and methylation peaks and in target genes for the histone deacetylase inhibitor belinostat [57]. These findings are in agreement with previous results involving chromatin modifications in maintenance of anxiety behavior and nociception and in visceral hypersensitivity induced by early-life stress [58, 59]. Additionally, top findings also include non-coding RNAs, an epigenetic mechanism that has been involved in regulation of genes related with visceral pain response and intestinal permeability [60–62]. These results add additional evidence towards the role of epigenetic programming in inflammation, visceral pain as well as in intestinal permeability, sensibility and motility in both humans and animal models of IBS [58, 59, 63, 64].

Despite many of the findings pointing out neurobiological processes and mental disorders, we also detected links between IBS and gastrointestinal-related phenotypes. Fine mapping showed that 38% of the credible variants were eQTLs for at least one digestive tissue and that credible sets were located in genes enriched in different digestive disorders, including ulcerative colitis and inflammatory bowel disease. In addition, positive genetic correlations were found between IBS and gastric reflux, diverticulitis, hiatus hernia, cholelithiasis/gallstones and gastric/stomach ulcers, among others, which adds evidence on the overlap between the genetic risk for IBS and for other digestive-related disorders and traits. These findings may reflect the multi-factorial etiology proposed for IBS involving psychological factors, abnormal brain functioning and dysregulation of brain-gut interactions [15, 65–67], as previously proposed in different psychiatric disorders such as depression [68].

IBS-associated signals were also enriched in target genes of relevant drugs, including l-lysine or S-adenosylmethionine. L-lysine acts as partial serotonin 5-HT4 receptor antagonist and inhibits serotonin-mediated intestinal pathologies in rats, including anxiety and stress-induced fecal excretion and severity of diarrhea [69]. Interestingly, l-lysine, and other 5-HT4 receptor antagonists, are promising targets for the treatment of diarrhea-predominant IBS [70, 71] and may aminorate serotonin disturbances in gut and brain that account for part of intestinal and mental disorders [69]. Additional

drugs of interest include S-adenosylmethionine, involved in neurotransmission signaling that has a putative antidepressant effect [72, 73] or allopurinol that improves inflammatory bowel disease clinical outcomes [74], among others.

Despite the high prevalence of psychiatric comorbidities reported in patients with IBS, particularly anxiety and depression, a clear temporal relationship between them has not been well established. We found evidence for a bidirectional causal effect between IBS and neuroticism or depression when accounting for correlated pleiotropy, which strengthens previous evidence [18]. In addition, we found evidence for a causal effect of the genetic liability of IBS on anxiety. These findings support that IBS increases the risk of subsequent depressive and anxiety disorders described in longitudinal study designs [75] and also previous evidence supporting that prior depression raises the risk of developing IBS [76, 77]. We found, however, no evidence for a causal effect of the genetic liability of anxiety on IBS when accounting for correlated pleiotropy, in line with previous results [18]. Although the sample size for anxiety was more limited and these results may also reflect lack of statistical power. Long term follow-up studies as well as larger datasets and sensitivity analyses are required to confirm the robustness of these results and to better understand the temporal relationship between IBS and comorbid mental conditions.

A major strength of our study is the substantial larger sample size compared with previous studies. By conducting meta-analysis of GWAS summary statistics for IBS and comorbid mental conditions with MTAG we increased the effective sample size from 486,601 in the original IBS dataset to 887,490 individuals and the number of IBS genome-wide significant associated SNPs from six in the single-trait analysis to 42. Thirty-eight of them were novel for IBS and 11 were not associated with any of the mental conditions under study, which highlight that MTAG combining GWAS on IBS and mental conditions is a robust strategy to identify trait specific genetic associations. In addition, four of the previously six identified SNPs were also significant in the present study [18]. Even though two identified SNPs demonstrated less association here, their associations were still suggestive ($P < 5E\text{-}07$) and in concordance in the direction of the effect with the original GWAS study on IBS, which supports validity of the findings across studies.

The study, however, should be considered in the context of some limitations: (i) We did not account for phenotypic overlap and cannot discard that comorbid conditions may have biased the observed results. Also, IBS is considered a highly heterogenous disorder with pathophysiological differences observed among clinical subtypes, between genders, and across age groups and geographic locations [1]. Accounting for such factors may contribute to better characterize the disorder, capture its genetic background and identify overlap with other comorbid disorders that may impact on IBS risk, prognosis and clinical outcome [6]; (ii) Despite the strong genetic correlation between IBS and the three mental conditions under study, MiXeR was unable to assess the genetic overlap between IBS, depression and anxiety probably due to the high polygenicity and low SNP heritability estimates for these traits (0.083 and 0.099, respectively) and the limited sample size of the original GWAS on anxiety. We cannot discard, either, that due to lack of power we did not detect IBS signals previously reported for anxiety in the original GWAS or evidence for anxiety increasing the risk for IBS in the causality analyses; (iii) gene-based analyses may be inflated as suggested by the lambda over 1, although given the increased power of gene-based over single SNP analyses and the lack of residual stratification or confounding inflation in the MTAG–IBS results, this inflation may just reflect high polygenicity; (iv) Combining GWAS that differ a great deal in power may lead to inflation of FDR, according to MTAG authors [24]. In this study we combined GWAS with different sample sizes, however their mean chi-squared was similar and accordingly the max-FDR estimated in our IBS analysis was 0.02, which suggested no inflation of our results. Moreover, despite increasing considerably the effective sample size for IBS through the addition of multiple mental conditions, a number of outcomes were gastrointestinal-related phenotypes, which further supports this approach.

In summary, we identified novel risk loci for IBS, reveal new insights of its polygenic architecture and extended previous knowledge on the genetic overlap and causal relationships between IBS, neuroticism, depression and anxiety. Overall, we advance our understanding of the biological mechanisms underlying IBS, highlighted candidate genes related to brain development and function as well as transcriptional regulation and provide insight into the association between IBS and comorbid mental disorders.

Abbreviations

ADHD	Attention-deficit/hyperactivity disorder
AIC	Akaike information criterion
ATC	Anatomical therapeutic chemical
CAUSE	Causal analysis using summary effect estimates
IBS	Irritable bowel syndrome
FDR	False discovery rate
ELPD	Expected log pointwise posterior density
FUMA	Functional Mapping and Annotation of GWAS
GCTA	Genome-wide complex trait analysis

GWAS	Genome-wide association study
h^2_{SNP}	SNP heritability
LDSC	Linkage disequilibrium score regression
MAGMA	Generalized gene-set analysis of GWAS data
MTAG	Multi-trait analysis of GWAS

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-023-04107-5.

Additional file 1. Supplementary materials.

Additional file 2: Figure S1. MiXeR results for IBS and neuroticism. A) Venn diagram depicting the estimated number of trait-influencing variants shared (gray) between IBS and neuroticism. Unique variants for each trait are depicted in blue for IBS and orange for neuroticism. The number of trait-influencing variants in thousands is shown, with the standard error in thousands provided in parentheses. The size of the circles reflects the polygenicity of each phenotype, with larger circles corresponding to greater polygenicity. The estimated genetic correlation (r_g) is shown in the bar. Red color indicates positive genetic correlation. B) and C) depict conditional Q–Q plots of observed versus expected −log10 p-values in the primary trait as a function of significance of association with a secondary trait at the level of $p \leq 0.1$ (orange lines), $p \leq 0.01$ (green lines), $p \leq 0.001$ (red lines). Blue line indicates all SNPs. Dotted lines in blue, orange, green, and red indicate model predictions for each stratum. Black dotted line is the expected Q–Q plot under null (no SNPs associated with the phenotype). D) Log-likelihood curves highlighting the goodness of model fit. The minimum point indicates the best-fitting model estimate of the number of influencing variants shared between two traits (Supplementary Table 1). **Figure S2.** LD Score regression plot with the MTAG-IBS results. Each point represents an LD score quantile. The x-axis represents the mean LD score for the variants included in the quantile and the y-axis represents the mean χ2 of variants in that quantile. The black line is the LD score regression line fitted by a linear regression model with mean χ2 as the outcome variable and mean LD score for each bin as the independent variable (Coefficient=0.011, p=2E-16). **Figure S3.** Regional Plots of the 42 lead SNPs identified in the MTAG-IBS analysis. In red, genes mapped by SNPs in the credible sets based on physical proximity, chromatin interaction and/or eQTLs using FUMA. **Figure S4.** Gene-based test QQ plot. Observed versus expected gene-based test p-values on the-log10 scale are shown. Lambda: 1.6855. **Figure S5.** Enrichment of genes mapped to MTAG-IBS variants with credible sets on Differentially Expressed Genes (DEG) in brain tissue. Results from hypergeometric test evaluating enrichment of the 289 mapped genes by credible variants in DEG in brain tissue representing different brain developmental stages in BrainSpan. Significant enrichment at Bonferroni corrected P-value ≤ 0.05 are coloured in red. **Figure S6.** MAGMA tissue expression analysis using GTEx v.8. Results from MAGMA gene-property analysis between gene-based MTAG-IBS associations and tissue specific gene expression profiles. (A) GTEx v.8 54 tissues. (B) GTEx v.8 30 general tissues. Red bars indicate significant results. **Figure S7.** MAGMA tissue expression analysis using Brainspan. Results from MAGMA gene-property analysis between gene-based MTAG-IBS results and tissue specific gene expression profiles in Brainspan. (A) BrainSpan 29 ages. (B) Brainspan 11 developmental stages. Red bars indicate significant results. **Figure S8.** Scatter plots of the causal analysis. Scatter plots of exposure versus outcome effect sizes for: the sharing model (left) illustrating the pattern induced by a shared factor (correlated pleiotropy, eta) without a causal effect; the causal model (middle) illustrating the pattern induced when including also a causal effect (gamma); and the expected log pointwise posterior density (DEPLD) contribution from each variant for each causal relationship tested.

Additional file 3: Table S1. a Univariate and bivariate MiXeR output for IBS vs. neuroticism. **Table S2.** Results from association analyses of lead and secondary lead variants conditioned on the lead variant using COJO. For locus with more than two secondary variants, we further check independency of the secondary variants among each other. P-value of the secondary variants in MTAG-IBS and P-value after conditional analysis using COJO are given. The last column in the right indicates whether the secondary variant was considered as an independent signal. **Table S3.** Results in the original GWAS on IBS of the lead SNPs from MTAG-IBS. **Table S4.** Credible variants for each of the 37 independent loci or IBS identified in the cross-trait analysis using MTAG. Variants included herein were identified by all three fine-mapping methods (PAINTOR, CAVIARBF and FINEMAP). The prosterior probability of the variant being causal estimated by PAINTOR, CAVIARBF and FINEMAP is indicated in the last three columns. Chromosome (CHR), base position (BP), and SNP of the index variants. Effect alle (A1) and non-effect allele (A2) with respect to the beta (BETA). The standard error of the beta (SE) and the association P-value. **Table S5.** Functional annotation of the credible variants for each of the 42 lead variants identified in the cross-trait analysis of IBS using MTAG. Chromosome (CHR), SNP of the index variants, base position (BP), combined Annotation Dependent Depletion score (CADD), RegulomeDB score (RDB) and chromatin state (minChrState) are indicated. **Table S6.** GWAS Catalog results for MTAG-IBS credible set variants associated with other traits in previous GWAS. **Table S7.** Variants in credible sets mapped to genes associated with eQTLs. **Table S8.** Genes mapped to sets of credible variants in FUMA. **Table S9.** Enrichment of genes mapped to variants in credible sets. **Table S10.** Results of the gene-based association analysis on MTAG-IBS using MAGMA. Only gene-wide signficant genes are shown. The p-value for gene-wide significance after Bonferroni correction was 0.05/18,135=2.757x10-6. **Table S11.** Results of the gene-set analysis on MTAG-IBS using MAGMA. **Table S12.** Enrichment analysis on druggable genes for MTAG-IBS genes using data from PharmaKG.Categories according to the Anatomical Therapeutic Chemical (ATC) classification system are provided. **Table S13.** Results of the partitioned heritability analysis using LDSC. **Table S14.** Genetic correlations between MTAG-IBS and 28 phenotypes inlcuding digestive, immnological and psychiatric disorders from MR-Base. **Table S15**. a CAUSE model comparison.

Author contributions

SA, MSA, JC, LVR, NLL and MR designed and supervised the study. SA analyzed the data. SA, MSA, JC and MR wrote the manuscript draft. All authors contributed to the interpretation of results and critically reviewed the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-00444 and 2017SGR-1461); the Ministry of Science, Innovation and Universities (RYC2021-031324-I to J.C.D) and the European Union H2020 Programme (H2020/2014–2020) under grant agreements no. 848228 (DISCOvERIE). SA acknowledge a Miguel Servet contract (CP22/00026) awarded by the Instituto de Salud Carlos III and co-funded by the European Union Found: Fondo Social Europeo Plus, FSE+. M.S.A acknowledge a Miguel Servet contract (CP22/00128) awarded by the Instituto de Salud Carlos III and co-funded by the European Union Found: Fondo Social Europeo Plus, FSE+. This document is also an output of a project grant (Grant Agreement nº: 848228, DISCOvERIE) funded under H2020 Research Programme of the European Commission. The content of this document represents the views of the author(s) only and is his/her/their sole responsibility; it cannot be considered to reflect the views of the European Commission or any other body of the European Union. The European Commission do not accept any responsibility for use that may be made of the information it contains.

Availability of data and materials

All data used in the current study is publicly available. Summary statistics for IBS can be download from European Bioinformatics Institute GWAS Catalog (https://www.ebi.ac.uk/gwas/). Summary statistics for neuroticism can be downloaded from https://ctg.cncr.nl/software/summary_statistics/ and http://www.ccace.ed.ac.uk. Summary statistics for depression can be downloaded from https://datashare.ed.ac.uk/handle/10283/3203. Summary statistics for anxiety can be downloaded from http://www.nealelab.is/uk-biobank. Genotype tissue expression (GTEx v8) portal: http://www.gtexportal.org/home/datasets. BRAINEAC: http://www.braineac.org. eQTL catalogue: https://www.ebi.ac.uk/eqtl/Methods/. PsychENCODE: http://resource.psychencode.org. CommonMind Consortium (CMC/CMC): https://www.synapse.org//#!Synapse:syn5585484. WEB-based GEne SeT AnaLysis Toolkit (WebGestAlt): http://www.

webgestalt.org. SNP heritability and genetic correlations: https://github.com/bulik/ldsc. MiXeR: https://github.com/precimed/mixer. Conditional analysis: https://yanglab.westlake.edu.cn/software/gcta/#COJO. Multi-Trait Analysis of GWAS (MTAG): https://github.com/omeed-maghzian/mtag). Fine-mapping: https://github.com/mulinlab/CAUSALdb-finemapping-pip. Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA): https://fuma.ctglab.nl/. Partitioned heritability: https://github.com/bulik/ldsc/wiki/Partitioned-Heritability. MR-Base database: https://github.com/MRCIEU/mrbase_casestudies. Causal Analysis Using Summary Effect estimates (CAUSE): https://jean997.github.io/cause/pipeline.html. The use of each software tools has been described in the Methods section. Analysis code and scripts used in the current study are available upon request from the corresponding authors.

Declarations

Ethics approval and consent to participate

This article contains results derived from data from human participants collected by several studies performed by previous studies. All participants gave informed consent in all the corresponding original studies. Our study is based on the large-scale GWAS datasets, and not the individual-level data. Hence, ethical approval was not applicable.

Consent for publication

Not applicable.

Competing interests

JARQ was on the speakers bureau and/or acted as consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Technofarma, Rubió and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Mental Health chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen-Cilag, Shire, Oryzon, Roche, Psious, and Rubió. ARU was on the speakers bureau and/or acted as consultant for Janssen-Cilag and Organon in tha last two years. All other authors declare no biomedical financial interests or conflicts of interest. JS has served as consultant for Noventure SL, Devintecpharma, Aboca, Reckitt, Ipsen and Pileje and discloses present and past recent scientific collaborations with Salvat, Norgine, Alfa-Sigma, Cosmo, Adare, Ordesa and Danone that do not constitute a conflict of interest in developing the content of the present manuscript. MS was on the speakers bureau and/or acted as consultant/advisory Board member for Tillotts, Menarini, Kyowa Kirin, Takeda, Biocodex, AlfaSigma, Sanofi, Janssen Immunology, Pfizer, Ferrer, BioGaia, Falk Foundation, Danone Nutricia Research, Ironwood, Genetic Analysis AS, DSM, Arena, Adnovate and Pharmanovia. He also was funded by Glycom/DSM research Grants.

Author details

[1]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Passeig Vall d'Hebron, 119-129, 08035 Barcelona, Spain. [2]Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [3]Biomedical Network Research Centre On Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. [5]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [6]Laboratory of Neuro-Immuno-Gastroenterology, Digestive System Research Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Barcelona, Spain. [7]Department of Gastroenterology, Vall d'Hebron Barcelona Hospital Campus, Vall d'Hebron Hospital Universitari, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain. [8]Department of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [9]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERHED), Instituto de Salud Carlos III, Madrid, Spain. [10]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [11]Centre for Functional GI and Motility Disorders, University of North Carolina, Chapel Hill, NC, USA.

Received: 30 January 2023 Accepted: 5 April 2023
Published online: 21 April 2023

References

1. Canavan C, West J, Card T. The epidemiology of irritable bowel syndrome. Clin Epidemiol. 2014;6:71–80.
2. Sperber AD, Bangdiwala SI, Drossman DA, Ghoshal UC, Simren M, Tack J, et al. Worldwide prevalence and burden of functional gastrointestinal disorders, results of rome foundation global study. Gastroenterology. 2021;160:99-114.e3.
3. Yeh H-W, Chien W-C, Chung C-H, Hu J-M, Tzeng N-S. Risk of psychiatric disorders in irritable bowel syndrome-A nationwide, population-based, cohort study. Int J Clin Pract. 2018;72:e13212.
4. Ford AC, Moayyedi P, Chey WD, Harris LA, Lacy BE, Saito YA, et al. American college of gastroenterology monograph on management of irritable bowel syndrome. Am J Gastroenterol. 2018;113:1–18.
5. Zamani M, Alizadeh-Tabari S, Zamani V. Systematic review with meta-analysis: the prevalence of anxiety and depression in patients with irritable bowel syndrome. Aliment Pharmacol Ther. 2019;50:132–43.
6. Lee S, Wu J, Ma YL, Tsang A, Guo WJ, Sung J. Irritable bowel syndrome is strongly associated with generalized anxiety disorder: a community study. Aliment Pharmacol Ther. 2009;30:643–51.
7. Lee C, Doo E, Choi JM, Ho JS, Ryu HS, Lee JY, et al. The Increased Level of Depression and Anxiety in Irritable Bowel Syndrome Patients Compared with Healthy Controls: Systematic Review and Meta-analysis. J Neurogastroenterol Motil. 2017;23:349.
8. Fond G, Loundou A, Hamdani N, Boukouaci W, Dargel A, Oliveira J, et al. Anxiety and depression comorbidities in irritable bowel syndrome (IBS): a systematic review and meta-analysis. Eur Arch Psychiatry Clin Neurosci. 2014;264:651–60.
9. Zhang QE, Wang F, Geng Q, Zheng W, Ng CH, Ungvari GS, et al. Depressive symptoms in patients with irritable bowel syndrome: a meta-analysis of comparative studies. Int J Biol Sci. 2018;14:1504–12.
10. Hazlett-Stevens H, Craske MG, Mayer EA, Chang L, Naliboff BD. Prevalence of irritable bowel syndrome among university students: the roles of worry, neuroticism, anxiety sensitivity and visceral anxiety. J Psychosom Res. 2003;55:501–5.
11. Lydiard RB. Irritable bowel syndrome, anxiety, and depression: what are the links? J Clin Psychiatry. 2001;62(Suppl 8):38–45.
12. Nisticò V, Rossi RE, D'Arrigo AM, Priori A, Gambini O, Demartini B. Functional neuroimaging in irritable bowel syndrome: a systematic review highlights common brain alterations with functional movement disorders. J Neurogastroenterol Motil. 2022;28:185–203.
13. Saffouri GB, Shields-Cutler RR, Chen J, Yang Y, Lekatz HR, Hale VL, et al. Small intestinal microbial dysbiosis underlies symptoms associated with functional gastrointestinal disorders. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-09964-7.
14. Aguilera-Lizarraga J, Hussein H, Boeckxstaens GE. Immune activation in irritable bowel syndrome: what is the evidence? Nat Rev Immunol. 2022. https://doi.org/10.1038/s41577-022-00700-9.
15. Ford AC, Sperber AD, Corsetti M, Camilleri M. Irritable bowel syndrome. The Lancet. 2020;396:1675–88.
16. Smoller JW. The genetics of stress-related disorders: PTSD, depression, and anxiety disorders. Neuropsychopharmacology. 2016;41:297–319.
17. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157:1552–62.
18. Eijsbouts C, Zheng T, Kennedy NA, Bonfiglio F, Anderson CA, Moutsianas L, et al. Genome-wide analysis of 53,400 people with irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders. Nat Genet. 2021;53:1543–52.
19. Wu Y, Murray GK, Byrne EM, Sidorenko J, Visscher PM, Wray NR. GWAS of peptic ulcer disease implicates Helicobacter pylori infection, other gastrointestinal disorders and depression. Nat Commun. 2021;12:1146.
20. Nagel M, Jansen PR, Stringer S, Watanabe K, De Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet. 2018;50:920–7.
21. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22:343–52.
22. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.

23. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat Commun. 2019;10:1–11.
24. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet. 2018;50:229–37.
25. Maihofer AX, Choi KW, Coleman JRI, Daskalakis NP, Denckla CA, Ketema E, et al. Enhancing discovery of genetic variants for posttraumatic stress disorder through integration of quantitative phenotypes and trauma exposure information. Biol Psychiatry. 2022;91:626–36.
26. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44:369-75,S1-3.
27. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32:1493–501.
28. Greenbaum J, Deng HW. A statistical approach to fine mapping for the identification of potential causal variants related to bone mineral density. J Bone Miner Res. 2017;32:1651–8.
29. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, et al. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. Genetics. 2015;200:719–36.
30. Wang J, Huang D, Zhou Y, Yao H, Liu H, Zhai S, et al. CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucleic Acids Res. 2020;48:D807–16.
31. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
33. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47:W199–205.
34. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015. https://doi.org/10.1038/ng.3404.
35. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
36. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.
37. Nagel M, Jansen PR, Stringer S, Watanabe K, De Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet. 2018;50:920–7.
38. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018;50:6–11.
39. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018. https://doi.org/10.1126/SCIENCE.AAP8757.
40. Boutwell B, Hinds D, Agee M, Alipanahi B, Auton A, Bell RK, et al. Replication and characterization of CADM2 and MSRA genes on human behavior. Heliyon. 2017. https://doi.org/10.1016/J.HELIYON.2017.E00349.
41. Lüthi A, Laurent JP, Figurov A, Mullert D, Schachnert M. Hippocampal long-term potentiation and neural cell adhesion molecules L1 and NCAM. Nature. 1994;372:777–9.
42. Downer EJ, Cowley TR, Lyons A, Mills KHG, Berezin V, Bock E, et al. A novel anti-inflammatory role of NCAM-derived mimetic peptide. FGL Neurobiol Aging. 2010;31:118–28.
43. Turner CA, Lyons DM, Buckmaster CL, Aurbach EL, Watson SJ, Schatzberg AF, et al. Neural cell adhesion molecule peptide mimetics modulate emotionality: pharmacokinetic and behavioral studies in rats and non-human primates. Neuropsychopharmacology. 2018;44:356–63.
44. Fregeau B, Kim BJ, Hernández-García A, Jordan VK, Cho MT, Schnur RE, et al. De Novo mutations of RERE cause a genetic syndrome with features that overlap those associated with proximal 1p36 deletions. Am J Hum Genet. 2016;98:963.
45. Jordan VK, Fregeau B, Ge X, Giordano J, Wapner RJ, Balci TB, et al. Genotype–phenotype correlations in individuals with pathogenic RERE variants. Hum Mutat. 2018;39:666.
46. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50:668–81.
47. Mbarek H, Milaneschi Y, Hottenga JJ, Ligthart L, De Geus EJC, Ehli EA, et al. Genome-wide significance for PCLO as a gene for Major depressive disorder. Twin Res Hum Genet. 2017;20:267–70.
48. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48:1031–6.
49. Lee PH, Anttila V, Won H, Feng YCA, Rosenthal J, Zhu Z, et al. Genomic relationships, novel Loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell. 2019;179:1469-1482.e11.
50. Ayano G. Dopamine: receptors, functions, synthesis pathways, locations and mental disorders review of literatures. J Ment Disord Treat. 2016. https://doi.org/10.4172/2471-271x.1000120.
51. Kurnik-Łucka M, Pasieka P, Łączak P, Wojnarski M, Jurczyk M, Gil K. Gastrointestinal dopamine in inflammatory bowel diseases: a systematic review. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms222312932.
52. Tolstanova G, Deng X, Ahluwalia A, Paunovic B, Prysiazhniuk A, Ostapchenko L, et al. Role of dopamine and D2 dopamine receptor in the pathogenesis of inflammatory bowel disease. Dig Dis Sci. 2015;60:2963–75.
53. Katsuki A, Kakeda S, Watanabe K, Igata R, Otsuka Y, Kishi T, et al. A single-nucleotide polymorphism influences brain morphology in drug-naïve patients with major depressive disorder. Neuropsychiatr Dis Treat. 2019;15:2425–32.
54. Wang Y, Ma T, Zhu YS, Chu XF, Yao S, Wang HF, et al. The KSR2-rs7973260 polymorphism is associated with metabolic phenotypes, but not psychological phenotypes, in Chinese elders. Genet Test Mol Biomarkers. 2017;21:416–21.
55. Manuel AM, Dai Y, Jia P, Freeman LA, Zhao Z. A gene regulatory network approach harmonizes genetic and epigenetic signals and reveals repurposable drug candidates for multiple sclerosis. Hum Mol Genet. 2023. https://doi.org/10.1093/HMG/DDAC265.
56. Schipper M, Posthuma D. Demystifying non-coding GWAS variants: an overview of computational tools and methods. Hum Mol Genet. 2022;31:R73–83.
57. Beck HC, Nielsen EC, Matthiesen R, Jensen LH, Sehested M, Finn P, et al. Quantitative proteomic analysis of post-translational modifications of human histones. Mol Cell Proteomics. 2006;5:1314–25.
58. Aguirre JE, Winston JH, Sarna SK. Neonatal immune challenge followed by adult immune challenge induces epigenetic-susceptibility to aggravated visceral hypersensitivity. Neurogastroenterol Motil. 2017;29:e13081.
59. Moloney RD, Stilling RM, Dinan TG, Cryan JF. Early-life stress-induced visceral hypersensitivity and anxiety behavior is reversed by histone deacetylase inhibition. Neurogastroenterol Motil. 2015;27:1831–6.
60. Vicario M, Martínez C, Santos J. Role of microRNA in IBS with increased gut permeability. Gut. 2010;59:710–2.
61. Martínez C, Rodiñō-Janeiro BK, Lobo B, Stanifer ML, Klaus B, Granzow M, et al. Original article: miR-16 and miR-125b are involved in barrier function dysregulation through the modulation of claudin-2 and cingulin expression in the jejunum in IBS with diarrhoea. Gut. 2017;66:1537.
62. Wohlfarth C, Schmitteckert S, Härtle JD, Houghton LA, Dweep H, Fortea M, et al. miR-16 and miR-103 impact 5-HT4 receptor signalling and correlate with symptom profile in irritable bowel syndrome. Sci Rep. 2017;7:19.
63. Hong S, Zheng G, Wiley JW. Epigenetic regulation of genes that modulate chronic stress-induced visceral pain in the peripheral nervous system. Gastroenterology. 2015;148:148-157.e7.
64. Tran L, Schulkin J, Ligon CO, Greenwood-Van MB. Epigenetic modulation of chronic anxiety and pain by histone deacetylation. Mol Psychiatry. 2015;20:1219–31.
65. Staudacher HM, Mikocka-Walus A, Ford AC. Common mental disorders in irritable bowel syndrome: pathophysiology, management, and

considerations for future randomised controlled trials. Lancet Gastroenterol Hepatol. 2021;6:401–10.
66. Carabotti M, Scirocco A, Maselli MA, Severi C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Ann Gastroenterol. 2015;28:203.
67. Fadgyas-Stanculete M, Buga A-M, Popa-Wagner A, Dumitrascu DL. The relationship between irritable bowel syndrome and psychiatric disorders: from molecular changes to clinical manifestations. J Mol Psychiatry. 2014. https://doi.org/10.1186/2049-9256-2-4.
68. Evrensel A, Ceylan ME. The gut-brain axis: the missing link in depression. Clin Psychopharmacol Neurosci. 2015;13:239–44.
69. Smriga M, Torii K. L-Lysine acts like a partial serotonin receptor 4 antagonist and inhibits serotonin-mediated intestinal pathologies and anxiety in rats. Proc Natl Acad Sci USA. 2003;100:15370–5.
70. Bharucha AE, Zinsmeister AR, Camilleri M, Haydock S, Ferber I, Burton D, et al. Effects of a serotonin 5-HT4 receptor antagonist SB-207266 on gastrointestinal motor and sensory function in humans. Gut. 2000;47:667–74.
71. Tam FS-F, Hillier K, Bunce KT. Characterization of the 5-hydroxytryptamine receptor type involved in inhibition of spontaneous activity of human isolated colonic circular muscle. Br J Pharmacol. 1994;113:143–50.
72. Papakostas GI, Mischoulon D, Shyu I, Alpert JE, Fava M. S-adenosyl methionine (SAMe) augmentation of serotonin reuptake inhibitors for antidepressant nonresponders with major depressive disorder: a double-blind, randomized clinical trial. Am J Psychiatry. 2010;167:942–8.
73. Sharma A, Gerbarg P, Bottiglieri T, Massoumi L, Carpenter LL, Lavretsky H, et al. S-adenosylmethionine (SAMe) for neuropsychiatric disorders: a clinician-oriented review of research. J Clin Psychiatry. 2017;78:e656–67.
74. Moreau B, Clement P, Theoret Y, Seidman EG. Allopurinol in combination with thiopurine induces mucosal healing and improves clinical and metabolic outcomes in IBD. Therap Adv Gastroenterol. 2017;10:819–27.
75. Lee YT, Hu LY, Shen CC, Huang MW, Tsai SJ, Yang AC, et al. Risk of psychiatric disorders following irritable bowel syndrome: a nationwide population-based cohort study. PLoS ONE. 2015;10:e0133283.
76. Sibelli A, Chalder T, Everitt H, Workman P, Windgassen S, Moss-Morris R. A systematic review with meta-analysis of the role of anxiety and depression in irritable bowel syndrome onset. Psychol Med. 2016;46:3065–80.
77. Goodwin L, White PD, Hotopf M, Stansfeld SA, Clark C. Life course study of the etiology of self-reported irritable bowel syndrome in the 1958 British birth cohort. Psychosom Med. 2013;75:202–10.

Publisher's Note

American College of Neuropsychopharmacology

www.nature.com/npp

ARTICLE OPEN

Shared genetic architecture between attention-deficit/hyperactivity disorder and lifespan

Laura Vilar-Ribó[1,2,3,10], Judit Cabana-Domínguez[1,2,3,10], Lourdes Martorell[3,4], Josep Antoni Ramos-Quiroga[1,2,3,5], Sandra Sanchez-Roige[6,7], Abraham A. Palmer[6,7], Elisabet Vilella[3,4], Marta Ribasés[1,2,3,8], Gerard Muntané[3,4,9,10]✉ and María Soler Artigas[1,2,3,8,10]✉

There is evidence linking ADHD to a reduced life expectancy. The mortality rate in individuals with ADHD is twice that of the general population and it is associated with several factors, such as unhealthy lifestyle behaviors, social adversity, and mental health problems that may in turn increase mortality rates. Since ADHD and lifespan are heritable, we used data from genome-wide association studies (GWAS) of ADHD and parental lifespan, as proxy of individual lifespan, to estimate their genetic correlation, identify genetic loci jointly associated with both phenotypes and assess causality. We confirmed a negative genetic correlation between ADHD and parental lifespan (rg = −0.36, $P = 1.41e{-}16$). Nineteen independent loci were jointly associated with both ADHD and parental lifespan, with most of the alleles that increased the risk for ADHD being associated with shorter lifespan. Fifteen loci were novel for ADHD and two were already present in the original GWAS on parental lifespan. Mendelian randomization analyses pointed towards a negative causal effect of ADHD liability on lifespan ($P = 1.54e{-}06$; Beta = −0.07), although these results were not confirmed by all sensitivity analyses performed, and further evidence is required. The present study provides the first evidence of a common genetic background between ADHD and lifespan, which may play a role in the reported effect of ADHD on premature mortality risk. These results are consistent with previous epidemiological data describing reduced lifespan in mental disorders and support that ADHD is an important health condition that could negatively affect future life outcomes.

Neuropsychopharmacology (2023) 48:981–990; https://doi.org/10.1038/s41386-023-01555-x

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that emerges in childhood and often persists into adulthood, affecting approximately 5.3% of children and adolescents and 2.8% of adults [1, 2]. It is characterized by age-inappropriate symptoms of inattention, impulsivity, and hyperactivity, which have a severe impact on the individual's social, emotional and psychological functioning, often representing an entry point into a poor life trajectory [3].

There is increasing evidence linking ADHD to a shorter life expectancy and mortality rates in individuals with ADHD are two to five times higher than in individuals without ADHD [4, 5]. Besides natural causes [5], the higher risk of early mortality in individuals with ADHD appears to be largely due to misadventure including a high propensity for accidents and injuries and an elevated risk of suicide [4–9].

ADHD often co-occurs with other mental and somatic comorbid disorders, traits and behaviors that are likely to increase mortality rates [10, 11]. These include (i) mental health problems, such as oppositional defiant disorder, conduct disorder, mood and anxiety disorders, and substance use disorder [8, 12–14]; (ii) comorbid somatic disorders such as obesity [3], asthma [15] and diabetes [16, 17]; (iii) harmful lifestyle behaviors, such as unhealthy eating habits or smoking [18–20]; and (iv) educational underachievement, low income, social adversity, delinquency and aggression [3, 21]. However, despite the high rate of comorbid conditions also linked to an excess mortality [12], these do not fully explain the risk of death observed in individuals with ADHD [4, 5, 8], indicating that ADHD itself is a health condition that confers an increased risk of mortality. For instance, inattention and impulsivity may directly increase proneness to risk-taking behaviors and therefore risk for accidental injuries, leading to reduced lifespan [5].

ADHD and lifespan are complex traits influenced both by genetic and environmental factors. Heritability is estimated to be around 70–80% for ADHD and 7–16% for human lifespan [22]. Genome-wide association studies (GWAS) identified genetic loci associated with both ADHD and lifespan [23, 24], although a large

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [4]Hospital Universitari Institut Pere Mata, Institut d'Investigació Sanitària Pere Virgili (IISPV)-CERCA, Universitat Rovira i Virgili, Reus, Spain. [5]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [6]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [7]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [8]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. [9]Institut de Biologia Evolutiva (UPF-CSIC), Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain. [10]These authors contributed equally: Laura Vilar-Ribó, Judit Cabana-Domínguez, Gerard Muntané, María Soler Artigas.
✉email: muntaneg@peremata.com; maria.soler@vhir.org

Received: 31 May 2022 Revised: 3 February 2023 Accepted: 20 February 2023
Published online: 11 March 2023

part of their heritability still remains to be explained. Some genes known to be related with the risk of developing ADHD, such as those related to the dopaminergic system, have also been associated with shorter life expectancy [19], and a negative genetic correlation between ADHD and parental age at death has been described [24]. These data support observational studies showing an association between ADHD with both elevated mortality risk and reduced estimated life expectancy in adulthood, and suggest that the underlying genetic background of ADHD and lifespan may overlap.

In the present study, we aim to examine the shared genetic architecture and the nature of the relationship between ADHD and parental lifespan, as a proxy for individual lifespan, using available GWAS data on both phenotypes by: (i) estimating their genetic correlation; (ii) performing a cross-trait analysis and (iii) testing the causal role of ADHD on lifespan.

MATERIALS AND METHODS

GWAS samples and data processing

GWAS summary statistics on ADHD were obtained from Demontis et al., comprising a total of 19,099 individuals with ADHD and 34,194 healthy controls, all of European ancestry [24]. Summary statistics on parental lifespan were obtained from Timmers et al. and included data from about 1 million individuals of European ancestry [23]. The GWAS summary statistics were referenced to a set of 9,546,816 SNPs generated from the 1000 Genomes Project Phase 3 European reference panel (http://www.internationalgenome.org/). SNPs that were non-biallelic, without rsIDs, duplicated, or with strand-ambiguous alleles were removed. We also filtered out SNPs with INFO scores <0.9 in the summary statistics files, those mapping to the extended major histocompatibility complex (MHC, genomic position in hg 19; chr6:25,119,106–33,854,733) and the 8p23.1 region (chr8:7,200,000–12,500,000), which are prone to rearrangements, SNPs located on the X, Y and mitochondrial chromosomes, and SNPs with sample sizes 5 standard deviations away from the mean. Finally, a common set of 3,206,697 SNPs were kept in both summary statistics. All *P* values were adjusted for standard genomic control (GC).

Genetic correlation and pleiotropy assessment

Linkage Disequilibrium Score Regression (LDSC) was used to calculate genome-wide genetic correlations across the traits studied [25].

The shared polygenic architecture between the two traits was assessed by means of stratified cross-phenotype Q-Q plots. *P* values for the primary trait were plotted conditioning on different association strengths ($P < 1$, 0.1, 0.01 and 1e−03) with the secondary trait. Thus, the visualization of a leftward shift in the primary trait of interest, as a function of increasingly strict *P* value thresholds in the secondary trait, was an indicator of a shared polygenic architecture between the two traits. To test for SNP-based heritability enrichment of a trait conditioned on different association strengths with a secondary trait, we used stratified LDSC [26]. As a reference panel for linkage disequilibrium (LD) we used the 1000 Genomes Project Phase 3 European reference panel [27].

Cross-trait analysis

To identify genetic loci jointly associated with ADHD and parental lifespan we estimated the conjunction FDR (conjFDR), defined as "the posterior probability that a given SNP is null for both phenotypes simultaneously when the *P* values for both phenotypes are as small as or smaller than the observed *P* values", using pleioFDR (https://github.com/precimed/pleiofdr) [28]. We kept all SNPs with conjFDR <0.1 for functional studies and reported independent SNPs with a conjFDR <0.05. Independent genomic loci were identified through clumping ($r^2 = 0.05$, kb = 500) using the 1000 Genomes Project Phase 3 European as the reference panel for LD computation and PLINK 1.09 [27, 29]. We evaluated the directional effects of shared loci between ADHD and parental lifespan by comparing z-scores between the original GWAS summary statistics. Overlap between hits from the cross-trait analysis and genome-wide significant associations reported in the original GWAS results ($P < 5e{-}08$) was assessed according to distance (+/−250 kb) and linkage disequilibrium ($r^2 > 0.1$) between of the cross-trait lead SNPs and previous reported hits.

Functional annotation

Functional annotation of all SNPs with a conjFDR value <0.10 and an LD $r^2 \geq 0.6$ with one of the independent significant SNPs was performed in FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies, https://fuma.ctglab.nl/) [30]. We combined data from the Combined Annotation Dependent Depletion (CADD) scores [31], which predict how deleterious the SNP effect is on protein structure/function based on 63 functional annotations, and RegulomeDB scores [32], a categorical score that estimates the regulatory functionality of SNPs based on existing functional data (annotation to cis-eQTLs, expression quantitative trait loci) and evidence for transcription factor binding. CADD ≥ 12.37 was considered as the threshold for deleterious variants, RegulomeDB scores <3 were likely to have a regulatory function, and minimum chromatin states between 1–7 were considered open chromatin states. The NHGRI-EBI GWAS catalog [33], the release from the 15th of September 2021, was used to identify traits previously associated with the SNPs of interest and we queried SNPs for known brain eQTLs using the Genotype-Tissue Expression (GTEx) v8 [34] and BRAINEAC [35]. We also used FUMA to map SNPs to genes based on physical proximity (using default parameters) and eQTL in brain (based on GTEx v8 and BRAINEAC) to test for enrichment on gene ontology and biological pathways of the mapped genes. All analyses were corrected for multiple comparisons using False Discovery Rate.

Causality analyses

GWAS summary statistics. To assess the causal effect of the genetic liability of ADHD on lifespan, the summary statistics from Pilling et al. [36], generated from a linear mixed-effects model in 208,118 individuals of European ancestry, was used instead of Timmers et al. [23]. The latter used a survival analysis to study parental lifespan, which may produce a significant bias in MR analyses [37]. LDSC was used to calculate genetic correlations between both studies and between ADHD and parental lifespan in the study by Pilling et al. [25, 36]. For mediation analyses, as defined below we used summary statistics from Sanchez-Roige et al. for total impulsivity score, lack of premeditation and positive urgency [38].

Mendelian randomization. Causality between ADHD (as the exposure), and parental lifespan (as the outcome), was assessed by two-sample MR using the TwoSampleMR and MRPRESSO R packages [39, 40]. After clumping ($r^2 = 0.05$, kb = 500) with PLINK 1.09 [29], independent SNPs were selected using a *P* value threshold of 5e−08 in the ADHD GWAS to be used as instruments. The multiplicative random effects inversed-variance weighted (IVW) was used as the main method to obtain the average effect across genetic variants. For IVW results to be valid, genetic variants used as instruments must meet three assumptions: (i) robust association with the exposure, (ii) absence of horizontal pleiotropy, or association with the outcome through an exposure-independent pathway, and (iii) independence of confounders influencing exposure and outcome. Additional MR methods were implemented, as sensitivity analyses, for significant IVW results (IVW $P < 0.05$) to assess the robustness of the findings under weaker assumptions: (i) the weighted median method, which under equal weights requires at least half of the variants to be valid instruments and is robust to outliers [41]; (ii) the MR-PRESSO method, which tests for horizontal pleiotropy (MR-PRESSO global test), and if detected, eliminates horizontal pleiotropic outliers and then performs the IVW method using the remaining instruments [40]; and (iii) the MR-Egger method, which is affected by outlier data points but allows all genetic variants to have pleiotropic effects assuming that these effects are independent of variant–exposure associations [42]. MR-Egger also implements a pleiotropy test, however, when the NO Measurement Error (NOME) assumption is violated ($I^2gx < 0.9$) MR-Egger causal estimates are biased towards the null, and the type I error of the pleiotropy test can be inflated. For this reason, when $I^2gx < 0.6$, MR-Egger results were disregarded. In addition, heterogeneity tests and leave-one-out analyses were performed and scatter, funnel and forest plots were generated. The analysis in the opposite direction, testing the effect of parental lifespan on ADHD, was considered as a negative control. Finally, as an alternative method, we also used the ρ-HESS implementation to prioritize putative causal models between pairs of traits, as previously described [43, 44].

Multivariate Mendelian Randomization (MVMR). To explore whether a putative causal effect of the genetic liability of ADHD on lifespan is mediated by impulsive personality traits we used MVMR and summary

statistics from the GWAS by Sanchez-Roige et al. undertaken in over 20,000 individuals [38]. Out of the ten traits analyzed by Sanchez-Roige et al. [38]. we selected as potential mediators those reported to have a significant genetic correlation with ADHD, prioritizing total scores when available, namely total impulsivity score (measured with the BIS-11 [45]), lack of premeditation and positive urgency (both measured using the UPPS-P Impulsive Behavior Scale [46]). We used the same clumping parameters as for the main analysis ($r^2 = 0.05$, kb = 500) and, given that no genome-wide significant SNPs were available, a threshold of 5e−06 was chosen to select independent SNPs for these traits.

Causal Analysis Using Summary Effect estimates (CAUSE). We also explored causality between ADHD and lifespan using the *cause* R package [47] considering independent variants ($r^2 = 0.05$, kb = 500). The CAUSE method uses a more permissive threshold for variant selection than MR ($P < 1e{-}03$). In addition, CAUSE allows all variants to show uncorrelated pleiotropy, also accounted for by MR-Egger or MR-PRESSO, but it also allows a subset of variants to show correlated pleiotropy, when they affect exposure and outcome through a shared heritable factor. CAUSE compares two nested models by measuring how well the posterior distributions of a particular model fit the data: (i) a sharing model, which only allows for pleiotropic effects and no causal effects; and (ii) a causal model, which also allows both for pleiotropic and causal effects.

RESULTS

We found strong evidence of negative SNP-based genetic correlation between ADHD and parental lifespan, used as a proxy for individual lifespan (rg = −0.36, $P = 1.41e{-}16$). Partitioning ADHD SNP heritability on different parental lifespan *P* value thresholds showed enrichment of ADHD SNP heritability ($P = 2.73e{-}07$ and $P = 1.85e{-}03$ conditioning ADHD on parental lifespan $P < 0.1$ and $P < 0.01$, respectively; Supplementary Table 1a). Furthermore, conditioning parental lifespan SNPs on ADHD *P* values showed parental lifespan SNP heritability enrichment ($P = 1.09e{-}06$ and 7.69e−03 conditioning parental lifespan on ADHD $P < 0.1$ and $P < 0.01$, respectively; Supplementary Table 1b). These enrichments were consistent with stratified cross-phenotype Q-Q plots showing a stronger leftward deflection from the null expectation when conditioning ADHD on increasing levels of association for parental lifespan and vice-versa (Fig. 1).

The cross-trait analysis showed a total of 19 independent genomic loci associated with both ADHD and parental lifespan with a conjFDR < 0.05 (Fig. 2A and Supplementary Fig. 1), 15 of which were not identified in the original GWAS on ADHD and two were already present in the original GWAS on parental lifespan (Table 1). Functional annotation of all SNPs with conjFDR < 0.1 at these 19 loci ($n = 479$ SNPs) revealed that 92.1% of these loci lay on regions of open chromatin and most of the signals were intergenic, intronic or located in intronic non-coding RNA (ncRNA) genes. In addition, several SNPs in these genomic risk loci were likely to affect the binding of transcription factors or had CADD scores >12.37, suggesting high deleteriousness (Fig. 3, Table 1 and Supplementary Table 2). In addition, we found that 42% of the SNPs were eQTL for at least one gene in one brain area, according to GTEx v8 [34] and BRAINEAC [35] (Supplementary Table 3). Finally, 23 SNPs at 11 different genomic risk loci were previously associated with different traits, mainly related to lifetime risky behaviors (e.g., smoking, general risk tolerance and number of sexual partners), psychiatric disorders (e.g., schizophrenia, ADHD and depression) and metabolic alterations (e.g., metabolic syndrome, triglyceride or cholesterol levels and blood pressure; Supplementary Table 4). The 19 risk loci identified mapped 40 genes (Supplementary Table 5), which were enriched in genes previously associated with cognitive performance, smoking and metabolite levels according to the GWAS catalog [33] (Supplementary Fig. 2), but no association with either biological pathways or differential tissue expression from GTEx were found.

To explore the landscape of pleiotropic effects further, we examined the direction of the effects of the lead SNPs of all independent loci from the cross-trait analysis on both ADHD and parental lifespan and found an opposite direction of effect for 95% of them ($n = 18$), with alleles that increased the risk for ADHD also shortening lifespan (Fig. 2B). ADHD hits [24] showed a significant negative correlation with parental lifespan with p-HESS, whereas no significant correlation was found between parental lifespan hits [23] and ADHD, which is consistent with a putative causal relationship of ADHD on shortened lifespan (Supplementary Figure 3**).** In line with these results, MR analyses showed evidence of a negative causal effect of ADHD liability on lifespan (IVW Beta = −0.07 and $P = 1.54e{-}06$; weighted median Beta = −0.05 and $P = 6.52e{-}03$; Supplementary Table 6 and Fig. 4). There was no evidence of horizontal pleiotropy according to MR-PRESSO (global test $P = 0.43$) or heterogeneity ($I^2 = 11.15$). The results were not driven by a single SNP (Supplementary Fig. 4). Variants included in the MR analyses are shown in Supplementary Table 7. MR testing the causal relationship of parental lifespan on ADHD, as a negative control, showed no significant results (IVW $P = 0.66$). MVMR analyses showed that the effect of ADHD liability on lifespan remained unchanged when accounting for total impulsivity score or lack of premeditation, however it decreased (from −0.07 to −0.06) when taking into account the effect of positive urgency, suggesting that positive urgency may be mediating around 11% of the effect of ADHD liability on lifespan (Supplementary Table 8). CAUSE did not provide evidence of correlated pleiotropy (Supplementary Table 9) or of a causal effect of ADHD liability on lifespan (Beta = −0.01 and CI = (−0.03, 0)) (Supplementary Tables 6 and 9). Given that survival analyses, including the study conducted by Timmers et al. used in the present study, may bias MR results, we conducted the causality analyses using summary statistics on parental lifespan from a second GWAS by Pilling et al. [36]. The genetic correlation between both summary statistics on parental lifespan was very high (rg = −0.93, $P = 2.98e{-}184$), and parental lifespan from the latest was also negatively correlated with ADHD (rg = −0.41, $P = 2.66e{-}14$).

DISCUSSION

The present study provides the first evidence of a common genetic background between ADHD and lifespan. Extensive literature supports that individuals with ADHD have an increased risk of premature death and a shorter life expectancy, which increases in females and depends on age at first ADHD diagnosis [4, 8]. In addition, the presence of other comorbid psychiatric conditions, such as oppositional defiant disorder, conduct disorder or substance use disorder, further increases ADHD mortality [4, 8]. This excess mortality and decreased life expectancy in ADHD subjects are mainly driven by unnatural causes being accidents the most common cause of early death [4, 9]. In addition, an increased risk of suicide, traffic violations or trauma have been also described in ADHD [8, 48–52]. At the same time, the impact of the disorder on other adverse life-course outcomes, such as poorer educational attainment, unemployment, delinquency or lower socioeconomic status, may also increase the risk of mortality [11]. Strengthening the results of observational studies, we provide evidence of shared genetic signatures and negative genetic correlation between ADHD and lifespan, which further supports the existing evidence that ADHD represents an entry point into a negative life trajectory [3].

To provide potential pleiotropic molecular mechanisms underlying this association, we performed a cross-trait analysis on ADHD and parental lifespan and identified 19 independent loci jointly associated with both traits, including 15 novel hits for ADHD [23, 24]. All of them but one (95%) showed consistent direction of the effect, with risk alleles for ADHD shortening lifespan, which is

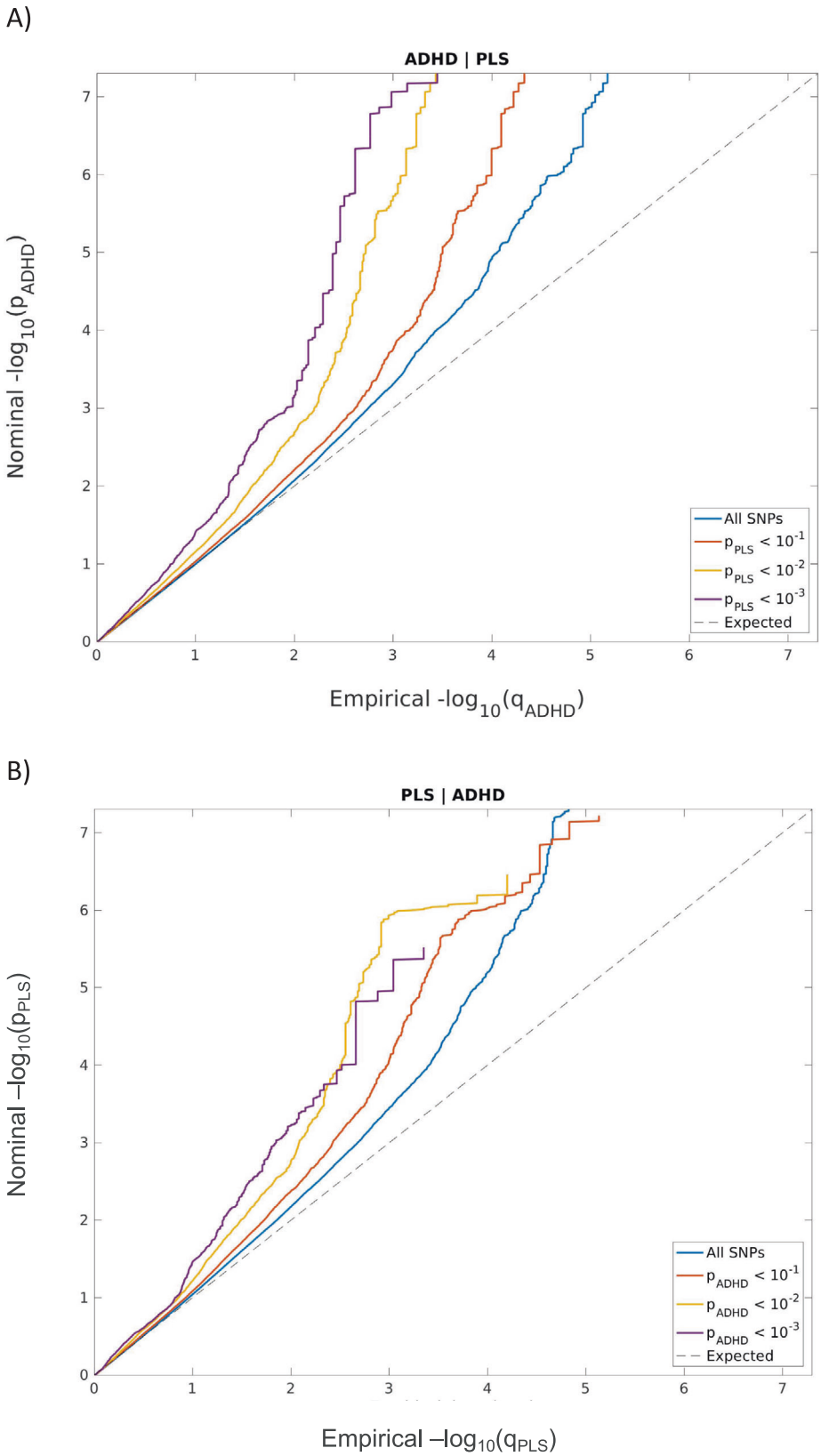


Fig. 1 Stratified cross-phenotype Q-Q plots. Nominal versus empirical ($-\log_{10}$) *P* values (corrected for inflation) are shown in **A** ADHD as a function of significance with parental lifespan and **B** parental lifespan as a function of significance with ADHD, at the level of $P < 0.1$ (red line), $P < 0.01$ (yellow line), and $P < 1e{-}03$ (purple line). The blue line indicates the standard enrichment of **A** ADHD or **B** parental lifespan including all SNPs, irrespective of their association with the secondary trait (i.e., parental lifespan or ADHD, respectively). The gray dashed line indicates the null distribution of *P* values. LD score regression intercepts for ADHD and parental lifespan full summary statistics were 1.04 and 1.05 respectively. PLS parental lifespan.

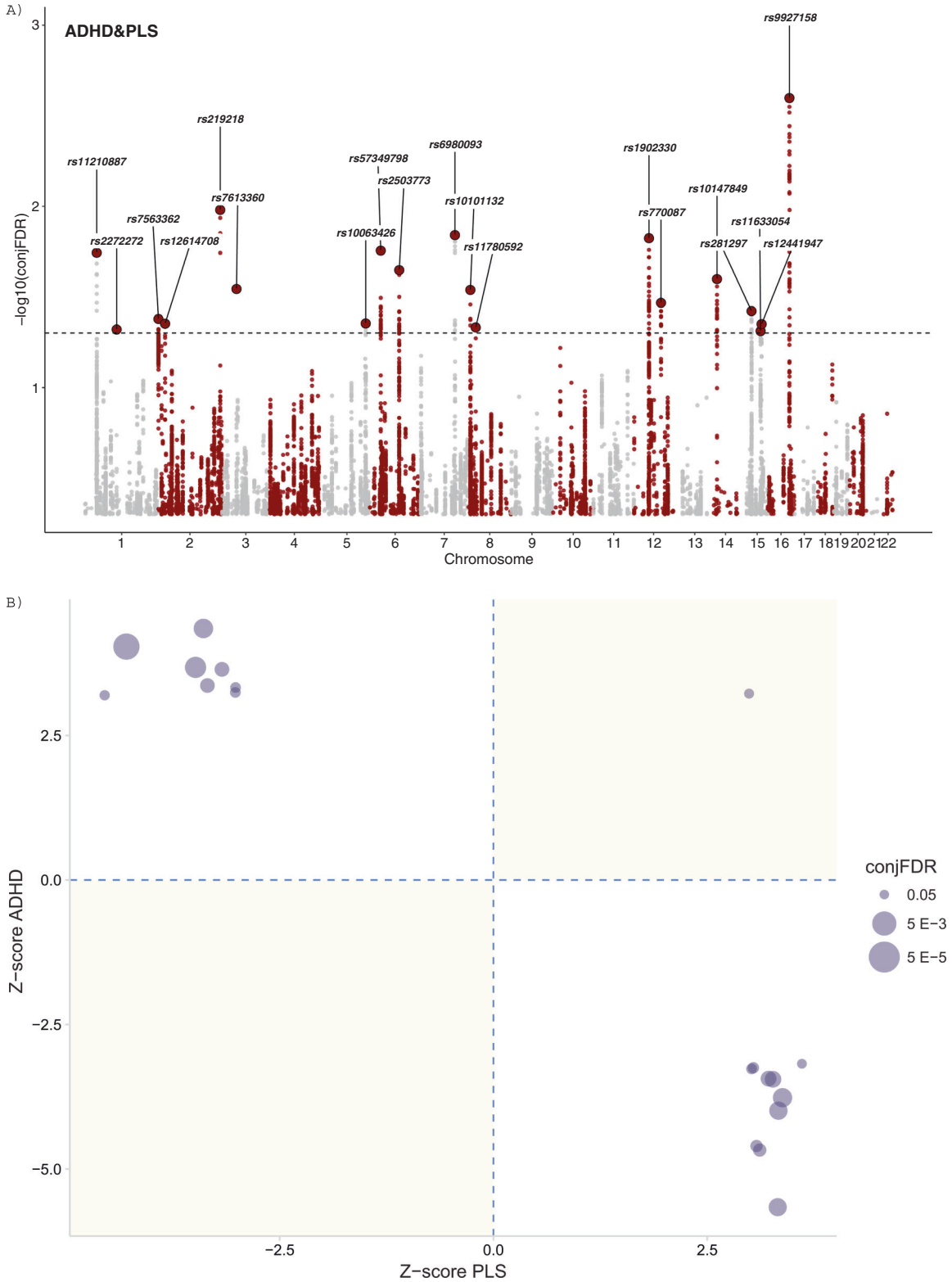


Fig. 2 Cross-trait analysis results. A Manhattan plot for independent ($r^2 < 0.05$) loci associated with both ADHD and parental lifespan after excluding SNPs in the MHC and the 8p23.1 regions. The dashed black line represents the P_{FDR} threshold of 0.05. **B** Pleiotropy plot. For lead SNPs ($n = 19$), *P* values and the direction of the effects (z-scores) of the derived alleles are plotted for parental lifespan (x-axis) against ADHD (y-axis). PLS parental lifespan.

consistent with the negative genetic correlation between ADHD and lifespan, adding further evidence for the contribution of a shared biological architecture. Interestingly, functional annotation of top hits from the cross-trait analysis highlighted loci previously associated with ADHD and/or reduced life expectancy-related phenotypes, including other psychiatric disorders (e.g., schizophrenia and major depression), lifetime risk behaviors (e.g., number of sexual partners and risk tolerance) or metabolic

Table 1. List of loci jointly associated with ADHD and parental lifespan (PLS) at conjunction FDR < 0.05.

Locus	SNP	CHR	BP	nearestGene	ANNOVAR	CADD	RDB	minChr State	Effect allele	Z-score ADHD	*P* value ADHD	Z-score PLS	*P* value PLS	conjFDR ADHD_PLS	New locus in	
															ADHD	PLS
1	rs11210887	1	44076019	PTPRF	Intronic	12.97	6	4	A	−6.41406	1.48E−10	3.84263	1.22E−04	1.80E−02	n	y
2	rs2272272	1	110009802	SYPL2	Intronic	22.2	NA	1	A	3.65022	2.63E−04	3.45524	5.50E−04	4.77E−02	y	y
3	rs7563362	2	620297	AC068490.2	ncRNA_intronic	0.133	6	5	A	−3.67479	2.35E−04	3.51328	4.43E−04	4.18E−02	y	n
4	rs12614708	2	22568238	AC009965.2	Intergenic	0.259	6	5	A	3.68099	2.39E−04	−3.4884	4.86E−04	4.43E−02	y	y
5	rs219218	2	205007037	AC093326.3	Intergenic	0.041	5	2	T	4.15596	3.13E−05	−4.0265	5.66E−05	1.05E−02	y	y
6	rs7613360	3	49916710	ACTBP13	Intergenic	2.031	5	5	T	4.11314	3.71E−05	−3.6695	2.43E−04	2.85E−02	y	y
7	rs10063426	5	154775154	CTC-447K7.1	Intergenic	1.814	6	9	A	−3.69334	2.17E−04	3.48961	4.84E−04	4.42E−02	y	y
8	rs57349798	6	37486052	RP11-436D23.1	ncRNA_intronic	1.314	NA	9	A	−4.5238	6.27E−06	3.852	1.17E−04	1.75E−02	y	y
9	rs2503773	6	98537145	RP1-153P14.8	ncRNA_intronic	2.452	5	2	A	−3.9189	9.52E−05	3.7788	1.58E−04	2.24E−02	y	y
10	rs6980093	7	114162740	FOXP2	Intronic	0.41	6	5	A	4.9354	8.39E−07	−3.9201	8.85E−05	1.44E−02	n	y
11	rs10101132	8	9616553	TNKS	Intronic	9.165	7	4	A	3.80143	1.40E−04	−3.8661	1.11E−04	2.89E−02	y	y
12	rs11780592	8	27418747	GULOP	ncRNA_intronic	1.308	1f	4	A	3.62034	2.97E−04	−5.2548	1.48E−07	4.65E−02	y	y
13	rs1902330	12	49939953	KCNH3	Intronic	8.355	5	5	A	−4.26642	2.00E−05	3.90708	9.34E−05	1.50E−02	y	y
14	rs770087	12	89744773	DUSP6	Exonic	23	NA	1	A	−5.29553	1.23E−07	3.59893	3.20E−04	3.40E−02	n	y
15	rs10147849	14	33304431	AKAP6	Intergenic	6.449	6	5	T	−3.89787	1.00E−04	3.71885	2.00E−04	2.52E−02	y	y
16	rs281297	15	47685504	SEMA6D	Intronic	4.125	5	5	T	−5.22589	1.89E−07	3.556	3.77E−04	3.78E−02	n	y
17	rs11633054	15	77747276	HMG20A	Intronic	7.219	7	4	A	−3.59311	3.20E−04	4.16936	3.05E−05	4.87E−02	y	y
18	rs12441947	15	81031886	ABHD17C	ncRNA_intronic	0.885	7	5	T	3.75846	1.64E−04	−3.4859	4.90E−04	4.46E−02	y	y
19	rs9927158	16	72256156	RP11-328J14.1	Intergenic	0.365	7	1	T	4.56195	4.86E−06	−4.961	7.01E−07	2.52E−03	y	n

ANNOVAR functional variant classification based on position in or outside of a gene, *CADD* Combined Annotation-Dependent depletion score, which predict how deleterious the SNP effect is on protein structure/function, in bold scores >12.37 suggesting high deleteriousness, *RBD* RegulomeDB scores predict likelihood of regulatory functionality, in bold scores lower than 3 that indicate higher likelihood. Further information about RDB scores can be found in Table 2 from the original paper [32]; *minChrState* minimum chromatin state across 127 tissue types, in bold scores lower than 6 that indicate more open chromatin, Also shown the effect allele (used to calculate beta), *p* values and effect sizes from the original summary statistics on Attention-Deficit/Hyperactivity disoder (ADHD) and parental lifespan (PLS), *NA* Not available.

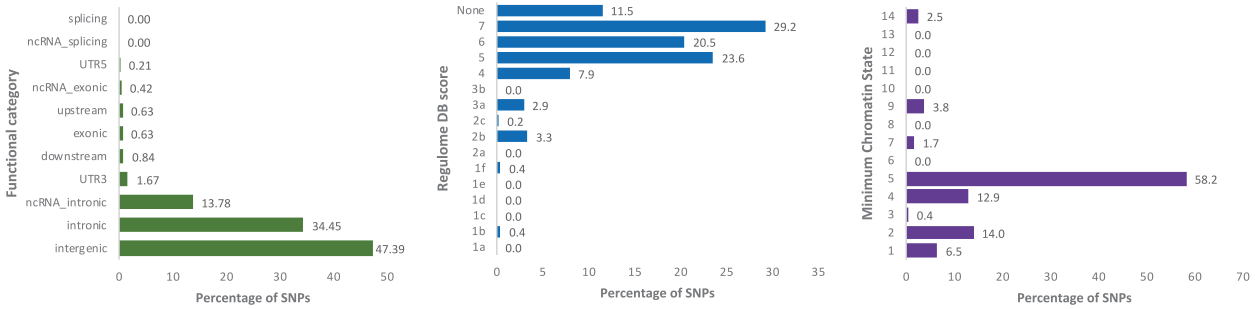


Fig. 3 Functional categories, Regulome DB scores, and Minimum Chromatin States for SNPs within loci jointly associated with ADHD and parental lifespan. Regulome DB score predicts likelihood of regulatory functionality, where lower scores indicate higher likelihood. Further information can be found in Boyle et al. 2012 [32]. Minimum Chromatin State across 127 tissue and cell types, lower scores indicate higher accessibility, with states 1–7 referring to open chromatin states.

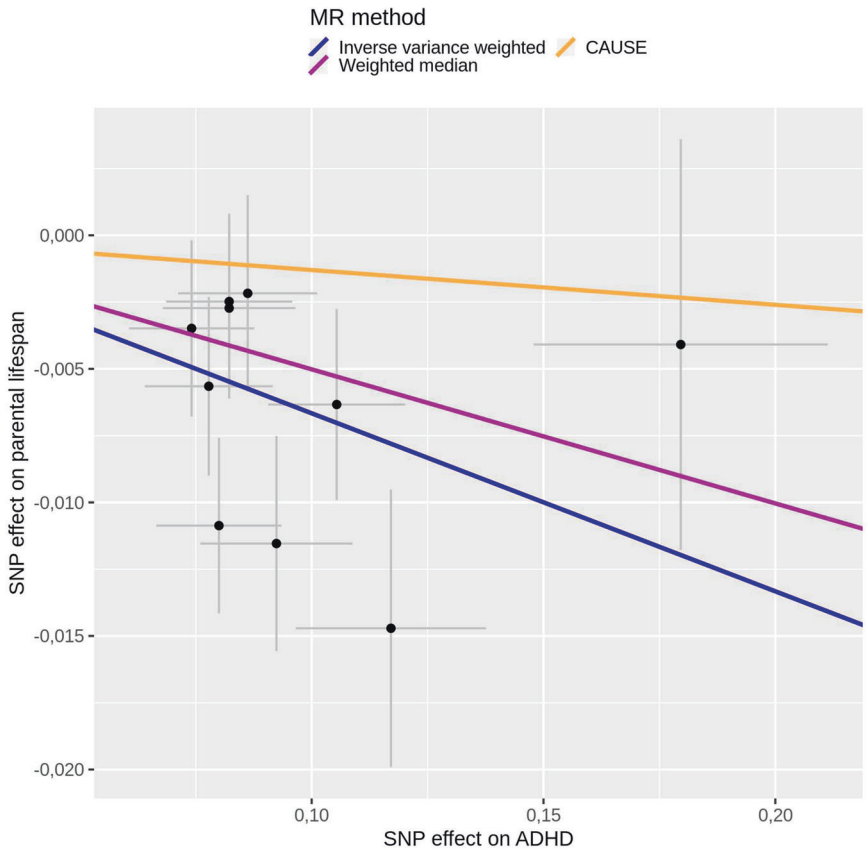


Fig. 4 Mendelian randomization results with ADHD as exposure and parental lifespan as outcome. Scatter plot of SNP effect estimates on ADHD vs. effect estimates on parental lifespan. Lines are drawn for inverse-variance weighted, weighted median and CAUSE, with the slope of each line corresponding to the estimated causal effect. Given that $Igx^2 = 0.36$, MR-Egger results are not reported.

alterations (e.g., metabolic syndrome, and triglyceride and cholesterol levels). This further supports that shared risk factors may not be specific of ADHD or lifespan but also contribute to other related disorders and behaviors that could mediate, in part, this relationship. For instance, ADHD may directly increase propensity for risky behaviors and thus lead to a shorter life expectancy [5].

The pleiotropic risk loci identified were enriched in introns and open chromatin regions, and almost half of them (42%) were cis-eQTLs tags for at least one gene in the brain, supporting their putative causal effect. Among the identified risk loci, we highlight new relevant candidate genes. For instance, the *TNKS* gene encoding tankyrase1, a protein involved in telomere maintenance [53], which is an essential cellular function closely related to ageing and longevity [54]. Interestingly, several genetic variants in this locus are cis-eQTL for *TNKS* in putamen, a key area in the basal ganglia previously related with ADHD [55–57]. In addition, *TNKS* was previously associated with multiple psychiatric conditions (e.g., risk tolerance and neuroticism), neurological (e.g., Alzheimer's disease and epilepsy) and metabolic disorders (e.g., blood pressure, obesity, diabetes and stroke) [58–64], which are highly comorbid with ADHD and associated with increased morbidity and mortality [65–67]. We also found genetic variants in other two loci encompassing *AKAP6* and *SEMA6D* genes, previously

associated with increased risk of schizophrenia, lower cognitive ability and educational attainment [68], all of them related with shorter life expectancy [69–72]. In addition, we found other loci including genes with relevant brain functions: *SYPL2*, which encodes a vesicular transmembrane protein that belongs to the synaptophysin family, key elements for the regulation of neuronal synaptic vesicles [73, 74]; and *HMG20A*, a non-histone chromosomal factor that regulates gene expression through changes in chromatin conformation, also involved in the regulation of neuronal differentiation [75, 76].

Our results seem to reflect an effect of ADHD on premature mortality risk and are consistent with previous epidemiological data describing a shortened lifespan in many mental disorders [4, 12]. A recent study suggests that there is no causal relationship between schizophrenia and parental lifespan [69], which can indicate that our results may not be extensive to other psychiatric disorders. Further studies are needed to clarify how specific the effect on parental lifespan is to ADHD. We also found suggestive evidence of a mediating role for positive urgency (i.e. the proclivity for rash action when feeling positive emotion) in the relationship between the genetic liability of ADHD and lifespan. This is consistent with a reported predictive effect of positive urgency on ADHD-related traits such as illegal drug use and risky sexual behavior [77]. Nevertheless these findings should be interpreted with caution, since the evidence found for a causal relationship of ADHD liability on shorter life expectancy was not conclusive. While IVW MR, together with all the sensitivity analyses performed and putative causality inferred with p-HESS supported a negative causal effect of genetic liability of ADHD on lifespan, CAUSE did not confirm this causal relationship. This highlights that inferring causal relationships between related traits remains a challenge to date and that new methods and larger sample sizes are needed to understand the common genetic architecture underlying such complex relationships.

Our results, however, should be interpreted in the context of some considerations:

First, parental lifespan of genotyped subjects was used as a surrogate trait for individual lifespan in the present generation. This kinship cohort design is supported by previous results reporting that parental lifespan predicts the long-term mortality of their offspring. However, the use of indirect genotypes reduces the effective sample size of the study [23, 78]. This, together with the modest heritability estimates for human lifespan (ranging from 7% to 16%) [23, 79] limits the power of our study and precludes the identification of shared genetic variants with smaller effect sizes.

Second, to avoid the potential bias introduced by using summary statistics generated through a survival analysis [37], such as that conducted by Timmers et al. [23], we selected the results of Pilling et al. [36] to assess causality between ADHD and parental lifespan. The smaller sample size of this study ($n = 208{,}118$) further reduced the statistical power and may have led to the inconsistent results observed [36]. CAUSE was developed with the aim of avoiding false positive results due to correlated horizontal pleiotropy, but in the present study it did not detect a causal effect but neither did it detect this kind of pleiotropy. Recent studies suggest that in some scenarios CAUSE tends to underestimate causal effect sizes in comparison to MR and produce overly conservative *P* values [80, 81], which may suggest, in our case, limited power rather than lack of a causal relationship, supporting further studies with larger sample sizes to clarify this issue.

Third, despite the shared genetic architecture underlying ADHD in children and adults [82], the risk of premature death also depends on age at diagnosis, where individuals diagnosed with ADHD in adulthood appear to have a higher risk of death than do those diagnosed in childhood or adolescence [4]. These data may suggest that persistent ADHD represents a more severe form of the disorder and that the role of ADHD on the overall life expectancy across age groups deserves further investigation. In addition, we did not control our analysis for other potential confounders that may bias our results. Increased mortality in individuals with ADHD depends on sex or the presence of co-occurring disorders [4, 8]. Also severity of ADHD symptoms, impairment and/or pharmacological treatment may influence life outcomes [83–85] and their role in the relationship between ADHD, premature death and reduced overall life expectancy deserve further investigation.

In summary, our results are in agreement with observational studies supporting ADHD as the entrance into a detrimental life trajectory, support negative genetic correlation between ADHD and lifespan and show common genetic loci shared between them, most of which showed risk-increasing effects on ADHD and reduced overall life expectancy. These results confirm the general pattern of increased mortality rates and reduced overall life expectancy associated with ADHD and highlight the need for further studies on larger datasets to better understand the common genetic architecture underlying these complex relationships.

REFERENCES

1. Polanczyk G, De Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry. 2007;164:942–8.
2. Fayyad J, Sampson NA, Hwang I, Adamowski T, Aguilar-Gaxiola S, Al-Hamzawi A, et al. The descriptive epidemiology of DSM-IV Adult ADHD in the World Health Organization World Mental Health Surveys. ADHD Atten Deficit Hyperact Disord. 2017;9:47–65.
3. Franke B, Michelini G, Asherson P, Banaschewski T, Bilbow A, Buitelaar JK, et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. Eur Neuropsychopharmacol. 2018;28:1059–88.
4. Dalsgaard S, Ostergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet. 2015;385:2190–6.
5. Sun S, Kuja-Halkola R, Faraone SV, D'Onofrio BM, Dalsgaard S, Chang Z, et al. Association of psychiatric comorbidity with the risk of premature death among children and adults with attention-deficit/hyperactivity disorder. JAMA Psychiatry. 2019;76:1141–9.
6. Prasad V, West J, Sayal K, Kendrick D. Injury among children and young people with and without attention-deficit hyperactivity disorder in the community: the risk of fractures, thermal injuries, and poisonings. Child Care Health Dev. 2018;44:871–8.
7. Stickley A, Koyanagi A, Ruchkin V, Kamio Y. Attention-deficit/hyperactivity disorder symptoms and suicide ideation and attempts: findings from the Adult Psychiatric Morbidity Survey 2007. J Affect Disord. 2016;189:321–8.
8. Barbaresi WJ, Colligan RC, Weaver AL, Voigt RG, Killian JM, Katusic SK. Mortality, ADHD, and psychosocial adversity in adults with childhood ADHD: a prospective study. Pediatrics. 2013;131:637–44.
9. Catalá-López F, Hutton B, Page MJ, Driver JA, Ridao M, Alonso-Arroyo A, et al. Mortality in persons with autism spectrum disorder or attention-deficit/hyperactivity disorder. JAMA Pediatr. 2022;176:e216401.
10. Faraone SV. Attention deficit hyperactivity disorder and premature death. Lancet. 2015;385:2132–3.
11. Erskine HE, Norman RE, Ferrari AJ, Chan GCK, Copeland WE, Whiteford HA, et al. Long-term outcomes of attention-deficit/hyperactivity disorder and conduct disorder: a systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry. 2016;55:841–50.
12. Nordentoft M, Wahlbeck K, Hällgren J, Westman J, Ösby U, Alinaghizadeh H, et al. Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. PLoS One. 2013;8:e55176.
13. Wilens TE, Spencer TJ. Understanding attention-deficit/hyperactivity disorder from childhood to adulthood. Postgrad Med. 2010;122:97.
14. Kendler KS, Ohlsson H, Sundquist J, Sundquist K. Alcohol use disorder and mortality across the lifespan: a longitudinal cohort and co-relative analysis. JAMA Psychiatry. 2016;73:575–81.
15. Cortese S, Sun S, Zhang J, Sharma E, Chang Z, Kuja-Halkola R, et al. Association between attention deficit hyperactivity disorder and asthma: a systematic review and meta-analysis and a Swedish population-based study. The Lancet. Psychiatry. 2018;5:717–26.

16. Chen MH, Pan TL, Hsu JW, Huang KL, Su TP, Li CT, et al. Risk of type 2 diabetes in adolescents and young adults with attention-deficit/hyperactivity disorder: A nationwide longitudinal study. J Clin Psychiatry. 2018;79:17m11607.
17. Kapellen TM, Reimann R, Kiess W, Kostev K. Prevalence of medically treated children with ADHD and type 1 diabetes in Germany - analysis of two representative databases. J Pediatr Endocrinol Metab. 2016;29:1293–7.
18. Ptacek R, Stefano GB, Weissenberger S, Akotia D, Raboch J, Papezova H, et al. Attention deficit hyperactivity disorder and disordered eating behaviors: links, risks, and challenges faced. Neuropsychiatr Dis Treat. 2016;12:571–9.
19. Barkley RA, Smith KM, Fischer M. ADHD risk genes involved in dopamine signaling and metabolism are associated with reduced estimated life expectancy at young adult follow-up in hyperactive and control children. Am J Med Genet Part B Neuropsychiatr Genet. 2019;180:175–85.
20. Groenman AP, Oosterlaan J, Rommelse N, Franke B, Roeyers H, Oades RD, et al. Substance use disorders in adolescents with attention deficit hyperactivity disorder: a 4-year follow-up study. Addiction. 2013;108:1503–11.
21. Mohr-Jensen C, Steinhausen HC. A meta-analysis and systematic review of the risks associated with childhood attention-deficit hyperactivity disorder on long-term outcome of arrests, convictions, and incarcerations. Clin Psychol Rev. 2016;48:32–42.
22. van den Berg N, Beekman M, Smith KR, Janssens A, Slagboom PE. Historical demography and longevity genetics: back to the future. Ageing Res Rev. 2017;38:28–39.
23. Timmers PRHJ, Mounier N, Lall K, Fischer K, Ning Z, Feng X, et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. Elife. 2019;8:1–40.
24. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.
25. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
26. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–35.
27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
28. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet. 2013;92:197–209.
29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
30. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1–11.
31. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
32. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.
33. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45:D896–901.
34. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.
35. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014;17:1418–28.
36. Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging. 2017;9:2504–20.
37. Cho Y, Rau A, Reiner A, Auer PL. Mendelian randomization analysis with survival outcomes. Genet Epidemiol. 2021;45:16–23.
38. Sanchez-Roige S, Fontanillas P, Elson SL, Gray JC, de Wit H, MacKillop J, et al. Genome-wide association studies of impulsive personality traits (BIS-11 and UPPSP) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. J Neurosci. 2019;39:2562–72.
39. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
40. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8.
41. Bowden J, Davey, Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.
42. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
43. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48:709–17.
44. Major M, Freund MK, Burch KS, Mancuso N, Ng M, Furniss D, et al. Integrative analysis of Dupuytren's disease identifies novel risk locus and reveals a shared genetic etiology with BMI. Genet Epidemiol. 2019;43:629–45.
45. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol. 1995;51:768–74.
46. Cyders MA, Littlefield AK, Coffey S, Karyadi KA. Examination of a short English version of the UPPS-P Impulsive Behavior Scale. Addict Behav. 2014;39:1372–6.
47. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.
48. Ertan C, Özcan ÖÖ, Pepele MS. Paediatric trauma patients and attention deficit hyperactivity disorder: correlation and significance. Emerg Med J. 2012;29:911–4.
49. Balazs J, Kereszteny A. Attention-deficit/hyperactivity disorder and suicide: a systematic review. World J Psychiatry. 2017;7:44.
50. Curry AE, Metzger KB, Pfeiffer MR, Elliott MR, Winston FK, Power TJ. Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder. JAMA Pediatr. 2017;171:756–63.
51. London AS, Landes SD. Attention Deficit Hyperactivity Disorder and adult mortality. Prev Med. 2016;90:8–10.
52. Chen VCH, Chan HL, Wu SI, Lee M, Lu ML, Liang HY, et al. Attention-deficit/hyperactivity disorder and mortality risk in Taiwan. JAMA Netw Open. 2019;2:e198714.
53. Damale MG, Pathan SK, Shinde DB, Patil RH, Arote RB, Sangshetti JN. Insights of tankyrases: a novel target for drug discovery. Eur J Med Chem. 2020;207:112712.
54. Riera CE, Merkwirth C, De Magalhaes Filho CD, Dillin A. Signaling networks determining life span. Annu Rev Biochem. 2016;85:35–64.
55. Wang Y, Zuo C, Xu Q, Hao L, Zhang Y. Attention-deficit/hyperactivity disorder is characterized by a delay in subcortical maturation. Prog Neuro-Psychopharmacology. Biol Psychiatry. 2021;104:110044.
56. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LSJ, et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. The Lancet. Psychiatry 2017;4:310–9.
57. Ceceli AO, Natsheh JY, Cruz D, Tricomi E. The neurobehavioral mechanisms of motivational control in attention-deficit/hyperactivity disorder. Cortex 2020;127:191–207.
58. Wachinger C, Nho K, Saykin AJ, Reuter M, Rieckmann A. A longitudinal imaging genetics study of neuroanatomical asymmetry in Alzheimer's Disease. Biol Psychiatry. 2018;84:522–30.
59. Song M, Liu J, Yang Y, Lv L, Li W, Luo XJ. Genome-wide meta-analysis identifies two novel risk loci for epilepsy. Front Neurosci. 2021;15:722592.
60. Sung YJ, Winkler TW, de las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. Am J Hum Genet. 2018;102:375–400.
61. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet. 2019;51:245–57.
62. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018;50:6.
63. Hill WD, Weiss A, Liewald DC, Davies G, Porteous DJ, Hayward C, et al. Genetic contributions to two special factors of neuroticism are associated with affluence, higher intelligence, better health, and longer life. Mol Psychiatry. 2020;25:3034.
64. Baselmans BML, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP, et al. Multivariate genome-wide analyses of the well-being spectrum. Nat Genet 2019 513. 2019;51:445–51.
65. Spann SJ, Ottinger MA. Longevity, Metabolic Disease, and Community Health. Prog Mol Biol Transl Sci. 2018;155:1–9.
66. Mackenbach JP, Karanikolos M, Looman CWN. The rise of mortality from mental and neurological diseases in Europe, 1979-2009: Observational study. BMC Public Health. 2014;14:840.
67. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry. 2015;72:334.

68. Lam M, Hill WD, Trampush JW, Yu J, Knowles E, Davies G, et al. Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. Am J Hum Genet. 2019;105:334–50.
69. Muntané G, Farré X, Bosch E, Martorell L, Navarro A, Vilella E. The shared genetic architecture of schizophrenia, bipolar disorder and lifespan. Hum Genet. 2021;140:441–55.
70. Olfson M, Gerhard T, Huang C, Crystal S, Stroup TS. Premature mortality among adults with schizophrenia in the United States. JAMA Psychiatry. 2015;72:1172–81.
71. Danler C, Pfaff K. The impact of an unequal distribution of education on inequalities in life expectancy. SSM Popul Heal. 2021;16:100954.
72. Halpern-Manners A, Raymo JM, Warren JR, Johnson KL. School performance and mortality: the mediating role of educational attainment and work and family trajectories across the life course. Adv Life Course Res. 2020;46:100362.
73. White DN, Stowell MHB. Room for two: the synaptophysin/synaptobrevin complex. Front Synaptic Neurosci. 2021;13:47.
74. Kwon SE, Chapman ER. Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons. Neuron. 2011;70:847–54.
75. Ceballos-Chávez M, Rivero S, García-Gutiérrez P, Rodríguez-Paredes M, García-Domínguez M, Bhattacharya S, et al. Control of neuronal differentiation by sumoylation of BRAF35, a subunit of the LSD1-CoREST histone demethylase complex. Proc Natl Acad Sci USA. 2012;109:8085–90.
76. Artegiani B, Labbaye C, Sferra A, Quaranta MT, Torreri P, Macchia G, et al. The interaction with HMG20a/b proteins suggests a potential role for β-dystrobrevin in neuronal differentiation. J Biol Chem. 2010;285:24740–50.
77. Zapolski TCB, Cyders MA, Smith GT. Positive urgency predicts illegal drug use and risky sexual behavior. Psychol Addict Behav. 2009;23:348–54.
78. Vågerö D, Aronsson V, Modin B. Why is parental lifespan linked to children's chances of reaching a high age? A transgenerational hypothesis. SSM - Popul Heal. 2018;4:45–54.
79. Kaplanis J, Gordon A, Shor T, Weissbrod O, Geiger D, Wahl M, et al. Quantitative analysis of population-scale family trees with millions of relatives. Science. 2018;360:171–5.
80. Darrous L, Mounier N, Kutalik Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. Nat Commun. 2021;12:7274.
81. Yuan Z, Liu L, Guo P, Yan R, Xue F, Zhou X. Likelihood-based Mendelian randomization analysis with automated instrument selection and horizontal pleiotropic modeling. Sci Adv. 2022;8:eabl5744.
82. Rovira P, Demontis D, Sánchez-Mora C, Zayats T, Klein M, Mota NR, et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology. 2020;45:1617–26.
83. McCarthy S, Cranswick N, Potts L, Taylor E, Wong ICK. Mortality associated with attention-deficit hyperactivity disorder (ADHD) drug treatment: A retrospective cohort study of children, adolescents and young adults using the general practice research database. Drug Saf. 2009;32:1089–96.
84. Chang Z, Lichtenstein P, D'Onofrio BM, Sjölander A, Larsson H. Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication a population-based study. JAMA Psychiatry. 2014;71:319–25.
85. Lichtenstein P, Halldner L, Zetterqvist J, Sjölander A, Serlachius E, Fazel S, et al. Medication for Attention Deficit–Hyperactivity Disorder and Criminality. N Engl J Med. 2012;367:2006–14.

AUTHOR CONTRIBUTIONS

GM, MR and MSA conceived the project, GM, MR and MSA participated in the study design, JCD, GM and LVR and participated in data acquisition, JCD, GM, MSA and LVR undertook the statistical analyses, JARQ, JCD, LM, GM, MR, MSA, LVR, EV, SSR and AAP. participated in the manuscript preparation. All authors contributed to the interpretation of the findings and revised and approved the final version of the manuscript.

FUNDING

This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-00444 and 2017SGR-1461), the Instituto de Salud Carlos III (PI18/01788, P19/01224, PI20/00041, PI21/00612, PI22/00464, FI18/00285 to LVR and CP22/00128 to MSA), the Ministry of Science, Innovation and Universities (IJC2018-035346-I to MSA and RYC2021-031324-I to JCD); the European Regional Development Fund (ERDF), the European Union H2020 Programme (H2020/2014–2020) under grant agreements no. 728018 (Eat2beNICE), no. 848228 (DISCOvERIE) and no. 2020604 (TIMESPAN) and the ECNP Network 'ADHD across the Lifespan' and "la Marató de TV3" (202228-30 and 202228-31). SSR was supported by funds from the California Tobacco-Related Disease Research Program (TRDRP; Grant Number T29KT0526 and T32IR5226), and NIH/NIDA DP1DA054394. AAP was supported by NIH grant P50DA037844 and R01AA029688 and Tobacco-Related Disease Research Program Grant number: 28IR-0070.

COMPETING INTERESTS

JARQ was on the speakers bureau and/or acted as consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Technofarma, Rubió and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry Mental Health chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious, and Rubió. AAP is on the scientific advisory board, owns stock options and a patent (https://patents.google.com/patent/US20160038559A1) for Vivid Genomics (https://www.vividgenomics.com/). All other authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41386-023-01555-x.

Correspondence and requests for materials should be addressed to Gerard Muntané or María Soler Artigas.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

International Journal of Epidemiology, 2022, 1–17
https://doi.org/10.1093/ije/dyac128
Original article

Original article

Mendelian randomization analysis for attention deficit/hyperactivity disorder: studying a broad range of exposures and outcomes

María Soler Artigas [1,2,3,4]* Cristina Sánchez-Mora,[1,2,3,4] Paula Rovira,[1,2,5,6] Laura Vilar-Ribó,[1,2,3] Josep Antoni Ramos-Quiroga[1,2,3,7] and Marta Ribasés[1,2,3,4]

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain, [2]Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain, [3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain, [4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain, [5]Vicerectorat de Recerca, postdoctoral researcher Margarita Salas, Universitat de Barcelona, Barcelona, Spain, [6]Departament of Psychiatry, Faculty of Medicine, Universidad de Granada, Spain and [7]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

*Corresponding author. Vall d'Hebron Research Institute (VHIR), Passeig Vall d'Hebron 119-129 08035 Barcelona, Spain. E-mail: maria.soler@vhir.org

Received 31 March 2021; Editorial decision 13 May 2022; Accepted 26 May 2022

Abstract

Background: Attention deficit/hyperactivity disorder (ADHD) is a highly prevalent neurodevelopmental disorder caused by a combination of genetic and environmental factors and is often thought as an entry point into a negative life trajectory, including risk for comorbid disorders, poor educational achievement or low income. In the present study, we aimed to clarify the causal relationship between ADHD and a comprehensive range of related traits.

Methods: We used genome-wide association study (GWAS) summary statistics for ADHD ($n = 53\,293$) and 124 traits related to anthropometry, cognitive function and intelligence, early life exposures, education and employment, lifestyle and environment, longevity, neurological, and psychiatric and mental health or personality and psychosocial factors available in the MR-Base database ($16\,067 \leq n \leq 766\,345$). To investigate their causal relationship with ADHD, we used two-sample Mendelian randomization (MR) with a range of sensitivity analyses, and validated MR findings using causal analysis using summary effect estimates (CAUSE), aiming to avoid potential false-positive results.

Results: Our findings strengthen previous evidence of a causal effect of ADHD liability on smoking and major depression, and are consistent with a causal effect on odds of decreased average total household income [odds ratio (OR) = 0.966, 95% credible interval (CrI) = (0.954, 0.979)] and increased lifetime number of sexual partners [OR = 1.023, 95%

CrI = (1.013, 1.033)]. We also found evidence for a causal effect on ADHD for liability of arm predicted mass and weight [OR = 1.452, 95% CrI = (1.307, 1.614) and OR = 1.430, 95% CrI = (1.326, 1.539), respectively] and time spent watching television [OR = 1.862, 95% CrI = (1.545, 2.246)], and evidence for a bidirectional effect for age of first sexual intercourse [beta = −0.058, 95% CrI = (−0.072, −0.044) and OR = 0.413, 95% CrI = (0.372, 0.457), respectively], odds of decreased age completed full-time education [OR = 0.972, 95% CrI = (0.962, 0.981) and OR = 0.435, 95% CrI = (0.356, 0.533), respectively] and years of schooling [beta = -0.036, 95% CrI = (−0.048, −0.024) and OR = 0.458, 95% CrI = (0.411, 0.511), respectively].

Conclusions: Our results may contribute to explain part of the widespread co-occurring traits and comorbid disorders across the lifespan of individuals with ADHD and may open new opportunities for developing preventive strategies for ADHD and for negative ADHD trajectories.

Key words: ADHD, Mendelian randomization, causal analysis using summary effect estimates

Key Messages

- Our results are consistent with a causal effect of attention deficit/hyperactivity disorder (ADHD) genetic liability decreasing average total household income and increasing lifetime number of sexual partners.
- We detect a positive effect of the liability of anthropometric traits (arm predicted mass and weight) and of time spent watching television on ADHD.
- We show evidence for a bidirectional negative effect between liability of ADHD and of education outcomes (years of schooling and age completed full-time education), age of first sexual intercourse and past tobacco smoking for non-heavy smokers.

Introduction

Attention deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder with a prevalence of around 5.3% in childhood and 2.8% in adulthood.[1,2] The aetiology of ADHD involves a combination of genetic and environmental factors, with an estimated heritability of ∼70–80%.[3,4] Potential environmental risk factors include pre- and perinatal risk factors (maternal smoking or alcohol consumption, low birthweight, prematurity), exposure to environmental toxins, unfavourable psychosocial conditions (severe early childhood deprivation, maternal hostility) or low socioeconomic status.[5,6]

ADHD is characterized by a persistent pattern of inattentive, hyperactive and impulsive behaviour; however, its clinical presentation is heterogeneous, with a wide spectrum of severity and symptoms that often overlap with other conditions.[7] There are a number of traits that are not part of the core diagnostic criteria for ADHD, which can nevertheless influence severity, persistence and treatment decisions. For instance, individuals with ADHD often have a poor cognitive performance in executive functions, such as response inhibition, vigilance, working memory or planning, personality profiles with low effortful control and high neuroticism,[8–11] or emotion dysregulation problems such as irritability or temper outbursts.[12] In addition, up to 70–80% of ADHD patients suffer from comorbid disorders across their lifespan.[13] These include other psychiatric conditions, such as major depressive, oppositional defiant, bipolar or substance use disorders, but also somatic diseases such as obesity, sleep disorders or migraine.[7] The presence of comorbidities in ADHD worsens symptom progression, disorder course and outcome, and also increases mortality rates.[8,14] In this context ADHD can be thought of as an entry point into a negative life trajectory with higher risks for poor educational achievement, unemployment or criminality, among others.[7]

Most studies undertaken to date have reported association between ADHD and comorbid traits, but inferring causality can be more challenging due to the potential effect of confounding factors or reverse causality. Different strategies have been developed to overcome these inference problems, and the causality for ADHD and a number of

traits have been tested using: (i) longitudinal analyses, some of them undertaken in twins; (ii) Mendelian randomization (MR), which uses genetic variants as proxies for an exposure (instrumental variables) to test for a causal effect on an outcome[15]; and (iii) the latent causal variable (LCV) model, based on a latent variable that mediates the genetic correlation between two traits and quantifies the degree of causality between them.[16] When using only one approach, longitudinal analyses have reported a causal role for low family income in early childhood on ADHD[6] and of ADHD on lower educational achievement.[17] In addition, MR studies have reported an effect of the liability of low birthweight increasing the risk for ADHD,[18] an effect of higher intelligence lowering the risk for ADHD[19] and an effect of the genetic liability to ADHD increasing the risk for asthma[20] and coronary artery disease, as well as a positive bidirectional effect for childhood obesity.[21] When more than one approach was used, consistent results were found for an effect for ADHD liability on major depression[22] and inconsistent results, which require further investigation, were identified for body mass index (BMI),[23–25] phone use,[26] smoking, cannabis and alcohol use.[27–30]

Overall, the evidence from causal inference analyses undertaken for ADHD is hard to interpret, in some cases due to the limited number of strategies applied[19,20] or because inconsistent findings are found when using different approaches.[29,30] In the present study we aim to clarify the relationship between ADHD and a comprehensive range of related traits, using genome-wide association studies (GWAS) datasets available,[31,32] following current guidelines on two-sample MR analyses[33] and a range of sensitivity analyses. In addition, we validated MR findings using causal analysis using summary effect estimates (CAUSE), a recently developed method to account for horizontal pleiotropy that acts through a common shared heritable factor, and avoids potential false-positive results.[34]

Methods

GWAS dataset selection

We selected GWAS summary statistics available in the MR-Base database[31,32] in May 2020 ($n = 31\,772$) which fulfilled the following inclusion criteria: (i) sample size [N effective = 4 × N cases × N controls/(N cases + N controls) for binary traits] >5000; (ii) results available in more than 450 000 genetic variants; (iii) European ancestry; (iv) non sex-specific; and (v) more than three independent genome-wide significant signals ($P < 5.00e-08$). This reduced the number of traits to 1259 (Supplementary Table S1, available as Supplementary data at *IJE* online). At this point we removed traits related to diet ($n = 52$), to metabolites' levels ($n = 132$), to procedural metrics or biological samples ($n = 76$) and to clinical traits ($n = 600$), except those related to neurological, psychiatric and mental health; and selected traits related to anthropometry, cognitive function and intelligence, early life exposures, education and employment, lifestyle and environment, longevity, neurological, psychiatric and mental health or personality and psychosocial factors. We then removed duplicated and unordered categorical traits ($n = 150$, Supplementary Table S1). Finally, in order to reduce further the number of traits analysed to those with causal relationships most likely to be identified by MR methods, we removed traits with a heritability Z score ≤4 and an ADHD genetic correlation Z score ≤2, obtained using LD score regression,[35,36] ending up with a total of 124 traits included in the analysis (Supplementary Tables S1 and S2, available as Supplementary data at *IJE* online). The MR-Base summary statistics used for these 124 traits were obtained from different sources: European Bioinformatics Institute (EBI) database,[37] Neale Lab and Integrative Epidemiology Unit (IEU) analyses of UK Biobank phenotypes [http://www.nealelab.is/uk-biobank][38] and manually collected and curated data from different consortia for MR-Base.[19,39–47]

Mendelian randomization

Main analysis

We ran two-sample MR in both directions to assess the potential causal relationship between ADHD and every selected trait, using the multiplicative random effects inverse variance weighted (IVW) method as the main analysis.[48] MR analyses only included independent single nucleotide polymorphisms (SNPs) ($r^2 < 0.001$ or distance >10 000 kb) with $P < 5.00e-08$ in the exposure, and variants meeting the threshold for both the exposure and the outcome were removed. For exposure variants not found in the outcome, GWAS proxies were used instead ($r^2 \geq 0.8$, obtained using 1000 Genomes European sample). ADHD liability genetic instruments are provided in Supplementary Table S3 (available as Supplementary data at *IJE* online). False-discovery rate (FDR) across all tests considered was used to correct for multiple testing.

MR sensitivity analyses

For IVW results with FDR corrected $P < 0.05$, we undertook sensitivity analyses to assess the robustness of the findings under weaker assumptions, given that the genetic variants used as instruments must meet three assumptions for IVW results to be valid: (i) robust association with the exposure; (ii) no horizontal pleiotropy, or association with the outcome through a pathway independent of the

exposure; and (iii) independence of confounders that influence the exposure and the outcome.

We used weighted median and weighed mode methods, which only require a subset of variants to be valid instruments and are robust to outliers.[49,50] Under equal weights, the weighted median requires at least half of the variants to be valid instruments, and the weighted mode requires more variants to estimate the true causal effect than any other quantity.[49,50] We used MR-PRESSO, which performs a test to detect horizontal pleiotropy (MR-PRESSO global test), and if detected, it removes horizontal pleiotropic outliers and then performs the IVW method using the remaining instruments.[51] We also applied MR-Egger, which is affected by outlying data points but allows all genetic variants to have pleiotropic effects, assuming that these effects are independent of the variant-exposure associations.[52,53] MR-Egger also implements a pleiotropy test, however, when the assumption of no measurement error in the SNP-exposure effect estimates (NOME assumption) is violated (I^2_{GX} <0.9). MR-Egger causal estimates are biased towards the null, and the pleiotropy test type I error can be inflated.[54] For this reason, when 0.6 <I^2_{GX}< 0.9 we implemented the method of simulation extrapolation (SIMEX) to obtain an adjusted estimate of the causal effect.[54] When I^2_{GX} <0.6, we disregarded MR-Egger results. We also calculated F statistics for continuous exposures to measure the strength of the instruments used; as a 'rule of thumb', F statistics >10 indicate strong-enough instruments.[55] In order to avoid results due to reverse causation, we also applied Steiger filtering, removing instruments from the analysis if they did not explain substantially more of the variance in the exposure trait than in the outcome and undertaking the IVW method on the remaining set of instruments.[56] Variance explained for binary traits was estimated using Equation 10 from Lee *et al.*,[57] as implemented by get_r_from_lor, R function within the TwoSampleMR package. In addition, we ran heterogeneity tests and leave-one-out analyses and generated scatter, funnel and forest plots. TwoSampleMR v0.5.5 and MRPRESSO R packages were used for these analyses.[31,51]

We considered that there was evidence of a causal relationship if: (i) IVW FDR P <0.05; (ii) the same direction of effect as the IVW beta estimate and P <0.05 was found for the weighted median and mode, MR PRESSO, IVW after Steiger filtering and when there was also evidence of pleiotropy (MR-Egger intercept P <0.05), MR-Egger or SIMEX if I^2_{GX} >0.9 or 0.6<I^2_{GX}<=0.9, respectively; and (iii) F statistic >10 for continuous exposures.

Causal analysis using summary effect estimates

For those traits that met the MR sensitivity criteria, we also ran causal analysis using summary effect estimates (CAUSE).[34] Uncorrelated horizontal pleiotropy occurs when the effects on the exposure and the outcome are uncorrelated and it can be accounted for by methods such as MR-Egger or MR-PRESSO. Correlated horizontal pleiotropy takes place when the effects on the exposure and the outcome act through a shared heritable factor, and is harder to account for by current MR methods. CAUSE can deal with both kinds of horizontal pleiotropy, avoiding therefore potential false-positive findings. This method uses expected log pointwise posterior density (ELPD), a Bayesian approach, to compare two nested models: a sharing model where the causal effect (γ) is zero but allows for horizontal pleiotropic effects; and a causal model where γ is a free parameter.[34] Independent variants (r^2 <0.01; distance >250 kb) associated to the exposure with P <1.00e-03 were included, and the cause v1.2.0 R package was used for these analyses.[34] A Bonferroni correction for the number of tests undertaken was used to correct for multiple testing at this stage (P <1.21e-03).

Results

Anthropometric measures

The IVW analyses showed findings with FDR P <0.05 for one anthropometric trait when ADHD was the exposure, 17 when ADHD was considered as the outcome and seven in both directions (Figure 1; Supplementary Table S4, available as Supplementary data at *IJE* online). After sensitivity analyses there was evidence of a positive effect of the genetic liability of eight traits (arm, leg, whole body and trunk fat-free mass, arm and trunk predicted mass, whole body water mass and weight) on ADHD (Table 1), showing moderate heterogeneity (40.128%<I^2<46.114%, Supplementary Table S5a, Figure S1, available as Supplementary data at *IJE* online). Also, evidence meeting the multiple comparison correction was found with CAUSE for a causal effect of arm predicted mass and weight on ADHD [odds ratio (OR)= 1.452, 95% CrI = (1.307, 1.614), ΔELPD P = 5.32e-04 and OR = 1.430, 95% credible interval (CrI) = (1.326, 1.539), ΔELPD P = 1.01e-06, respectively], with also suggestive evidence (ΔELPD P <0.05) of a positive effect for the remaining traits (Figure 2, Table 2).

Cognitive function and intelligence

In the IVW analysis, a negative effect of ADHD genetic liability on mean time to correctly identify matches, a measure of raw processing and reaction speed, was detected as well as a negative effect of intelligence when ADHD was the outcome, and for cognitive performance and fluid

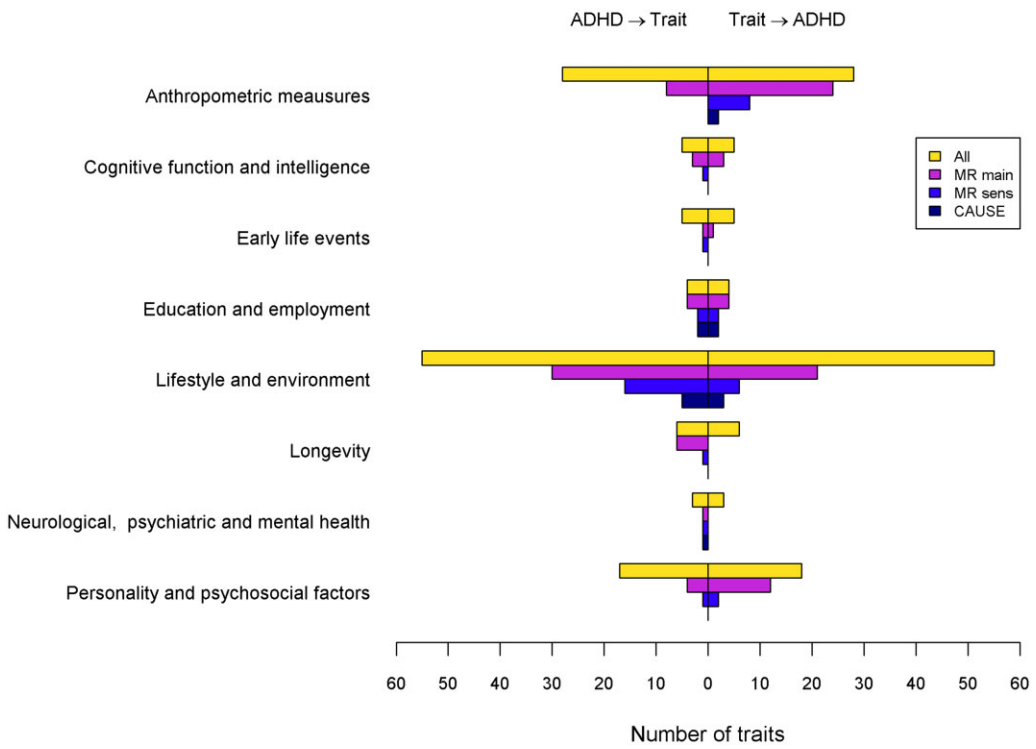


Figure 1 Number of traits included in each stage of the analysis for each direction and category. Total number of traits analysed ('All') are presented as well as number of traits meeting the significance criteria in the MR main analysis ('MR main'), in the MR sensitivity analyses ('MR sens') and in CAUSE ('CAUSE'). ADHD, attention deficit/hyperactivity disorder; CAUSE, causal analysis using summary effect estimates; MR, Mendelian randomization

intelligence in both directions (IVW FDR $P<0.05$) (Table 1; Supplementary Table S4; Figure 1). Of them, the effect of ADHD as exposure on cognitive performance survived the sensitivity analyses, but showed high heterogeneity (I^2 = 84.231%, Supplementary Table S5b, Supplementary Figure S1a–h). When outliers were removed in the MR-PRESSO analysis, the magnitude of the causal effect estimate increased slightly (Supplementary Table S5b,) and the heterogeneity went down to an I^2 of 68.844%. The CAUSE analysis did not show evidence for the effect of ADHD liability on cognitive performance (ΔELPD $P=0.069$, Table 2).

Early life events

Out of all early life exposure traits, a positive effect between maternal smoking around birth and ADHD was identified in both directions in the IVW analysis; but only when ADHD was considered as exposure were the sensitivity criteria met (Figure 1; Supplementary Table S4). There was moderately high heterogeneity for this analysis (I^2 = 63.896%), and although removing one outlier in the MR-PRESSO analysis reduced the heterogeneity and the causal effect estimate, measured as odds of maternal smoking per unit increase in log odds ratio [log(OR)] of ADHD, it remained significant [from OR = 1.037, 95% confidence interval (CI) = (1.023, 1.052) to OR = 1.03, 95% CI = (1.018, 1.043), I^2 = 36.476%, Supplementary Table S5c, Supplementary Figure S1am]. The effect of ADHD on maternal smoking around birth, however, did not survive multiple comparisons correction with CAUSE (ΔELPD $P=2.19e\text{-}03$, Table 2).

Education and employment

The IVW analysis detected results with FDR $P<0.05$ in both directions for all traits analysed, with a negative effect for age completed full-time education and years of schooling and a positive effect for job involving heavy manual, physical work or mainly walking or standing (Figure 1, Table 1; Supplementary Table S4). Only age completed full-time education and years of schooling met the sensitivity analysis criteria in both directions. However, when ADHD was the outcome, the Steiger analysis provided evidence for reverse causation,with reduced effect sizes after filtering, which suggests that at least part of the observed effect was due to ADHD liability causing the educational outcomes (Supplementary Table S5d). The heterogeneity for these analyses was moderate (I^2 <50%) except for the effect of ADHD liability on years of schooling, which was

Table 1 Mendelian randomization results

Trait[a]	ADHD → Trait					Trait → ADHD				
	Number of SNPs	Effect size	95% CI	FDR *P*	Meets sensitivity?	Number of SNPs	OR	95% CI	FDR *P*	Meets sensitivity?
Anthropometric measures										
Arm fat-free mass (left)	9	0.026	(0.005, 0.046)	3.09E-02	No	479	1.671	(1.449, 1.927)	2.00E-11	Yes
Arm predicted mass (left)	9	0.025	(0.004, 0.046)	4.38E-02	No	471	1.665	(1.438, 1.929)	1.11E-10	Yes
Whole body fat-free mass	9	0.016	(−0.006, 0.038)	2.31E-01	–	509	1.431	(1.251, 1.637)	1.13E-06	Yes
Whole body water mass	9	0.016	(−0.006, 0.038)	2.25E-01	–	518	1.456	(1.27, 1.669)	5.24E-07	Yes
Leg fat-free mass (left)	9	0.012	(−0.016, 0.04)	4.88E-01	–	460	1.461	(1.262, 1.69)	2.12E-06	Yes
Trunk fat-free mass	9	0.017	(−0.003, 0.037)	1.47E-01	–	520	1.397	(1.218, 1.602)	8.25E-06	Yes
Trunk predicted mass	9	0.017	(−0.003, 0.037)	1.58E-01	–	518	1.406	(1.225, 1.613)	6.09E-06	Yes
Weight	9	0.045	(0.002, 0.088)	7.67E-02	–	458	1.662	(1.485, 1.859)	1.54E-17	Yes
Cognitive function and intelligence										
Cognitive performance	9	−0.072	(−0.131, −0.013)	3.55E-02	Yes	134	0.600	(0.499, 0.722)	5.24E-07	No
Early life events										
Maternal smoking around birth[b]	9	1.037	(1.023, 1.052)	1.89E-06	Yes	15	143.936	(25.34, 817.581)	1.70E-07	No
Education and employment										
Age completed full-time education[b]	9	0.930	(0.909, 0.951)	2.52E-09	Yes	36	0.167	(0.102, 0.275)	2.00E-11	Yes
Years of schooling	9	−0.098	(−0.138, −0.059)	5.99E-06	Yes	297	0.298	(0.251, 0.354)	1.20E-41	Yes
Lifestyle and environment										
Alcohol intake frequency[b]	9	1.108	(1.049, 1.171)	8.51E-04	Yes	92	1.399	(1.168, 1.676)	8.51E-04	No
Alcohol intake versus 10 years previously[b]	9	1.042	(1.026, 1.058)	1.13E-06	Yes	12	8.389	(2.293, 30.696)	3.62E-03	Yes
Alcohol usually taken with meals[b]	9	0.959	(0.942, 0.976)	2.08E-05	Yes	33	0.085	(0.033, 0.221)	2.19E-06	No
Average weekly champagne plus white wine intake[b]	9	0.955	(0.936, 0.975)	4.74E-05	Yes	4	0.148	(0.041, 0.528)	7.92E-03	No
Frequency of stair-climbing in past 4 weeks[b]	9	0.954	(0.918, 0.991)	3.21E-02	Yes	17	0.383	(0.204, 0.717)	7.00E-03	No
Nitrogen dioxide air pollution 2010	9	0.032	(0.016, 0.049)	5.37E-04	Yes	8	0.610	(0.159, 2.339)	5.43E-01	–
Particulate matter air pollution ($pm_{2.5}$) 2010	9	0.047	(0.029, 0.064)	1.55E-06	Yes	7	1.989	(0.38, 10.401)	5.01E-01	–
Age first had sexual intercourse	9	−0.130	(-0.18, -0.08)	1.74E-06	Yes	184	0.223	(0.187, 0.266)	8.34E-61	Yes
Lifetime number of sexual partners[b]	9	1.063	(1.029, 1.099)	8.61E-04	Yes	60	2.958	(1.76, 4.971)	1.63E-04	No
Current tobacco smoking[b]	9	1.041	(1.028, 1.055)	4.94E-09	Yes	32	12.305	(3.945, 38.38)	6.54E-05	No
Ever smoked[b]	9	1.032	(1.022, 1.043)	3.64E-09	Yes	73	15.251	(7.528, 30.898)	7.20E-13	No

(Continued)

Table 1 Continued

Trait[a]	ADHD → Trait					Trait → ADHD				
	Number of SNPs	Effect size	95% CI	FDR *P*	Meets sensitivity?	Number of SNPs	OR	95% CI	FDR *P*	Meets sensitivity?
Pack/years adult smoking as proportion of lifespan exposed to smoking	9	0.075	(0.039, 0.111)	1.73E-04	Yes	13	1.335	(0.971, 1.835)	1.28E-01	–
Past tobacco smoking[b]	9	0.905	(0.88, 0.931)	4.14E-11	Yes	87	0.381	(0.296, 0.491)	1.02E-12	Yes
Average total household income before tax[b]	9	0.909	(0.885, 0.933)	2.65E-11	Yes	44	0.329	(0.237, 0.455)	2.37E-10	No
Number of full sisters[b]	9	1.029	(1.016, 1.043)	8.78E-05	Yes	5	9.746	(0.4, 237.152)	2.38E-01	–
Townsend deprivation index at recruitment	9	0.075	(0.055, 0.095)	1.35E-12	Yes	17	5.400	(1.952, 14.941)	3.25E-03	Yes
Length of mobile phone use[b]	9	1.013	(0.975, 1.052)	5.87E-01	–	29	1.983	(1.314, 2.993)	3.14E-03	Yes
Time spent watching television (TV)[b]	9	1.030	(0.997, 1.065)	1.34E-01	–	107	3.495	(2.521, 4.847)	1.00E-12	Yes
Longevity										
Parental longevity (combined parental attained age, Martingale residuals)[c]	9	0.050	(0.029, 0.071)	1.07E-05	Yes	10	0.972	(0.462, 2.041)	9.62E-01	–
Neurological, psychiatric and mental health										
Major depressive disorder[b]	10	1.200	(1.122, 1.283)	7.12E-07	Yes	5	1.739	(1.048, 2.886)	6.25E-02	–
Personality and psychosocial factors										
Able to confide[b]	9	0.919	(0.881, 0.959)	3.51E-04	No	13	0.635	(0.46, 0.878)	1.39E-02	Yes
Frequency of tiredness/lethargy in past 2 weeks[b]	9	1.029	(1.004, 1.055)	5.16E-02	–	36	5.220	(2.809, 9.698)	1.13E-06	Yes
Frequency of unenthusiasm/disinterest in past 2 weeks[b]	9	1.034	(1.021, 1.047)	1.13E-06	yes	11	8.077	(1.753, 37.221)	1.69E-02	No

For all the traits with inverse variance weighted FDR $P < 0.05$ which also met the sensitivity criteria in at least one direction, the number of SNPs included, the inverse variance weighted results and information on whether the MR sensitivity criteria were met are presented for both directions.

MR, Mendelian randomization; ADHD, attention deficit/hyperactivity disorder; SNP, single nucleotide polymorphism; CI, 95% confidence interval; FDR, false-discovery rate; *P*, *P*-value.

[a]The unit reported for all continuous traits is SD, except for years of schooling, which was provided in years and cognitive performance, which was provided using a cognitive score.[44]

[b]The causal effect size provided for the comparison with ADHD as the exposure for these traits is odds raio (OR) per ADHD log(OR) unit increase, since they were all analysed as categorical ordered, except for maternal smoking around birth, ever smoked, alcohol usually taken with meals and major depressive disorder, which were analysed as binary.

[c]For parental longevity (combined parental attained age, Martingale residuals) a positive effect size indicates decreased attained age.[47]

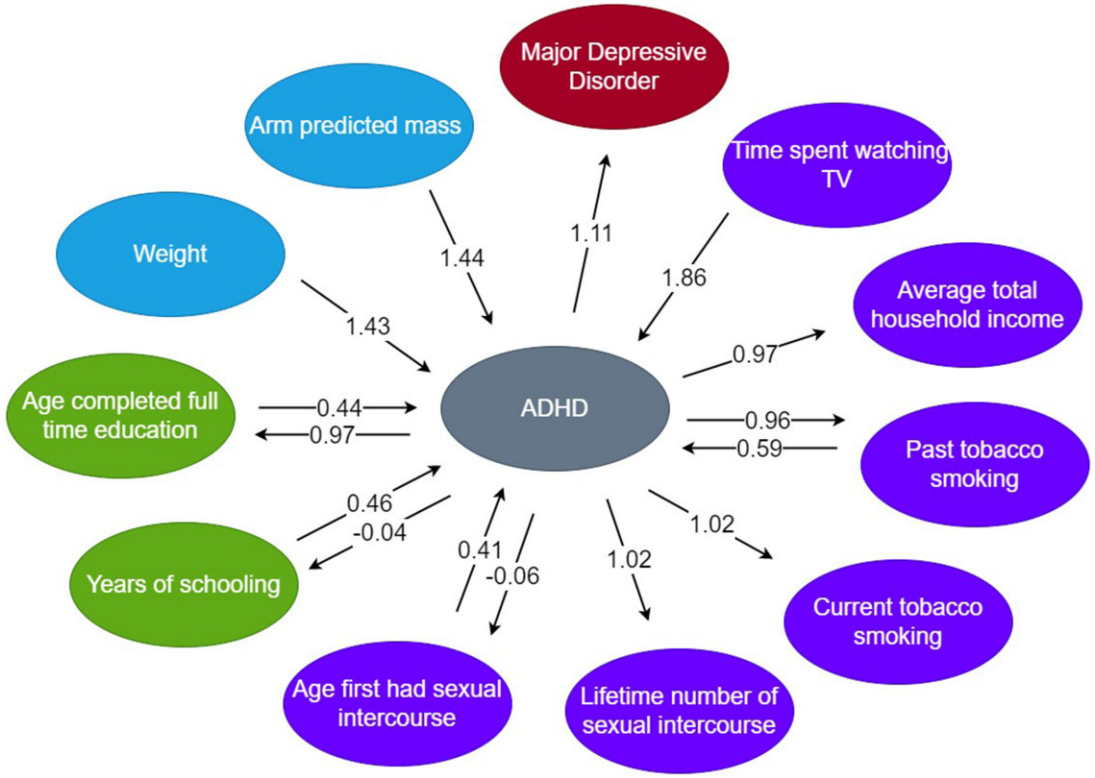


Figure 2 Diagram of the relationships found with consistent evidence across methods. Traits are coloured by category (education and employment in green; anthropometric measures in blue; neurological, psychiatric and mental health-related traits in red; and lifestyle and environment in purple) and CAUSE effect sizes are presented. Details on effect size units can be found in Table 2. ADHD, attention deficit/hyperactivity disorder; CAUSE, causal analysis using summary effect estimates; TV, television

considerable (I^2 = 87.572%, Supplementary Table S5d). After excluding outliers, MR-PRESSO results remained significant (Supplementary Table S5d, Supplementary Figure S1au) and the heterogeneity decreased, although it remained substantial (I^2 = 67.755%). CAUSE confirmed these results, providing strong evidence for a causal effect per unit increase in the log(OR) of ADHD on odds of decreased age completed full-time education and years of schooling [OR = 0.972, 95% CrI = (0.962, 0.981), ΔELPD P = 2.85e-04 and beta = -0.036, 95% CrI = (-0.048, -0.024), ΔELPD P = 8.04e-05, respectively] (Figure 2, Table 2). Evidence in the opposite direction was also found, with causal effect per unit increase in odds of increased age completed full-time education and years of schooling on ADHD odds [OR = 0.435, 95% CrI = (0.356, 0.533), ΔELPD P = 2.51e-06 and OR = 0.458, 95% CrI = (0.411, 0.511), ΔELPD P = 1.08e-010, respectively] (Figure 2, Table 2).

Lifestyle and environment

In the IVW analysis 12 lifestyle and environment traits had FDR P <0.05 when ADHD was the exposure, three when ADHD was the outcome and 18 in both directions (Figure 1; Supplementary Table S4). After the sensitivity analyses, we found evidence for an effect of ADHD as exposure on traits related to alcohol use (increased alcohol intake frequency, alcohol intake versus 10 years previously, decreased alcohol usually taken with meals and average weekly champagne plus white wine intake), physical exercise (decreased frequency of stair-climbing in past 4 weeks), air pollution (increased nitrogen dioxide and particulate matter), sexual behaviour (decreased age first had sexual intercourse and increased lifetime number of sexual partners), smoking (increased current tobacco smoking, ever smoking and pack/years adult smoking, and decreased past tobacco smoking) and sociodemographic information (decreased average total household income and increased Townsend deprivation index and number of sisters) (Table 1). Alcohol intake versus 10 years previously, age of first sexual intercourse, past tobacco smoking and Townsend deprivation index also had an effect when ADHD was used as outcome, as did length of mobile phone use and time spent watching television (Table 1).

When ADHD was considered as outcome, all analyses that met the sensitivity criteria also showed evidence of reverse causation, with smaller effect sizes after Steiger filtering particularly for Townsend deprivation index, suggesting that at least some the observed effect was due to

Table 2 Causal analysis using summary effect estimates results

Trait[a]	ADHD → Trait				Trait → ADHD			
	γ	95% CrI	ΔELPD (SE)	ΔELPD *P*	γ	95% CrI	ΔELPD (SE)	ΔELPD *P*
Anthropometric measures								
Arm fat-free mass (left)	–	–	–	–	1.441	(1.287,1.606)	−4.444 (1.555)	2.13E-03
Arm predicted mass (left)	–	–	–	–	1.452	(1.307,1.614)	−4.57 (1.396)	**5.32E-04**
Whole body fat-free mass	–	–	–	–	1.312	(1.175,1.464)	−3.275 (1.405)	9.87E-03
Whole body water mass	–	–	–	–	1.312	(1.176,1.464)	−3.14 (1.346)	9.83E-03
Leg fat-free mass (left)	–	–	–	–	1.371	(1.234,1.523)	−3.79 (1.4)	3.39E-03
Trunk fat-free mass	–	–	–	–	1.232	(1.1,1.381)	−2.51 (1.44)	4.06E-02
Trunk predicted mass	–	–	–	–	1.244	(1.108,1.398)	−2.505 (1.406)	3.74E-02
Weight	–	–	–	–	1.430	(1.326,1.539)	−5.71 (1.202)	**1.01E-06**
Cognitive function and intelligence								
Cognitive performance	−0.023	(−0.04,−0.006)	−2.681 (1.81)	6.92E-02	–	–	–	–
Early life events								
Maternal smoking around birth[b]	1.012	(1.007,1.017)	−4.886 (1.715)	2.19E-03	–	–	–	–
Education and employment								
Age completed full time education[b]	0.972	(0.962,0.981)	−5.655 (1.641)	**2.85E-04**	0.435	(0.356,0.533)	−6.159 (1.349)	**2.51E-06**
Years of schooling	−0.036	(−0.048,-0.024)	−6.5 (1.723)	**8.04E-05**	0.458	(0.411,0.511)	−7.184 (1.131)	**1.08E-10**
Lifestyle and environment								
Alcohol intake frequency[b]	1.033	(1.017,1.049)	−5.103 (2.079)	7.04E-03	–	–	–	–
Alcohol intake versus 10 years previously[b]	1.010	(1.003,1.016)	−2.66 (1.813)	7.12E-02	2.099	(1.294,3.384)	−3.324 (1.814)	3.34E-02
Alcohol usually taken with meals[b]	0.986	(0.98,0.993)	−4.543 (1.8)	5.80E-03	–	–	–	–
Average weekly champagne plus white wine intake[b]	0.985	(0.977,0.994)	−3.654 (1.905)	2.75E-02	–	–	–	–
Frequency of stair climbing in last 4 weeks[b]	0.991	(0.979,1.002)	−0.55 (1.288)	3.35E-01	–	–	–	–
Nitrogen dioxide air pollution 2010	0.003	(-0.005,0.012)	0.653 (0.47)	9.18E-01	–	–	–	–
Particulate matter air pollution ($pm_{2.5}$) 2010	0.008	(-0.001,0.016)	−0.936 (1.484)	2.64E-01	–	–	–	–
Age first had sexual intercourse	−0.058	(-0.072,−0.044)	−6.939 (1.562)	**4.47E-06**	0.413	(0.372,0.457)	−7.797 (1.105)	**8.70E-13**
Lifetime number of sexual partners[b]	1.023	(1.013,1.033)	−5.869 (1.911)	**1.06E-03**	–	–	–	–
Current tobacco smoking[b]	1.015	(1.01,1.021)	−5.488 (1.703)	**6.35E-04**	–	–	–	–
Ever smoked[b]	1.011	(1.006,1.016)	−5.368 (1.974)	3.27E-03	–	–	–	–

(Continued)

Table 2 Continued

Trait[a]	ADHD → Trait				Trait → ADHD			
	γ	95% CrI	ΔELPD (SE)	ΔELPD *P*	γ	95% CrI	ΔELPD (SE)	ΔELPD *P*
Pack years adult smoking as proportion of life span exposed to smoking	0.034	(0.019,0.047)	−4.232 (1.751)	7.81E-03	–	–	–	–
Past tobacco smoking[b]	0.963	(0.947,0.978)	−5.747 (1.829)	**8.37E-04**	0.588	(0.523,0.664)	−5.685 (1.399)	**2.41E-05**
Average total household income before tax[b]	0.966	(0.954,0.979)	−5.777 (1.809)	**7.01E-04**	–	–	–	–
Number of full sisters[b]	1.007	(0.999,1.013)	−1.207 (1.546)	2.17E-01	–	–	–	–
Townsend deprivation index at recruitment	0.024	(0.015,0.033)	−4.996 (1.701)	1.65E-03	1.870	(1.38,2.522)	−3.676 (1.546)	8.72E-03
Length of mobile phone use[b]	–	–	–	–	1.281	(1.084,1.516)	−2.874 (1.805)	5.56E-02
Time spent watching television (TV)[b]	–	–	–	–	1.862	(1.545,2.246)	−4.491 (1.475)	**1.16E-03**
Longevity								
Parental longevity (combined parental attained age, Martingale residuals)[c]	0.017	(0.007,0.026)	−3.663 (1.873)	2.52E-02	–	–	–	–
Neurological, psychiatric and mental health								
Major depressive disorder[b]	1.110	(1.073,1.148)	−5.828 (1.666)	**2.34e-04**	–	–	–	–
Personality and psychosocial factors								
Able to confide[b]	–	–	–	–	0.864	(0.739,1.012)	−0.836 (1.416)	2.77E-01
Frequency of tiredness/lethargy in past 2 weeks[b]	–	–	–	–	1.772	(1.388,2.25)	−3.717 (1.523)	7.34E-03
Frequency of unenthusiasm/disinterest in past 2 weeks[b]	1.013	(1.007,1.019)	−5.665 (1.982)	2.13E-03	–	–	–	–

For all analyses that met the MR sensitivity criteria, CAUSE results are presented, including an estimate of the causal effect size (γ) with 95% credible intervals and the difference between the ELPD in the causal and in the sharing models (**ΔELPD**) with its standard error and one-sided *P*-value. *P*-values meeting a Bonferroni corrected threshold are highlighted in bold (0.05/41=1.21E-03).

MR, Mendelian randomization; CAUSE, causal analysis using summary effect estimates; ADHD, attention deficit/hyperactivity disorder; Crl, credible interval; ELPD, expected log pointwise posterior density; SE, standard error; *P*, *P*-value.

[a]The unit reported for all continuous traits is SD, except for years of schooling, which was provided in years and cognitive performance, which was provided using a cognitive score.[44]

[b]The causal effect size (γ) provided for these traits is odds ratio (OR) per ADHD log(OR) unit increase when ADHD is the exposure, since they were all analysed as categorical ordered, except for maternal smoking around birth, ever smoked, alcohol usually taken with meals and major depressive disorder, which were analysed as binary.

[c]For parental longevity (combined parental attained age, Martingale residuals) a positive effect size indicates decreased attained age.[47]

ADHD liability causing the lifestyle and environmental outcomes (Supplementary Table S5e). Also, high heterogeneity (I^2 >70%) was detected for alcohol intake frequency, age of first sexual intercourse and lifetime number of sexual partners when ADHD was the exposure and for Townsend deprivation index in the other direction (Supplementary Figure S1s, af, ah, au, Supplementary Table S5e). After removing outliers there was still evidence of a causal effect and the heterogeneity decreased, although it remained around 60% for age of first sexual intercourse, lifetime number of sexual partners and Townsend deprivation index (Supplementary Table S5e).

After multiple comparisons correction, CAUSE analyses provided evidence with ADHD as the exposure for lower odds of increased average total household income [OR = 0.966, 95% CrI = (0.954, 0.979)], higher odds of a increased lifetime number of sexual partners [OR = 1.023, 95% CrI = (1.013, 1.033)] and a negative effect in both directions for age first had sexual intercourse [beta = −0.058, 95% CrI = (−0.072, −0.044) and OR = 0.413, 95% CrI = (0.372, 0.457) for ADHD as exposure and outcome, respectively] (Figure 2, Table 2). Evidence was also found for a positive effect of ADHD liability on current tobacco smoking [OR = 1.015, 95% CrI = (1.01, 1.021)], and a negative effect for past tobacco smoking in both directions [OR = 0.963, 95% CrI = (0.947, 0.978), OR = 0.588, 95% CrI = (0.523, 0.664) for ADHD as exposure and outcome, respectively] (Figure 2, Table 2). In addition, time spent watching television seemed to increase significantly the odds of ADHD [OR = 1.862, 95% CrI = (1.545, 2.246)] (Figure 2, Table 2). Suggestive evidence (ΔELPD P <0.05) was shown in the CAUSE analysis for a causal effect of the genetic liability of ADHD on the remaining smoking-related traits and on all alcohol-related traits, except for alcohol intake versus 10 years previously, which showed an effect in the opposite direction, and for Townsend deprivation index, which showed an effect in both directions (Table 2).

Longevity

In the IVW analysis, there was no evidence of a causal effect when ADHD was considered as outcome, but evidence for an effect of ADHD as exposure was found in all longevity-related traits, decreasing maternal, paternal and combined age of death or attained age and decreasing the odds of both parents being in the top 10% of survival (Figure 1, Table 1; Supplementary Table S4). However, only effects on combined parental attained age met the sensitivity analysis criteria and showed suggestive evidence of a causal relationship in the CAUSE analysis, although it did not meet multiple testing correction (ΔELPD P = 0.025, Table 2).

Neurological, psychiatric and mental health-related traits

We found evidence of an increase in the odds of major depression per unit increase in the log(OR) of ADHD in the IVW analysis [OR = 1.200, 95% CI=(1.122, 1.283), Table 1], which survived the sensitivity analyses and also met the Bonferroni correction for multiple testing in CAUSE [OR = 1.110, 95% CrI=(1.073, 1.148), ΔELPD P = 2.34e-04] (Figure 2, Table 2). An alternative and more broad definition of depression, however, did not reach statistical significance at any stage, and neither did the other analyses undertaken (Figure 1; Supplementary Table S4).

Personality and psychosocial factors

In the IVW analysis, we found evidence of an effect for ADHD as exposure on one trait, with ADHD as outcome for nine traits and in both directions for three traits (Supplementary Table S4; Figure 1). The effect of the genetic liability of ADHD on frequency of unenthusiasm/disinterest in past 2 weeks met the sensibility analysis criteria, but was only suggestive after multiple comparison correction in the CAUSE analysis (ΔELPD P = 2.13e-03, Table 2) and evidence of reverse causation was found in the opposite direction (Supplementary Table S5h). The effect of the genetic liability of frequency of tiredness/lethargy in past 2 weeks and of being able to confide on ADHD remained after the sensitivity analyses, but only suggestive evidence was found for being able to confide in the CAUSE analyses (ΔELPD P = 7.34e-03, Table 2).

Overall, the direction of the causal effect size estimates was consistent between IVW MR and CAUSE for all analyses undertaken and, in general, CAUSE effect size estimates were smaller than those estimated by IVW MR (Figure 3, Tables 1 and 2).

Discussion

In the present study we assessed for the first time the causal relationships between ADHD and a broad range of ADHD-related traits, applying complementary approaches. Through this comprehensive strategy we found consistent evidence across methods for a causal effect of the genetic liability of anthropometric measures and time spent watching television on ADHD, for the genetic liability of ADHD on average household income and major depressive disorder, and a bidirectional relationship with educational achievement, smoking and sexual behaviour.

Our findings give support to the relationship between ADHD and risk-taking behaviours and to existing evidence indicating that ADHD is an entry point into a harmful life

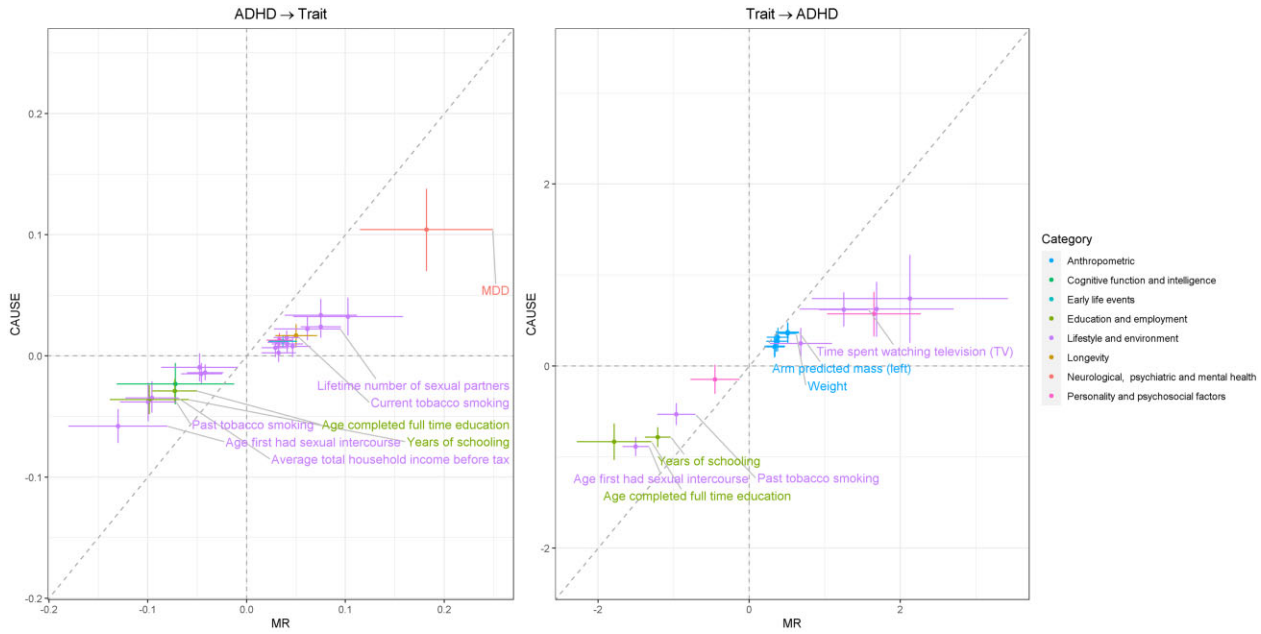


Figure 3 Causal effect estimates obtained by MR, with 95% CI, vs those obtained by CAUSE, with 95% CrI, in both directions. Effect sizes are provided as logOR for binary and categorical ordered traits. Different categories have been plotted in different colours, and traits with results meeting the Bonferroni corrected threshold in CAUSE are labelled. ADHD, attention deficit/hyperactivity disorder; CAUSE, causal analysis using summary effect estimates; CI, confidence interval; CrI, credible interval; MDD, major depressive disorder; MR, Mendelian randomization; TV, television; OR, odds ratio

trajectory, where ADHD individuals are more likely to engage in behaviours that put them at risk for negative outcomes, including smoking, problematic substance use or unsafe sexual behaviour.[58,59] We confirm previous findings for a causal effect of the genetic liability of ADHD on smoking behaviour.[29] In particular, we found a positive effect of ADHD liability on frequency of current smoking and a negative effect on past tobacco smoking, which indicates lower frequency of smoking in the past for individuals who were not heavy smokers when asked. These results may suggest that those individuals who were heavy smokers in the past carried on being heavy smokers when asked, and support that ADHD genetic liability may increase frequency of smoking and make smokers less likely to give up, which agrees with a reported negative effect of ADHD liability in smoking cessation.[29] A recent MR study, however, did not find evidence for a causal relationship from liability for ADHD to nicotine dependence, although their sample size was more limited and these results may just reflect lack of statistical power.[60] We also found suggestive evidence for an effect of the genetic liability of ADHD increasing smoking initiation and pack/years of smoking that support previous findings.[29,30] Also, despite not surpassing the strict Bonferroni correction applied, suggestive evidence of a causal effect of the genetic liability of ADHD increasing alcohol intake frequency and making individuals less likely to have their alcohol intake with meals, was found, which suggests increased risk for prejudicial use of alcohol. Although previous studies reported a weak effect of liability to ADHD on alcohol dependence, they failed to find a causal connection between ADHD and alcohol amount or alcohol use disorder.[29] Differences between the traits considered, sample sizes or the way of measuring alcohol consumption may account for inconsistent findings among studies. With respect to sexual behaviour, we show, for the first time, evidence for a causal role of the genetic liability to ADHD on lower age at first sexual intercourse and on increased lifetime number of sexual partners. In line with all these results, and with additional evidence from observational literature, we also found suggestive evidence for a causal effect of ADHD genetic liability on decreased longevity.[14]

Our results also support a positive effect of time spent watching television on ADHD, which goes in line with a reported association between ADHD and television usage and with evidence from longitudinal studies reporting an effect of increased screen time worsening a child's development and increasing risk for autism spectrum disorder.[61–63]

In addition, our findings strengthen previous evidence linking ADHD with academic, employment and financial problems.[17,64] In fact we report, for the first time, consistent evidence for a negative causal effect of ADHD liability on years of schooling, age completed full-time education and average total household income before tax, and suggestive evidence of a positive effect on Townsend deprivation index at recruitment. In addition, evidence was found in the other direction for the educational traits and suggestive evidence for Townsend deprivation index. Given that ADHD is a neurodevelopmental disorder, the latest

relationships are temporally implausible and results may reflect previous studies showing that children in families with lower parental education, family income or socioeconomic status are at higher risk for ADHD or ADHD symptoms.[6,65,66] This is consistent with dynastic effects, when genetic variants in parents may affect the next generation indirectly through their effect on the environment rather than through the inherited DNA, affecting our results.[67] For instance, it might be that parent's socioeconomic status could influence parenting skills, social development or stress levels and these, in turn, may impact on children's mental health.[66]

Along these lines, we also found evidence for liability of lower age at first sexual intercourse and of lower rate of past tobacco smoking in non-heavy smokers increasing the odds of ADHD. Evidence from the literature has linked related traits such as young parental age or maternal smoking with increased risk of ADHD in children,[5,68,69] suggesting that these chronological implausible results may also be driven by dynastic effects.

We also show evidence consistent with a positive effect of anthropometric measures on ADHD, a finding which is likely related to the effect of BMI liability on ADHD, reported in a previous MR study.[23] Later studies with longitudinal and family designs, however, pointed to this relationship being largely explained by a variety of psychosocial factors and shared genetic and environmental confounders, also including a role for parental education dynastic effects.[24,25]

Our findings also confirm the effect of ADHD liability on major depression and the lack thereof when using a broader definition,[22] but no evidence supporting previous results of a causal relationship of ADHD and birthweight or intelligence was found. These discrepancies may be explained by methodological differences between studies, including: (i) the selection of genetic instruments and additional covariates taken into account by authors considering birthweight[18]; or (ii) differences in sensitivity analyses undertaken when the protective effect of intelligence on ADHD was described.[19]

The results of the present study should be interpreted in the light of several limitations, as follows.

i. Given that we aimed to give an overview of potential causal relationships between ADHD and a considerable number of related traits, using publicly available summary statistics datasets, it was not feasible to tailor the analytical strategy separately for each trait or to carefully curate each phenotype. This may have prevented us from identifying additional evidence for causal relationships, as may be the case for birthweight mentioned above,[18] but also may have led to some spurious findings due to instrument misspecification.
ii. Aiming to avoid false-positive results, we designed a strict analysis pipeline. We undertook a comprehensive set of sensitivity analyses, including the weighted mode, recently reported to maintain the correct type I error rate in a diverse set of scenarios but also to be too conservative, particularly for large sample sizes.[70] In addition, we applied a strict multiple testing correction, despite the presence of correlated traits.
iii. Despite undertaking this range of different analyses, each one under a different set of assumptions, and selecting only results that were consistent across methods, we still identified some temporally implausible relationships. These associations could be explained statistically because the instruments were used as measures of the liability for a trait, not necessarily its observed manifestation,[70] although they may also indicate the invalidity of the genetic instruments. As discussed above, some of them, such the effect of socioeconomic status or smoking behaviour liability on ADHD, are likely driven by dynastic effects. In addition, for those traits with evidence of bidirectional causality, we cannot rule out a scenario where most of the heritable variation of both exposure and outcome is mediated by the same unobserved process, as acknowledged by CAUSE authors.[34]
iv. Some of the scenarios where MR sensitivity analyses have been carried out may not have been optimal for their performance: for instance, the limited number of genetic instruments available for ADHD (particularly relevant for MR Egger) or the difference in sample size between the exposure and the outcome in some comparisons (relevant for Steiger filtering). It may be that under more suitable circumstances, MR sensitivity analyses would be more efficient in detecting false-positive results.
v. In addition, genetic variants could be associated with more than one trait, which would make it difficult to ascertain which one is the true causal exposure. This is particularly relevant when analysing correlated traits, as it is the case in this study. Further sensitivity analyses, which were out of the scope of this study, excluding variants associated with other traits or undertaking mediation analysis would contribute to deepen our understanding and provide more robust evidence for the causal relationships identified. For instance, multivariable MR analyses have been used to detect an effect of educational attainment mediating the relationship between BMI and ADHD.[25]
vi. We must be cautious when comparing effect sizes between analyses with ADHD as exposure and outcome,

since they are often presented in different scales, and there are a number of assumptions that need to hold for reliable interpretation of causal effects for binary exposures.[71] Also, due to differences in sample size, the power was often different between analyses, which in turn makes it difficult to establish a prevailing direction of causality for traits with a bidirectional effect.

vii. There are some methodological issues that should also be considered. Given the large resources of GWAS summary statistics currently available and the flourishment of MR-related methods being developed, there is a huge potential for MR analyses to shed light into the causal relationships between many complex traits. We should, however, also bear in mind their limitations when designing studies and interpreting results. In order to avoid the effect of horizontal pleiotropy, which is relevant given that emerging results from GWAS point to a large number of genetic variants being associated with multiple traits,[51,71] we followed up MR findings with CAUSE.[34] Overall we observed consistency in direction of effects between both methods, with smaller effect sizes estimated by CAUSE than by MR and narrower intervals in general. The difference in effect sizes may be due to CAUSE also accounting for horizontal pleiotropy in its model. Although, CAUSE authors' also acknowledge that in the event of causality and no correlated pleiotropy, their causal estimates tend to shrink towards zero in comparison with other methods, partly due to prior distribution being centred at zero.[34] Under these circumstances they also report lower mean squared error for CAUSE compared with MR if causal effects are small, and there is also low power on the exposure, which seems to be the case when we consider ADHD as exposure.[34] Overall, despite the use of a range of complementary approaches in this study and of the evidence provided for causal relationships supported by the literature and by alternative study designs, such as the effect of ADHD on depression,[22] it seems that some of our results may still have been affected by biases such as dynastic effects. This highlights the caution that must still be exerted when interpreting MR findings and the need for other studies with alternative designs, such as those in families,[72] to triangulate their findings and confirm MR results.

Conclusion

Our results are consistent with a causal effect of the genetic liability of ADHD on average household income and major depressive disorder, of the genetic liability of anthropometric measures and time spent watching television on ADHD, and of a bidirectional relationship with educational achievement, smoking and sexual behaviour. Additional analyses with complementary study designs to avoid the effect of potential biases will be required to follow up these findings. However, our results may still contribute to explain part of the widespread co-occurring traits and comorbid disorders across the lifespan of individuals with ADHD, and may open new opportunities for developing preventive strategies for ADHD and for negative ADHD trajectories.

Ethics approval

Ethics approval was not required for this study, since no new data were generated and only summary statistics were analysed.

Data availability

The data underlying this article are available in the MR-Base database (doi: 10.1101/2020.08.10.244293 and doi : 10.7554/eLife.34408.001).

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

M.R. and M.S.A. designed the study and the analytical strategy and wrote the manuscript. M.S.A. conducted the analysis. C.S-M., J.A.R-Q., L.V-R., M.R., M.S.A. and P.R. interpreted the findings and reviewed critically the manuscript.

Funding

This work was supported by European Union H2020 Programme (H2020/2014–2020) under grant agreements no. 667302 (CoCA), no. 728018 (Eat2beNICE), no. 848228 (DISCOvERIE) and no. 2020604 (TIMESPAN); by the Instituto de Salud Carlos III (PI18/01788, PI19/00721, P19/01224, PI20/00041, FI18/00285 to L.V.R., CD15/00199 to C.S.M, CP09/00119 and CPII15/00023 to M.R.); by the Pla estratègic de recerca i innovació en salut (PERIS), Generalitat de Cata-lunya (METAL-Cat; SLT006/17/287); by the Agència de Gestiód'Ajuts Universitaris i de Recerca AGAUR, Generalitat de Catalunya (2017SGR1461, 2016FI_B00899 to P.R.); Ministry of Science, Innovation and Universities (IJC2018-035346-I to M.S.A); by the European Regional Development Fund (ERDF); and by 'la Marató de TV3' (092330/31) and by the ECNP Network 'ADHD across the Lifespan' [https://www.ecnp.eu/researchinnovation/ECNP-networks/List-ECNP-Networks/]. This paper reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained therein.

Conflict of interest

J.A.R.Q. was on the speakers' bureau and/or acted as consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Rubiand Raffo in the past 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, RubiShire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the past 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious and Rubió.

References

1. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry* 2007;**164**:942–48.
2. Fayyad J, Sampson NA, Hwang I *et al.*; WHO World Mental Health Survey Collaborators. The descriptive epidemiology of DSM-IV Adult ADHD in the World Health Organization World Mental Health Surveys. *Atten Defic Hyperact Disord* 2017;**9**: 47–65.
3. Faraone SV, Perlis RH, Doyle AE *et al.* Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 2005; **57**:1313–23.
4. Larsson H, Chang Z, D'Onofrio BM, Lichtenstein P. The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan. *Psychol Med* 2014; **44**:2223–29.
5. Faraone SV, Asherson P, Banaschewski T. Attention-deficit/hyperactivity disorder. *Nat Rev Dis Prim* 2015;**1**:15020.
6. Larsson H, Sariaslan A, Långström N, D'Onofrio B, Lichtenstein P. Family income in early childhood and subsequent attention deficit/hyperactivity disorder: a quasi-experimental study. *J Child Psychol Psychiatry* 2014;**55**:428–35.
7. Franke B, Michelini G, Asherson P *et al.* Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur Neuropsychopharmacol* 2018;**28**:1059–88.
8. Kooij JJS, Bijlenga D, Salerno L *et al.* Updated European Consensus Statement on diagnosis and treatment of adult ADHD. *Eur Psychiatry* 2019;**56**:14–34.
9. Martel MM, Nigg JT, Lucas RE. Trait mechanisms in youth with and without attention-deficit/hyperactivity disorder. *J Res Pers* 2008;**42**:895–913.
10. De Pauw SSW, Mervielde I. The role of temperament and personality in problem behaviors of children with ADHD. *J Abnorm Child Psychol* 2011;**39**:277–91.
11. Marx I, Hacker T, Yu X, Cortese S, Sonuga-Barke E. ADHD and the choice of small immediate over larger delayed rewards: a comparative meta-analysis of performance on simple choice-delay and temporal discounting paradigms. *J Atten Disord* 2021; **25**:171–87.
12. Shaw P, Stringaris A, Nigg J, Leibenluft E. Emotion dysregulation in attention deficit hyperactivity disorder. *Am J Psychiatry* 2014;**171**:276–93.
13. Kessler RC, Adler LA, Berglund P *et al.* The effects of temporally secondary co-morbid mental disorders on the associations of DSM-IV ADHD with adverse outcomes in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Psychol Med* 2014;**44**:1779–92.
14. Sun S, Kuja-Halkola R, Faraone SV *et al.* Association of psychiatric comorbidity with the risk of premature death among children and adults with attention-deficit/hyperactivity disorder. *JAMA Psychiatry* 2019;**76**:1141–49.
15. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
16. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018;**50**:1728–34.
17. de Zeeuw EL, van Beijsterveldt CEM, Ehli EA, de Geus EJC, Boomsma DI. Attention deficit hyperactivity disorder symptoms and low educational achievement: evidence supporting a causal hypothesis. *Behav Genet* 2017;**47**:278–89.
18. Orri M, Pingault J-B, Turecki G. Contribution of birth weight to mental health, cognitive and socioeconomic outcomes: two-sample Mendelian randomisation. *Br J Psychiatry* 2021;**219**: 507–514.
19. Savage JE, Jansen PR, Stringer S *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 2018;**50**:912–19.
20. Zhu Z, Zhu X, Liu C-L *et al.* Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. *Eur Respir J* 2019;**54**:1901507.
21. Leppert B, Riglin L, Wootton RE *et al.* The effect of ADHD on physical health outcomes: a two-sample Mendelian randomization study. *Am J Epidemiol.* 2020;**190**:1047–1055.
22. Riglin L, Leppert B, Dardani C *et al.* ADHD and depression: investigating a causal explanation. *Psychol Med* 2020;**51**: 1890–1897.
23. Martins-Silva T, Vaz JDS, Hutz MH *et al.* Assessing causality in the association between attention-deficit/hyperactivity disorder and obesity: a Mendelian randomization study. *Int J Obes* 2019; **43**:2500–08.
24. Donnchadha SÓ, Bramham J, Greene C. Rethinking the association between overweight/obesity and ADHD in children: a longitudinal and psychosocial perspective. *Ir J Psychol Med* 2020; **75**: 1–14.
25. Liu C-Y, Schoeler T, Davies NM *et al.* Are there causal relationships between attention-deficit/hyperactivity disorder and body mass index? Evidence from multiple genetically informed designs. *Int J Epidemiol* 2021;**50**:496–509
26. Byun Y-H, Ha M, Kwon H-J *et al.* Mobile phone use, blood lead levels, and attention deficit hyperactivity symptoms in children: a longitudinal study. *PLoS One* 2013;**8**:e59742.
27. Soler Artigas M, Sánchez-Mora C, Rovira P *et al.*; ADHD Group of the Psychiatric Genomics Consortium, International Cannabis Consortium. Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. *Mol Psychiatry* 2020;**25**:2493–503.
28. Vilar-Ribó L, Sánchez-Mora C, Rovira P *et al.* Genetic overlap and causality between substance use disorder and attention-deficit and hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2020;**186**:140–150.
29. Treur JL, Demontis D, Davey Smith G *et al.* Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. *Addict Biol* 2021; **26**:e12849.

30. Jang SK, Saunders G, Liu M, Jiang Y, Liu DJ, Vrieze S; 23andMe Research Team. Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychol Med* 2022;**52**:968–78.
31. Hemani G, Zheng J, Elsworth B *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**. https://elifesciences.org/articles/34408 (15 January 2022, date last accessed).
32. Elsworth B, Lyon M, Alexander T *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv* 2020. doi: 101101/20200810244293. http://biorxiv.org/content/early/2020/08/10/2020.08.10.244293 (15 January 2022, date last accessed).
33. Burgess S, Davey Smith G, Davies NM *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 2019;**4**:186.Available from: https://wellcomeopenresearch.org/articles/4-186/v2 (1 January 2022, date last accessed).
34. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet* 2020;**52**:740–47.
35. Bulik-Sullivan BK, Loh P-R, Finucane HK *et al.*; Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;**47**:291–95. Mar
36. Bulik-Sullivan B, Finucane HK, Anttila V *et al.*; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;**47**:1236–41.
37. Buniello A, MacArthur JAL, Cerezo M *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;**47**:D1005–12.
38. Elsworth BL, Mitchell R, Raistrick CA, Paternoster L, Hemani G, Gaunt T. MRC IEU UK Biobank GWAS pipeline version 1. MRC Integrative Epidemiology Unit, *University of Bristol*, 2017. https://doi.org/10.5523/bris.2fahpksont1zi26xosyamqo8rr (13 February 2021, date last accessed).
39. Berndt SI, Gustafsson S, Mägi R *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013;**45**:501–12.
40. Lambert JC, Ibrahim-Verbaas CA, Harold D *et al.*; European Alzheimer's Disease Initiative (EADI). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;**45**:1452–58.
41. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun* 2018;**9**:905.
42. Howard DM, Adams MJ, Shirali M *et al.*; 23andMe Research Team. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun* 2018;**9**:1470.
43. Wray NR, Ripke S, Mattheisen M *et al.*; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018;**50**:668–81.
44. Lee JJ, Wedow R, Okbay A *et al.*; Social Science Genetic Association Consortium. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;**50**:1112–21.
45. Clifton EAD, Perry JRB, Imamura F *et al.* Genome-wide association study for risk taking propensity indicates shared pathways with body mass index. *Commun Biol* 2018;**1**:36.
46. Demontis D, Walters RK, Martin J *et al.*; 23andMe Research Team. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019;**51**:63–75.
47. Pilling LC, Kuo C-L, Sicinski K *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)* 2017;**9**:2504–20.
48. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 2013;**178**:1177–84.
49. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;**46**:1985–98.
50. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.
51. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;**50**:693–98.
52. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.
53. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017;**32**:377–89.
54. Bowden J, Greco MF, Del Minelli C *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *Int J Epidemiol* 2016;**45**:1961–74.
55. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997;**65**:557.
56. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017;**13**:e1007081.
57. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* 2012;**36**:214–24.
58. Charach A, Yeung E, Climans T, Lillie E. Childhood attention-deficit/hyperactivity disorder and future substance use disorders: comparative meta-analyses. *J Am Acad Child Adolesc Psychiatry* 2011;**50**:9–21.
59. Schoenfelder EN, Kollins SH. Topical review: ADHD and health-risk behaviors: toward prevention and health promotion. *J Pediatr Psychol* 2016;**41**:735–40.
60. Vink JM, Treur JL, Pasman JA, Schellekens A. Investigating genetic correlation and causality between nicotine dependence and ADHD in a broader psychiatric context. *Am J Med Genet B Neuropsychiatr Genet* 2021;**186**:423–29.
61. Lingineni RK, Biswas S, Ahmad N, Jackson BE, Bae S, Singh KP. Factors associated with attention deficit/hyperactivity disorder

among US children: results from a national survey. *BMC Pediatr* 2012;**12**:50.

62. Kushima M, Kojima R, Shinohara R *et al.*; Japan Environment and Children's Study Group. Association between screen time exposure in children at 1 year of age and autism spectrum disorder at 3 years of age. *JAMA Pediatr* 2022;**176**:384–91.
63. Madigan S, Browne D, Racine N, Mori C, Tough S. Association between screen time and children's performance on a developmental screening test. *JAMA Pediatr* 2019;**173**:244–50.
64. Barkley RA. Major life activity and health outcomes associated with attention-deficit/hyperactivity disorder. *J Clin Psychiatry* 2002;**63**(Suppl 1):10–15.
65. Russell AE, Ford T, Williams R, Russell G. The association between socioeconomic disadvantage and attention Deficit/Hyperactivity Disorder (ADHD): a systematic review. *Child Psychiatry Hum Dev* 2016;**47**:440–58.
66. Torvik FA, Eilertsen EM, McAdams TA *et al.* Mechanisms linking parental educational attainment with child ADHD, depression, and academic problems: a study of extended families in The Norwegian Mother, Father and Child Cohort Study. *J Child Psychol Psychiatry* 2020;**61**:1009–18.
67. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. *Hum Mol Genet* 2019;**28**:R170–79.
68. Min X, Li C, Yan Y. Parental age and the risk of ADHD in offspring: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2021; **18**:4939.
69. Hvolgaard Mikkelsen S, Olsen J, Bech BH, Obel C. Parental age and attention-deficit/hyperactivity disorder (ADHD). *Int J Epidemiol* 2017;**46**:409–20.
70. Hemani G, Bowden J, Haycock P *et al.* Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv* 2017. doi: 101101/173682. https://www.biorxiv.org/content/early/2017/08/23/173682 (15 January 2022, date last accessed).
71. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018;**33**:947–52.
72. Brumpton B, Sanderson E, Heilbron K *et al.*; 23andMe Research Team. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* 2020;**11**:3519.

ARTICLE OPEN

Check for updates

Comprehensive analysis of omics data identifies relevant gene networks for Attention-Deficit/Hyperactivity Disorder (ADHD)

Judit Cabana-Domínguez[1,2,3,4]✉, María Soler Artigas[1,2,3,4], Lorena Arribas[1,2], Silvia Alemany[1,2], Laura Vilar-Ribó[1,2,3], Natalia Llonga[1,2], Christian Fadeuilhe[1,2,3,5], Montse Corrales[1,2,3,5], Vanesa Richarte[1,2,3,5], Josep Antoni Ramos-Quiroga[1,2,3,5] and Marta Ribasés[1,2,3,4]✉

Attention-deficit/hyperactivity disorder (ADHD) is a highly prevalent neurodevelopmental disorder that results from the interaction of both genetic and environmental risk factors. Genome-wide association studies have started to identify multiple genetic risk loci associated with ADHD, however, the exact causal genes and biological mechanisms remain largely unknown. We performed a multi-step analysis to identify and characterize modules of co-expressed genes associated with ADHD using data from peripheral blood mononuclear cells of 270 ADHD cases and 279 controls. We identified seven ADHD-associated modules of co-expressed genes, some of them enriched in both genetic and epigenetic signatures for ADHD and in biological pathways relevant for psychiatric disorders, such as the regulation of gene expression, epigenetics and immune system. In addition, for some of the modules, we found evidence of potential regulatory mechanisms, including microRNAs and common genetic variants. In conclusion, our results point to promising genes and pathways for ADHD, supporting the use of peripheral blood to assess gene expression signatures in psychiatric disorders. Furthermore, they highlight that the combination of multi-omics signals provides deeper and broader insights into the biological mechanisms underlying ADHD.

Translational Psychiatry (2022)12:409 ; https://doi.org/10.1038/s41398-022-02182-8

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a highly prevalent neurodevelopmental disorder that affects around 5–6% of children and adolescents worldwide, and in 40–65% of cases persist into adulthood [1]. It is mainly characterized by inattention and/or hyperactivity and high levels of impulsivity.

ADHD is a complex disorder that results from the interaction of both genetic and environmental risk factors, with an estimated heritability of 70–80% throughout the lifespan [2]. Several studies support the role of both common and rare genetic variants in the development of ADHD, although its etiology and pathogenesis still remain largely unknown [2]. The first genome-wide association study (GWAS) meta-analysis identifying genetic risk variants for ADHD (20,183 cases and 35,191 controls) was published in 2019 [3]. They identified 12 independent ADHD risk loci and estimated that common variants account for 22% of the total ADHD heritability. In addition, very recently, a larger GWAS meta-analysis on ADHD reported 21 new loci and a reduced estimated SNP heritability ($h^2_{SNP} = 14\%$) [4]. These data highlight that part of the genetic variance still needs to be explained, which may be accounted, in part, for gene by environment interactions [5]. Epigenetic processes (i.e. histone modifications, DNA methylation and microRNAs) are potential mechanisms by which environmental risk factors lead to changes on gene expression and long-lasting alterations in the neuronal circuits found in psychiatric disorders like ADHD [6]. Recently, the first epigenome-wide association study (EWAS) in peripheral blood mononuclear cells (PBMCs) from adults with ADHD was published, identifying four regions differentially methylated and located in genes previously related to autoimmune disorders, cancer, or neuroticism [7]. Additional EWAS in saliva and whole blood have been performed both in adults and children with ADHD diagnosis or ADHD symptoms, however, results among studies are not consistent and further studies with larger sample sizes are needed [8–12].

Although genetic and epigenetic factors that contribute to the etiology of ADHD have started to be identified through GWAS and EWAS, their biological relevance is difficult to characterize, in part, because genetic risk loci were usually associated with the nearest gene, which may not be necessarily the true causal one. In contrast, the analysis of gene expression profiles provides a closer physiological picture of the disorder that is easier to interpret, and reduces the burden of multiple testing. Transcriptome studies in ADHD, nevertheless, are limited by the inaccessibility of brain samples and have been focused on whole blood or PBMCs. To date, eight transcriptomic studies on ADHD have been performed, highlighting alterations in genes involved in several neuronal functions and the immune system [13–20]. However, the studies performed so far were based on differential gene expression

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. [5]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. ✉email: judit.cabana@vhir.org; marta.ribases@vhir.org

Received: 5 May 2022 Revised: 8 September 2022 Accepted: 13 September 2022
Published online: 24 September 2022

analyses between ADHD cases and controls, which assume that every gene acts as an independent unit in the expression landscape and select genes based on statistical significance. In contrast, gene co-expression network analyses use an unsupervised framework to identify groups of genes with similar expression patterns (co-expressed genes) independently of any phenotype and then correlate these gene modules with a phenotype of interest. This approach has been widely used to characterize patterns of co-expression in normal brain [21, 22] and both in brain and blood samples from several psychiatric disorders [23–28].

In the present study, we aimed to perform a multi-step analysis to identify and characterize modules of co-expressed genes associated with ADHD using expression data from PBMCs of ADHD cases and controls. To further understand the biological relevance and provide a more accurate picture of the regulatory mechanisms, we performed a comprehensive characterization of genes in each module and combined genomic and transcriptomic data to identify loci that may regulate the ADHD-associated co-expressed genes.

MATERIALS AND METHODS

Study design

A comprehensive and multi-step approach was applied to identify and characterize modules of co-expressed genes in PBMCs. In the first step we ran a Weighted Gene Correlation Network Analysis (WGCNA) on the processed transcriptomic data from 270 ADHD cases and 279 controls and assessed the association of the resulting co-expression modules with ADHD status. Subsequently, we disentangled the biological relevance of the ADHD-associated co-expression modules by (i) performing enrichment analyses in brain expression, functional pathways, druggable genes and miRNA target genes, (ii) combining results with ADHD genetic, transcriptomic, and epigenetic signatures, and (iii) running a co-expression module eQTL analysis to identify loci regulating the ADHD-associated modules of co-expressed genes (Fig. 1).

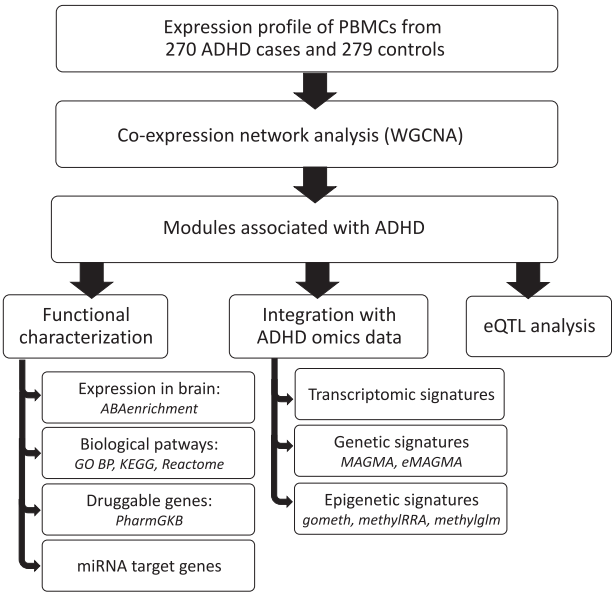


Fig. 1 Flowchart of the study. Modules of co-expressed genes were identified from peripheral blood mononuclear cells (PBMCs) of processed transcriptomic data from 270 ADHD cases and 279 controls by using Weighted Gene Correlation Network Analysis (WGCNA). Then, we assessed the association of the resulting modules with the ADHD status and investigated their biologically relevance by (i) performing enrichment analyses in brain expression (*ABAenrichment* R package), functional pathways, druggable genes and miRNA target genes using WebGestAlt webtool; (ii) integrating ADHD transcriptomic, genetic and epigenetic data from GWAS meta-analysis [3] and EWAS [7] on ADHD; and (iii) running a co-expression module eQTL analysis to identify loci regulating the ADHD-associated modules of co-expressed genes.

Participants

Analysis of co-expression modules was performed in an in-house sample of 270 ADHD cases (59.3% male, mean age = 34.2 years, s.d = 11.7) and 279 controls (56.9% male, mean age = 36.6 years, s.d = 9.9). All subjects were of European ancestry. Clinical assessment was conducted by structured interviews and self-reported questionnaires as previously described [19]. Detailed information is available in Supplementary Information. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, methods were performed in accordance with the relevant guidelines and regulations and written informed consent was obtained from all subjects before inclusion in the study.

Transcriptome profiling and weighted gene correlation network analysis (WGCNA)

RNA from PBMCs was isolated, hybridized to GeneChip Human Gene 1.1 ST 96-Array plate (Affymetrix) and data were analyzed as previously described [19] (Supplementary Information). Modules of co-expressed genes were identified from processed transcriptomic data by the *WGCNA* R-package [29]. A soft-thresholding power of 4 was selected (Fig. S1) and one-step network construction and module detection was performed considering an unsigned network type with default values (additional details in Supplementary Information). Gene expression for each module was represented by a module eigengene, derived from its first principal component and treated as a quantitative trait in the downstream analyses. The association between the module eigengenes and ADHD status or potential confounding factors (age, sex, RNA integrity number (RIN), and batch) was tested using regression analyses. Bonferroni correction was applied to correct for multiple testing considering the overall number of co-expression modules constructed ($P < 0.05/27$ modules <1E-03).

Within each module we examined the correlation between module membership (an indicator of the intramodular connectivity of a gene based on the association between its expression and the module eigengene) and gene significance (effect size of the association between each gene and ADHD) using Pearson correlation (Fig. S2).

Enrichment analyses in the ADHD-associated co-expression modules

We assessed whether genes in each ADHD-associated co-expression module were expressed in specific brain regions at different developmental stages using data from the Allen Human Brain Atlas with *ABAEnrichment* R-package [30] (additional details in Supplementary Information). Then, enrichment analyses with the webtool WebGestAlt (WEB-based GEne SeT AnaLysis Toolkit, http://www.webgestalt.org/) [31] were performed on: (i) Gene Ontology non-redundant Biological Process (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome, (ii) target genes of pharmacological drugs based on the information from PharmGKB and (iii) miRNA target genes. False discovery rate *P*-value (P_{FDR}) < 0.05 was set as the significance threshold.

In addition, the correlation between the identified miRNA and its corresponding module eigengene was tested using the non-parametric Spearman rank correlation test in a subset of 310 individuals included in the WGCNA (60% overlap; 150 ADHD cases and 160 controls) from whom miRNA expression profile data from PBMC were available as described in Sanchez-Mora et al. [16]. Expression was available and retrieved from a total of 27 mature miRNAs and Bonferroni correction was used to adjust for multiple testing ($P < 0.05/27$ tests <1.85E−03).

Integrative analysis of ADHD-associated co-expression modules and ADHD omics data

ADHD transcriptomic signatures. After quality control and sample processing, differential gene expression profiles between the 270 ADHD cases and 279 controls used in the WGCNA analysis were obtained with *Limma* R-package [32]. Only genes with $P_{FDR} < 0.05$ and fold change (FC) > |1.15| were considered differentially expressed and were used to test for enrichment in the ADHD-associated co-expression modules using a F-Fisher test and Bonferroni correction across all modules ($P < 0.05/7$ modules <7.1E−03).

ADHD genetic signatures. The identified ADHD-associated co-expression modules were used as gene sets to test for enrichment in ADHD genetic

signatures, considering the European ancestry GWAS summary statistics on ADHD described by Demontis et al. [3]. Gene-based analyses were run in MAGMA_v1.08 [33] using the SNP-wise mean model, and SNPs were assigned to genes based on a positional-approach and eMAGMA [34, 35]. Competitive gene-set analysis was performed using *P*-values obtained from each gene-based analysis. Bonferroni correction was applied to correct for multiple testing ($P < 7.1E-03$; Supplementary Information)

ADHD methylation signatures. To test for enrichment in ADHD epigenetic marks in modules of co-expressed genes we used the summary statistics of an EWAS on PBMCs from 103 ADHD patients and 100 controls [7] (90% sample overlap with the in-house sample used in WGCNA), setting the unadjusted *P*-value < 0.01 to select differentially methylated proves ($n = 3967$ CpG sites). We considered enrichment in epigenetic signatures in a co-expression module when the three approaches used (*methylglm, methylRRA* and *gsameth* [36]) were significant after applying the Bonferroni correction for multiple comparison corrections ($P < 7.1E-03$; Supplementary Information).

Gene-module eQTL analysis and functional annotation

Genetic information was available from a subset of 231 ADHD subjects and 264 controls included in the WGCNA. A GWAS with each module eigengene as the dependent variable were performed to identify genetic variants associated with each co-expression module. After ascertain normality of module eigengene (Table S1), seven gene-module eQTL analyses were run under an additive linear regression model using PLINK_1.09, adding as covariates the first 10 principal components, sex, age, and the genotyping wave. Lead SNPs were identified in each eQTL analysis considering a *P*-value < 1E−06 and functionally annotated using the FUMA protocol (Functional Mapping and Annotation of Genome-Wide Association Studies, https://fuma.ctglab.nl/) [37] (Supplementary Information).

Raw data from this article are not publicly available because of limitations in ethical approvals and the summary data will be available upon request.

RESULTS

The WGCNA identified a total of 27 modules of co-expressed genes with size ranging from 33 to 2191 genes (Fig. S1). 42.7% of genes ($N = 8114$) were not assigned to any module and remained in the module M0. Seven co-expression modules were associated with ADHD after multiple testing correction (modules M1–M7, Table S2). No association between module eigengenes and potential confounders, including age, sex, RIN, or batch, was detected for any module (Table S2). All modules were consistent across samples and have characteristic band structures suggestive of well-defined modules (Fig. S2). Interestingly, modules M1, M3, and M6 showed high module membership—gene significance correlation ($r^2 > 0.4$), suggesting that the higher the connectivity of a gene within the module, the stronger the association with ADHD (Fig. S3).

Different patterns of gene expression in the brain at different developmental stages were found across ADHD-associated co-expression modules. M2 genes are broadly expressed in the whole brain during the lifespan, while genes in M7 are expressed in a specific brain area, the M1C_primary motor cortex, only during the prenatal stage. Besides, genes in modules M1 and M4 show broader expression in different areas from the telencephalon during the prenatal period and are mainly expressed in the cerebellar cortex after birth (Table 1 and S3).

To explore the biological relevance of ADHD-associated co-expression modules further, we performed a functional enrichment analysis in genes in each module and found that several of them were enriched in genes involved in pathways previously related to psychiatric disorders [38], including the *posttranscriptional regulation of gene expression* and *epigenetics* (M1 and M7), *covalent chromatin remodeling* (M4) or *immune system* and *inflammatory response* (M5), among others (Table 1 and S4-6). We also performed an enrichment analysis in druggable genes in the ADHD-associated co-expression modules. For six out of the seven modules we identified enrichment in target genes of at least one drug, being *Antiinfective for systemic use* and *Antineoplastic and immunomodulating agents* the most common Anatomical Therapeutic Chemical classification categories across all modules (Table 1 and S7-8). Interestingly, module M5, enriched in genes involved in the immune system and inflammatory response, showed enrichment in drugs from all Anatomical Therapeutic Chemical categories, especially those related to the immune response, as expected.

Enrichment in miRNA target genes was identified in modules M1 and M7. Genes in M1 were targeted by 24 families of miRNAs, resulting on 40 mature miRNAs, and genes in M7 were targeted by five mature miRNAs (Table 1 and S9). Consistently, a significant correlation between the eigengene profile of module M1 and the expression of four out of 27 of these miRNAs (hsa-miR-142–5p, hsa-miR-181a-5p, hsa-miR-192–5p, and hsa-miR-215–5p) was found in a subset of 310 individuals (150 ADHD cases and 160 controls) from which miRNA and gene expression from PBMCs was available (Fig. S4 and Table S10).

Then, we further explored the ADHD-associated co-expression modules by integrating transcriptomics with genetic and epigenetic data on ADHD. We explored whether genes differentially expressed between ADHD cases and controls were grouped in any of the identified ADHD-associated modules of co-expressed genes, and found a significant enrichment in module M5 ($P < 2.2E-16$), which also remained significant when considering only highly connected genes (module membership > 0.8; Table 1). In addition, we found module M4 significantly enriched in genetics ($P_{MAGMA} = 1.8E-03$; $P_{eMAGMA} = 4.2E-03$) and epigenetics ($P_{gsameth} = 1.6E-04$; $P_{methylRRA} = 1.1E-03$; $P_{methylglm} = 8.1E-05$) signatures for ADHD, using data from GWAS meta-analysis [3] and EWAS [7] on ADHD (Table 1 and S11).

We performed a co-expression module eQTL analysis to identify loci regulating ADHD-associated modules in a subset of 495 individuals included in the WGCNA (91.3%) from whom genomic and gene expression profiles were available. After strict quality control criteria, we ran a GWAS on module eigengenes of each of the seven ADHD-associated co-expression modules independently (M1–M7; Fig. S5). QQ plots indicate minimal effects of genomic inflation, and consequently population substructure, on the analyses (Fig. S6). No SNP overcame the genome-wide significance threshold, but 12 independent genomic loci showed suggestive evidence of association ($P < 1E-06$) with different module eigengenes (Table 2 and Fig. S7). Functional annotations revealed that these loci lay on regions of open chromatin and that most of the signals were intergenic or intronic (Fig. 2). Several SNPs in these genomic risk loci were likely to affect the binding of transcription factors (RBD score = 2b; rs73866266, rs59928606, rs10830974 and rs36098630), had CADD scores > 12.37, suggesting high deleteriousness (rs73170573, rs13408514, rs1508617, rs9565360, and rs10830974; Fig. 2, Table 2 and S12) or were located in regulatory regions of the brain (rs73170578, rs62096513, and rs12583109), according to the information of enhancer and promotor histone marks from the HaploReg webtool [39] (Table 2). In addition, four SNPs, rs62096513, rs6707596, rs66506812 and rs2462337, lay in nearby genes encoding transcription factors (*ZSCAN30*, *SP3*, *CSRNP3* and *CUX1*, respectively; Table 2). Of them, rs6707596 nearby *SP3* and rs62096513 located in intron 1 of *ZSCAN30* were cis-eQTL of these genes in PBMCs in our sample (Fig. 3). Interestingly, the co-expression module M1, which showed suggestive evidence of association with rs62096513 that lies in blood and brain regulatory regions of *ZSCAN30* and is cis-eQTL in PBMCs, is enriched in target genes for this specific transcription factor (P = 1.27E-07), which suggest that *ZSCAN30* may be upstream regulator of the M1 module of co-expressed genes.

Table 1. Summary of the main results from the analyses on the ADHD-associated modules of co-expressed genes.

Module	No. genes	Association with ADHD[a]	Brain expression	Main biological function	ATC1 drug categories	miRNA target genes	ADHD transcriptomic signature	ADHD genetic signature[b]	ADHD epigenetic signature[c]
M1	1546	Effect= 9.53 *P* = 9.3E−06	Prenatal—telencephalon Postnatal—cerebellum	Posttranscriptional regulation of gene expression Methylation	A, B, C, L	40 miRNAs (24 miR families)	–	–	–
M2	1239	Effect = −6.49 *P* = 1.7E−03	Whole brain	mRNA processing Metabolism	A, B, C, D, J, L	–	–	–	***mglm* = 5.2E−03**
M3	63	Effect= 7.87 *P* = 1.8E−03	–	–	–	–	–	–	*gsam* = 0.0269 *mglm* = 7.9E−03
M4	885	Effect = −7.82 *P* = 1.8E−03	Prenatal - telencephalon Postnatal - cerebellum	Regulation of gene expression Covalent chromatin modification	B, L	-	-	***MA* = 1.8E−03** ***eM* = 4.2E−03**	***gsam* = 1.6E−04** ***mRRA* = 1.1E−03** ***mglm* = 8.1E−05**
M5	133	Effect = −11.17 *P* = 2.6E-06	–	Immune response	A, B, C, D, H, J, L, M, N, R, S	–	***P* = 2.2E−16**	–	–
M6	258	Effect= 11.58 *P* = 2.1E−07	–	mRNA processing	J	–	–	–	–
M7	185	Effect= 8.66 *P* = 4.4E-05	Prenatal - M1C	Posttranscriptional regulation of gene expression Epigenetics	B	5 miRNAs	-	-	*gsam* = 0.0157

In bold results from gene-set analyses that overcome multiple testing corrections for each method.
ATC1 (Anatomical Therapeutic Chemical) categories: A = Alimentary system and metabolism; B = Blood and hematopoietic organs; C = Heart therapy; D = Dermatics; H = Hormone preparations excl.sex; J = Antiinfective for systemic use; L = Antineoplastic and immunomodulating agents; M = musculo-skeletal system; N = Nervous system; R = Respiratory tract; S = Sensory organs. MA = MAGMA; eM = eMAGMA; mglm = methlyglm; gsam = gsameth; mRRA = methylRRA; DE = differentially expressed; M1C = human primary motor cortex.
[a]Logistic regression Effect = log(OR).
[b]Demontis et al. Nat Genet. 2019;51(1):63–75.
[c]Rovira et al. Transl Psychiatry. 2020;10(1):199.

Table 2. Top hits from GWAS on the module eigengenes from the seven ADHD-associated co-expression modules.

Module	SNP	GWAS P-value	MAF	intra/ intergenic	Nearby genes	TF[a]	Enrichment on target genes[b]	Regulatory region in[c]		eQTL in PBMCs[d]	eQTL whole blood[e]
								blood	brain		
M1	rs73170578	2.17E−07	0.061	intra	*CNTNAP2*	–	–	–	SNP in LD (rs73170573)	–	–
M1	rs62096513	5.40E−07	0.066	intra	*ZSCAN30*	YES	YES	YES	YES	YES	–
M3	rs75370437	2.56E−07	0.055	intra	*ACSM1, ACSM3*	–	–	–	–	–	*ACSM1*
M4	rs6707596	5.53E−07	0.344	intra	*SP3*	YES	NO	SNP in LD (rs13010038)	–	YES	*AC106900.6*
M4	rs72873859	5.65E−07	0.296	intra	*HEATR5B*	–	–	–	–	–	*STRN, PRKD3, RP11–288C18.1*
M5	rs66506812	3.35E−07	0.055	intra	*CSRNP3*	YES	NA	–	–	–	–
M5	rs10830974	7.21E−07	0.360	inter	*SLC36A4, MTNR1B*	–	–	–	–	–	–
M6	rs73866245	1.35E−07	0.023	inter	*PCOLCE2, PAQR9, U2SURP*	–	–	–	–	–	–
M6	rs114943986	1.37E−07	0.037	inter	*PRR16*	–	–	–	–	–	–
M6	rs2462337	3.52E−07	0.271	inter	*CUX1*	YES	NO	YES	–	–	–
M6	rs72806699	6.39E−07	0.010	intra	*MEAK7*	–	–	–	–	–	–
M7	rs12583109	6.00E−07	0.153	intra	*NBEA, MAB21L1*	–	–	–	SNP in LD (rs9565360)	–	–

[a]TF: the nearby gene is a transcription factor.
[b]Enrichment analysis of transcription factor's target genes in their corresponding module using F-Fisher test. NA: list of TF target genes not available.
[c]Haploreg v4.1 tool (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php), when one SNP in high LD ($r^2 > 0.8$) is located in a regulatory region is indicated.
[d]Results from the eQTL analysis in our samples.
[e]GTEx v8 (https://www.gtexportal.org/home/).

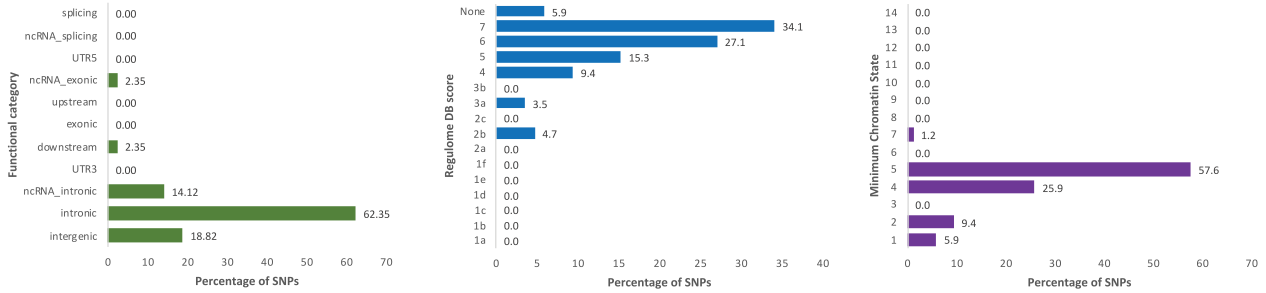


Fig. 2 Functional categories, Regulome DB scores, and minimum chromatin states for independent risk loci associated to any module eigengene. Regulome DB score predicts likelihood of regulatory functionality, lower scores indicate higher likelihood. Further information can be found in Boyle et al. [68]. Minimum Chromatin State across 127 tissue and cell types, lower scores indicate higher accessibility, with states 1–7 referring to open chromatin states.

DISCUSSION

In the present study, we used a network-based approach to identify novel ADHD-associated modules of co-expressed genes in PBMCs. To further investigate the biological significance of the ADHD-associated networks identified, we performed a comprehensive characterization of each module by performing enrichment analysis in biological pathways and drug or miRNA target genes. We also performed an integrative analysis by combining transcriptomic, genetic and epigenetic data on ADHD and run an eQTL analysis to identify genetic variants that could regulate the ADHD-associated modules of co-expressed genes. Our results identified seven ADHD-associated modules of co-expressed genes and support that the study of gene correlation networks may improve our understanding of the complex molecular systems underling ADHD.

Two of the ADHD-associated co-expression modules identified (M1 and M7), were enriched in genes involved in *posttranscriptional regulation of gene expression* and *epigenetic modifications*, two relevant pathways in the pathogenesis of ADHD [6]. In the same line, we found enrichment in target genes for several miRNAs in these two modules. In particular, the expression of four of them (hsa-miR-142–5p, hsa-miR-181a-5p, hsa-miR-192–5p and hsa-miR-215–5p) also correlates with the eigengene profile of module M1, pointing them as potential upstream regulatory mechanisms underlying the M1 co-expression network. Some of these miRNAs have been previously related to ADHD, like miR-192–5p upregulated in PBMCs of ADHD patients [16], and comorbid psychiatric disorders, such as miR-192–5p and miR-215–5p that were differentially expressed in the dorsolateral prefrontal cortex of major depression patients [40] or miR-181a-5p extensively related to drug addiction both in mice and human studies [41–45]. Interestingly, these miRNAs share many target genes, suggesting a complex and redundant regulatory system, particularly in the case of miR-291–5p and miR-215–5p which recognize the same seed sequence. Several of these miRNAs may regulate a number of central genes (those with high intramodular connectivity) from module M1, such as *CPSF6* encoding a subunit of a cleavage factor required for the RNA cleavage and polyadenylation processing, which was previously related to externalizing behaviors including ADHD [46], and *RICTOR*, which plays an essential role during the neurodevelopment and has been associated with hyperactivity and reduced anxiety-like behavior in conditional *knock-out* mice in the dorsal neural progenitor cells [47].

Module M1, as well as M2 and M6, were also enriched in genes that encode proteins involved in the *processing of messenger RNA* (mRNA), which includes any process related to the conversion of a primary mRNA transcript into one or more mature mRNAs. mRNA processing and alternative splicing are key processes for both the diversification of protein isoforms and the spatio-temporal control of transcripts, essential for the neuronal development, maturation, and synaptic function [48], and genetic variants in genes encoding these proteins have been related to rare neurodevelopmental disorders [49], as well as common psychiatric disorders like schizophrenia [50].

Module M5 was enriched in genes involved in *immune system* and *inflammatory response*, pathways known to play an important role in the development of neuropsychiatric disorders [38, 51, 52], particularly in ADHD [53]. Moreover, genes in module M5, and to a less extent in module M2, are targeted by a great variety of known therapeutic drugs, especially by those that target the immune system (including the Anatomical Therapeutic Chemical categories *Antiinfective for systemic use* and *Antineoplastic and immunomodulating agents*), pointing to genes in these co-expression networks as potential therapeutic targets. Importantly, a recent study that explored the druggable genome in ADHD also pointed to drugs to treat autoimmune disorders and malignancies as a potential novel path for the treatment of ADHD [54]. Besides, in module M5 we also found an enrichment in genes differentially expressed in ADHD patients compared with controls, suggesting that differentially expressed genes in ADHD cases are co-expressed and participate in the same biologic pathways. Furthermore, this enrichment was also significant when considering only highly connected genes, highlighting that the genes differentially expressed are central nodes highly connected in this network, reinforcing their relevance in the pathophysiology of ADHD.

The integrative analysis of transcriptomics, genomics, and epigenomics data on ADHD revealed that genes in module M4, also involved in the *regulation of gene expression* and *epigenetic mechanisms*, were enriched in both genetic and epigenetic signatures previously described for ADHD [3, 7]. We used two complementary approaches to assign ADHD-associated SNPs to genes, based on position or eQTL results, and found consistent results. *PNPLA2* and *IQSEC1* were the central genes in the module more significantly associated with ADHD using both methods. *PNPLA2* encodes an enzyme involved in the hydrolysis of triglycerides in adipose tissue, and has been related to obesity [55], a highly comorbid disorder in ADHD [56]. In addition, a recent study pointed *PNPLA2* as one of the most high-confidence causal genes for ADHD, after combining GWAS, eQTL and gene expression data [57]. *IQSEC1* encodes a guanine nucleotide exchange factor, essential for the maintenance of glutamatergic synapses [58], one of the key neurotransmitter systems involved in the pathophysiology of ADHD in combination with dopamine [59, 60].

The eQTL analysis did not reveal any genetic variant that overcame the genome-wide significance threshold, but we found 12 independent genomic loci that showed suggestive evidence of association ($P < 1e{-}06$) with the different module eigengenes. We identified a genetic variant associated with the co-expression module M1, rs62096513, which is located in a blood and brain

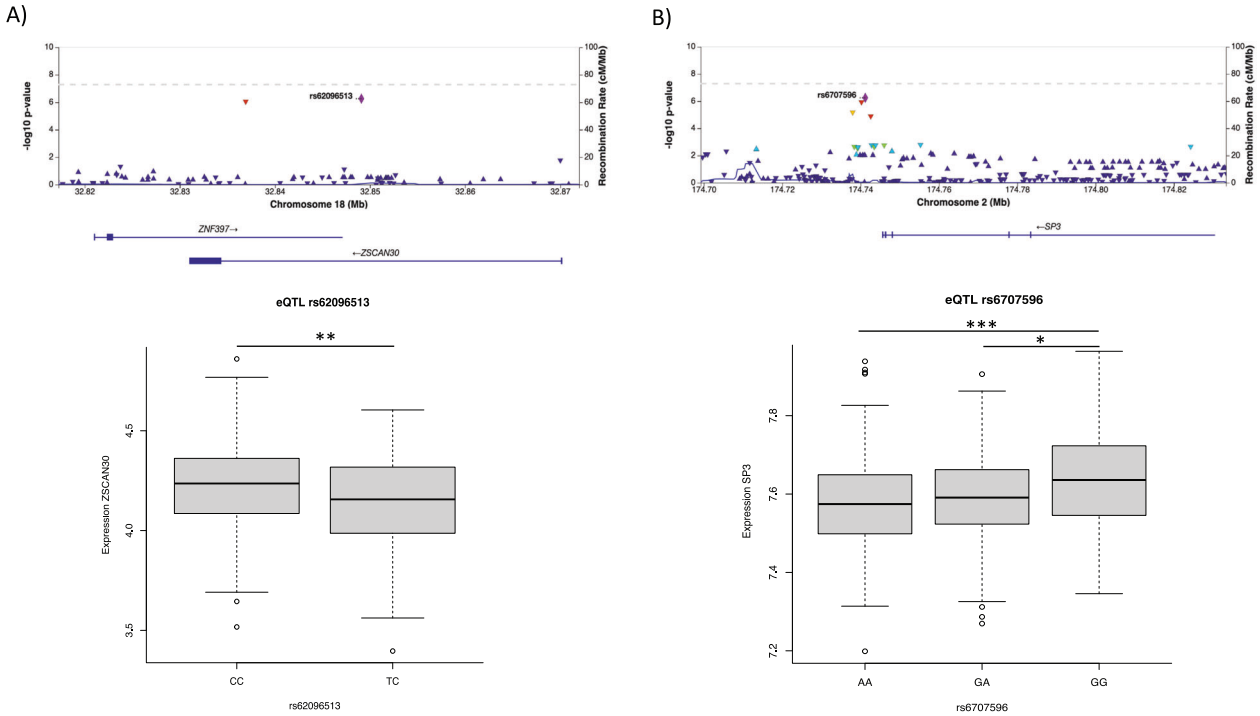


Fig. 3 Regional plots and cis-eQTL analyses in PBMCs in our in-house sample of 270 ADHD cases and 279 controls. In the top figure, regional plots for module eigengenes for the (**A**) rs62096513 and (**B**) rs6707596 loci and in the bottom figure, boxplots showing the effect of rs62096513 in *ZSCAN30* and rs6707596 in *SP3* gene expression. Results from the linear regression *p*-value are shown: $*P < 0.5$, $**P < 0.01$, $***P < 0.001$.

regulatory region of a transcription factor, *ZSCAN30*, and regulates its expression in PBMCs. Interestingly, module M1 was enriched in target genes for ZSCAN30 that is also included in the same module, suggesting that this transcription factor is an upstream regulator of the co-expressed genes in the module. Besides, we identified another genetic variant associated with the M1 module eigengene, rs73170578, located in *CNTNAP2*, which encodes a neuronal transmembrane protein member of the neurexin superfamily that function as cell adhesion molecules and receptors. Both rare and common genetic variants in *CNTNAP2* have been associated with neurodevelopmental disorders [61, 62], with a special relevance in ADHD and autism [63, 64]. In addition, module M4 was associated with rs6707596, that is an eQTL of the *SP3* gene in PBMCs, a transcription factor involved in synaptic plasticity [65]. Finally, we identified four genetic variants associated with M6 module, among them, rs2462337 is located in a blood regulatory region upstream the *CUX1* gene, a transcription factor involved in the control of neuronal differentiation and the regulation of dendritic branching, spine development, and synapse formation in cortical neurons [66].

Gene networks analyses reduce the dimensionality of genome-wide gene expression data without losing important biological information and alleviate the multiple testing burden associated with the traditional gene-based methods. Similar network-based studies have been performed using gene expression data in both brain and blood in several psychiatric disorders like autism, schizophrenia and bipolar disorder [23–28]. These studies were usually performed in small sample sizes ($n < 100$ individuals), limiting their statistical power. In contrast, we improved the resolution and robustness of gene networks by considering more than 500 subjects, which allowed the identification of seven ADHD-associated modules enriched in relevant and highly significant biological pathways. However, although our transcriptomic analyses were performed mainly in medication-naive ADHD patients without comorbid disorders (93.7% of all ADHD cases), we cannot discard that these conditions may have influenced the results of the present study. So, further studies in the same cell type are required to confirm our results. Additionally, the identified modules were based on expression data from PBMCs, a non-invasive peripheral tissue whose expression profile has been proposed as a surrogate for expression profiling in the central nervous system [67], and further evidence in the brain is required to confirm their role in the pathophysiology of the disorder.

In summary, we conducted a multi-step analysis to identify and characterize modules of co-expressed genes associated with ADHD using expression data from PBMCs in ADHD cases and controls. We identified seven ADHD-associated modules of co-expressed genes, some of them being enriched in both genetic and epigenetic signatures for ADHD and on biological pathways relevant for psychiatric disorders, such as the regulation of gene expression, epigenetic mechanisms and immune signaling. We also found preliminary evidence for some potential regulatory mechanisms, including microRNAs and genetic variants, for some of the ADHD-associated modules of co-expressed genes identified. These results pinpoint promising genes and pathways for ADHD, support the use of peripheral blood to assess gene expression signatures for the disorder and highlight that the combination of multi-omics signals provides deeper and broader insights into the biological mechanisms underlying the disorder.

REFERENCES

1. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry. 2018;5:175–86.
2. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. Mol Psychiatry. 2019;24:562–75.
3. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

4. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Farajzadeh L, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. medRxiv. 2022; https://doi.org/10.1101/2022.02.14.22270780.
5. Palladino VS, McNeill R, Reif A, Kittel-Schneider S. Genetic risk factors and gene-environment interactions in adult and childhood attention-deficit/hyperactivity disorder. Psychiatr Genet. 2019;29:63–78.
6. Mirkovic B, Chagraoui A, Gerardin P, Cohen D. Epigenetics and attention-deficit/hyperactivity disorder: new perspectives? Front Psychiatry. 2020;11. https://doi.org/10.3389/fpsyt.2020.00579.
7. Rovira P, Sánchez-Mora C, Pagerols M, Richarte V, Corrales M, Fadeuilhe C, et al. Epigenome-wide association study of attention-deficit/hyperactivity disorder in adults. Transl Psychiatry. 2020;10:1–12.
8. Wilmot B, Fry R, Smeester L, Musser ED, Mill J, Nigg JT. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. J Child Psychol Psychiatry Allied Discip. 2016;57:152–60.
9. Walton E, Pingault JB, Cecil CAM, Gaunt TR, Relton CL, Mill J, et al. Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. Mol Psychiatry. 2017;22:250–6.
10. Mooney MA, Ryabinin P, Wilmot B, Bhatt P, Mill J, Nigg JT. Large epigenome-wide association study of childhood ADHD identifies peripheral DNA methylation associated with disease and polygenic risk burden. Transl Psychiatry. 2020;10:1–12.
11. Meijer M, Klein M, Hannon E, van der Meer D, Hartman C, Oosterlaan J, et al. Genome-Wide DNA methylation patterns in persistent attention-deficit/hyperactivity disorder and in association with impulsive and callous traits. Front Genet. 2020;11:16.
12. van Dongen J, Zilhão NR, Sugden K, Heijmans BT, 't Hoen PAC, van Meurs J, et al. Epigenome-wide Association Study of attention-deficit/hyperactivity disorder symptoms in adults. Biol Psychiatry. 2019;86:599–607.
13. Marín-Méndez JJ, Patiño-García A, Segura V, Ortuño F, Gálvez MD, Soutullo CA. Differential expression of prostaglandin D2 synthase (PTGDS) in patients with attention deficit-hyperactivity disorder and bipolar disorder. J Affect Disord. 2012;138:479–84.
14. De Jong S, Newhouse SJ, Patel H, Lee S, Dempster D, Curtis C, et al. Immune signatures and disorder-specific patterns in a cross-disorder gene expression analysis. Br J Psychiatry. 2016;209:202–8.
15. Garcia-Martínez I, Sánchez-Mora C, Soler Artigas M, Rovira P, Pagerols M, Corrales M, et al. Gene-wide association study reveals RNF122 ubiquitin ligase as a novel susceptibility gene for attention deficit hyperactivity disorder. Sci Rep. 2017;7:1–13.
16. Sánchez-Mora C, Soler Artigas M, Garcia-Martínez I, Pagerols M, Rovira P, Richarte V, et al. Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. Neuropsychopharmacology. 2019;44:890–7.
17. Lorenzo G, Braun J, Muñoz G, Casarejos MJ, Bazán E, Jimenez-Escrig A. RNA-Seq blood transcriptome profiling in familial attention deficit and hyperactivity disorder (ADHD). Psychiatry Res. 2018;270:544–6.
18. Nuzziello N, Craig F, Simone M, Consiglio A, Licciulli F, Margari L, et al. Integrated analysis of microRNA and mRNA expression profiles: an attempt to disentangle the complex interaction network in attention deficit hyperactivity disorder. Brain Sci. 2019;9:288.
19. Mortimer N, Sánchez-Mora C, Rovira P, Vilar-Ribó L, Richarte V, Corrales M, et al. Transcriptome profiling in adult attention-deficit hyperactivity disorder. Eur Neuropsychopharmacol. 2020;41:160–6.
20. McCaffrey TA, St. Laurent G, Shtokalo D, Antonets D, Vyatkin Y, Jones D, et al. Biomarker discovery in attention deficit hyperactivity disorder: RNA sequencing of whole blood in discordant twin and case-controlled cohorts. BMC Med Genomics. 2020;13. https://doi.org/10.1186/s12920-020-00808-8.
21. Hu Y, Pan J, Xin Y, Mi X, Wang J, Gao Q, et al. Gene expression analysis reveals novel gene signatures between young and old adults in human prefrontal cortex. Front Aging Neurosci. 2018;10:259.
22. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. Nat Neurosci. 2008;11:1271–82.
23. Li X, Zhang Y, Wang L, Lin Y, Gao Z, Zhan X, et al. Integrated analysis of brain transcriptome reveals convergent molecular pathways in autism spectrum disorder. Front Psychiatry. 2019;10:706.
24. Maschietto M, Tahira AC, Puga R, Lima L, Mariani D, Da Silveira Paulsen B, et al. Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia Bioinformatic and algorithmical studies. BMC Med Genomics. 2015;8:1–15.
25. Gerring ZF, Gamazon ER, Derks EM. Consortium for the MDDWG of the PG. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. PLOS Genet. 2019;15:e1008245.
26. Radulescu E, Jaffe AE, Straub RE, Chen Q, Shin JH, Hyde TM, et al. Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. Mol Psychiatry. 2018;25:791–804.
27. Jalbrzikowski M, Lazaro MT, Gao F, Huang A, Chow C, Geschwind DH, et al. Transcriptome profiling of peripheral blood in 22q11.2 deletion syndrome reveals functional pathways related to psychosis and autism spectrum disorder. PLoS ONE 2015;10. https://doi.org/10.1371/journal.pone.0132542.
28. Hess JL, Tylee DS, Barve R, de Jong S, Ophoff RA, Kumarasinghe N, et al. Transcriptomic abnormalities in peripheral blood in bipolar disorder, and discrimination of the major psychoses. Schizophr Res. 2020;217:124.
29. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:1–13.
30. Grote S, Prüfer K, Kelso J, Dannemann M. ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain. Bioinformatics. 2016;32:3201–3.
31. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47:W199–W205.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
33. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
34. Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. Bioinformatics. 2021;37:2245–9.
35. Gerring ZF, Gamazon ER, Derks EM. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. PLoS Genet. 2019;15:e1008245.
36. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. 2021;22:1–26.
37. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.
38. O'dushlaine C, Rossin L, Lee PH, Duncan L, Parikshak NN, Newhouse S, et al. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci. 2015;18:199.
39. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016;44:D877–D881.
40. Yoshino Y, Roy B, Dwivedi Y. Differential and unique patterns of synaptic miRNA expression in dorsolateral prefrontal cortex of depressed subjects. Neuropsychopharmacology. 2021;46:900–10.
41. Sim MS, Soga T, Pandy V, Wu YS, Parhar IS, Mohamed Z. MicroRNA expression signature of methamphetamine use and addiction in the rat nucleus accumbens. Metab Brain Dis. 2017;32:1767–83.
42. Zhao Y, Zhang K, Jiang H, Du J, Na Z, Hao W, et al. Decreased expression of plasma MicroRNA in patients with methamphetamine (MA) use disorder. J Neuroimmune Pharm. 2016;11:542–8.
43. Xu W, Zhao M, Lin Z, Liu H, Ma H, Hong Q, et al. Increased expression of plasma hsa-miR-181a in male patients with heroin addiction use disorder. J Clin Lab Anal. 2020;34. https://doi.org/10.1002/jcla.23486.
44. Chandrasekar V, Dreyer JL. microRNAs miR-124, let-7d and miR-181a regulate Cocaine-induced Plasticity. Mol Cell Neurosci. 2009;42:350–62.
45. Chandrasekar V, Dreyer JL. Regulation of MiR-124, Let-7d, and MiR-181a in the accumbens affects the expression, extinction, and reinstatement of cocaine-induced conditioned place preference. Neuropsychopharmacology. 2011;36:1149–64.
46. Karlsson Linnér R, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci. 2021;24:1367–76.
47. Carson RP, Fu C, Winzenburger P, Ess KC. Deletion of rictor in neural progenitor cells reveals contributions of mTORC2 signaling to tuberous sclerosis complex. Hum Mol Genet. 2013;22:140–52.
48. Furlanis E, Scheiffele P. Regulation of neuronal differentiation, function, and plasticity by alternative splicing. Annu Rev Cell Dev Biol. 2018;34:451–69.
49. Fliedner A, Kirchner P, Wiesener A, van de Beek I, Waisfisz Q, van Haelst M, et al. Variants in SCAF4 cause a neurodevelopmental disorder and are associated with impaired mRNA processing. Am J Hum Genet. 2020;107:544–54.
50. Chapman RM, Tinsley CL, Hill MJ, Forrest MP, Tansey KE, Pardiñas AF, et al. Convergent evidence that ZNF804A Is a regulator of pre-messenger RNA processing and gene expression. Schizophr Bull. 2019;45:1267–78.
51. Jiang NM, Cowan M, Moonah SN, Petri WA. The impact of systemic inflammation on neurodevelopment. Trends Mol Med. 2018;24:794–804.

52. Han VX, Patel S, Jones HF, Dale RC. Maternal immune activation and neuroinflammation in human neurodevelopmental disorders. Nat Rev Neurol. 2021;17:564–79.
53. Dunn GA, Nigg JT, Sullivan EL. Neuroinflammation as a risk factor for attention deficit hyperactivity disorder. Pharmacol Biochem Behav. 2019;182:22–34.
54. Hegvik TA, Waløen K, Pandey SK, Faraone SV, Haavik J, Zayats T. Druggable genome in attention deficit/hyperactivity disorder and its co-morbid conditions. New avenues for treatment. Mol Psychiatry. 2021;26:4004–15.
55. La Cruz-Color LD, Hernández-Nazará ZH, Maldonado-González M, Navarro-Muñíz E, Domínguez-Rosales JA, Torres-Baranda JR, et al. Association of the PNPLA2, SCD1 and leptin expression with fat distribution in liver and adipose tissue from obese subjects. Exp Clin Endocrinol Diabetes. 2020;128:715–22.
56. Cortese S. The association between ADHD and obesity: Intriguing, progressively more investigated, but still puzzling. Brain Sci. 2019;9. https://doi.org/10.3390/brainsci9100256.
57. Fahira A, Li Z, Liu N, Shi Y. Prediction of causal genes and gene expression analysis of attention-deficit hyperactivity disorder in the different brain region, a comprehensive integrative analysis of ADHD. Behav Brain Res. 2019;364:183–92.
58. Scholz R, Berberich S, Rathgeber L, Kolleker A, Köhr G, Kornau HC. AMPA receptor signaling through BRAG2 and Arf6 critical for long-term synaptic depression. Neuron. 2010;66:768–80.
59. Huang X, Wang M, Zhang Q, Chen X, Wu J. The role of glutamate receptors in attention-deficit/hyperactivity disorder: From physiology to disease. Am J Med Genet Part B Neuropsychiatr Genet. 2019;180:272–86.
60. Cai Y, Xing L, Yang T, Chai R, Wang J, Bao J, et al. The neurodevelopmental role of dopaminergic signaling in neurological disorders. Neurosci Lett. 2021;741. https://doi.org/10.1016/j.neulet.2020.135540.
61. Poot M. Connecting the CNTNAP2 networks with neurodevelopmental disorders. Mol Syndromol. 2015;6:7–22.
62. Rodenas-Cuadrado P, Ho J, Vernes SC. Shining a light on CNTNAP2: complex functions to complex disorders. Eur J Hum Genet. 2014;22:171.
63. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry. 2010;15:637.
64. Vecchia ED, Mortimer N, Palladino VS, Kittel-Schneider S, Lesch KP, Reif A, et al. Cross-species models of attention-deficit/hyperactivity disorder and autism spectrum disorder: Lessons from CNTNAP2, ADGRL3, and PARK2. Psychiatr Genet. 2019;29:1–17.
65. Yamakawa H, Cheng J, Penney J, Gao F, Rueda R, Wang J, et al. The transcription factor Sp3 cooperates with HDAC2 to regulate synaptic function and plasticity in neurons. Cell Rep. 2017;20:1319–34.
66. Cubelos B, Sebastián-Serrano A, Beccari L, Calcagnotto ME, Cisneros E, Kim S, et al. Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. Neuron. 2010;66:523–35.
67. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet - Neuropsychiatr Genet. 2006;141 B:261–8.
68. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.

ACKNOWLEDGEMENTS

The authors are grateful to patients and controls who kindly participated in this research. The microarrays service was carried out at High Technology Unit (UAT) at Vall d'Hebron Research Institute (VHIR), Barcelona (Spain).

AUTHOR CONTRIBUTIONS

JCD, MR, and MSA conceived the project. CF, MC, VR, JARQ participated in the clinical assessment and in the recruitment of patients. JCD, LA, and LVR participated in the RNA isolation and preparation of samples. JCD undertook the statistical analyses. JCD, LVR, NL, SA, MR, and MSA participated in the study design and the discussion of results. JCD, MR, and MSA participated in the manuscript preparation. All authors contributed to the interpretation of the findings and revised and approved the final version of the manuscript.

FUNDING

This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-1461), the Instituto de Salud Carlos III (PI18/01788, P19/01224, PI20/00041 and FI18/00285 to L.V.R), the Ministry of Science, Innovation and Universities (IJC2018-035346-I to M.S.A), the Network Center for Biomedical Research (CIBER) to J.C.D.; the European Regional Development Fund (ERDF), the European Union H2020 Programme (H2020/2014-2020) under grant agreements no. 728018 (Eat2beNICE), no. 848228 (DISCOvERIE) and no. 2020604 (TIMESPAN) and the ECNP Network 'ADHD across the Lifespan'.

CONFLICT OF INTEREST

J.A.R.Q. was on the speakers bureau and/or acted as consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Technofarma, Rubió and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial, and Medice. The Department of Mental Health chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen-Cilag, Shire, Oryzon, Roche, Psious, and Rubió. C.F. and V.R. have received fees to give talks for Shire/Takeda and Rubió. All other authors declare no biomedical financial interests or conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02182-8.

Correspondence and requests for materials should be addressed to Judit Cabana-Domínguez or Marta Ribasés.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-022-01143-7

Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder

Veera M. Rajagopal[1,2,3], Jinjie Duan[1,2,3], Laura Vilar-Ribó[4,5,6], Jakob Grove[1,2,3,7], Tetyana Zayats[8,9,10], J. Antoni Ramos-Quiroga[4,5,6,11], F. Kyle Satterstrom[8,9], María Soler Artigas[4,5,6,11], Jonas Bybjerg-Grauholm[2,12], Marie Bækvad-Hansen[2,13], Thomas D. Als[1,2,3], Anders Rosengren[2,14], Mark J. Daly[8,9,15,16], Benjamin M. Neale[8,9], Merete Nordentoft[2,17], Thomas Werge[2,14], Ole Mors[2,18], David M. Hougaard[2,13], Preben B. Mortensen[2,19,20], Marta Ribasés[4,5,6,21], Anders D. Børglum[1,2,3] and Ditte Demontis[1,2,3] ✉

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder with onset in childhood (childhood ADHD); two-thirds of affected individuals continue to have ADHD in adulthood (persistent ADHD), and sometimes ADHD is diagnosed in adulthood (late-diagnosed ADHD). We evaluated genetic differences among childhood ($n = 14,878$), persistent ($n = 1,473$) and late-diagnosed ($n = 6,961$) ADHD cases alongside 38,303 controls, and rare variant differences in 7,650 ADHD cases and 8,649 controls. We identified four genome-wide significant loci for childhood ADHD and one for late-diagnosed ADHD. We found increased polygenic scores for ADHD in persistent ADHD compared with the other two groups. Childhood ADHD had higher genetic overlap with hyperactivity and autism compared with late-diagnosed ADHD and the highest burden of rare protein-truncating variants in evolutionarily constrained genes. Late-diagnosed ADHD had a larger genetic overlap with depression than childhood ADHD and no increased burden in rare protein-truncating variants. Overall, these results suggest a genetic influence on age at first ADHD diagnosis, persistence of ADHD and the different comorbidity patterns among the groups.

ADHD is a childhood neurodevelopmental disorder characterized by age-inappropriate levels of hyperactivity, impulsivity and inattention. The disorder affects around 5–6% of school-age children and around 3% of adults[1,2]. It is a complex disorder with both environmental and genetic factors contributing to risk. Genetic factors explain a large part of the etiology of ADHD, with an estimated twin heritability of 0.74 (ref. [3]), and the SNP heritability (that is, the contribution of common genetic variants) is substantial, explaining 22% of the phenotypic variance[4].

Around two-thirds of children diagnosed with ADHD will continue to have symptoms in adulthood[5], which is referred to as persistent ADHD. Persistent ADHD is associated with more severe outcomes compared with the one-third of individuals who no longer have ADHD as adults (remitters), for example, increased risk of substance use disorders[6,7], nicotine dependence[8] and comorbidity with other psychiatric disorders[9–11]. Several studies have reported a lower heritability for persistent ADHD than for childhood ADHD[12,13]; however, these findings have been questioned owing to methodological differences in assessment of children and adults[12,14]. It has also been suggested that persistence of symptoms has a genetic risk component specific to persistence rather than baseline symptoms[15] and that a trajectory of persistent symptoms is associated with a high load of common ADHD risk variants[16].

[1]Department of Biomedicine (Human Genetics), Aarhus University, Aarhus, Denmark. [2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [3]Center for Genomics and Personalized Medicine, Aarhus, Denmark. [4]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain. [5]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [6]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [7]BiRC, Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [8]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [9]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [10]PROMENTA, Department of Psychology, University of Oslo, Oslo, Norway. [11]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [12]Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [13]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [14]Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark. [15]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [16]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [17]Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark. [18]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Aarhus, Denmark. [19]National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark. [20]Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark. [21]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain. ✉e-mail: ditte@biomed.au.dk

Table 1 | Index variants for the genome-wide significant loci identified in the GWAS of childhood and late-diagnosed ADHD

SNP	CHR	BP	A1	A2	MAF	OR	s.e.	*P*	Nearest gene
Childhood ADHD									
rs7511800	1	44214269	T	A	0.31	0.91	0.01	7.4×10^{-11}	*ST3GAL3*
rs12653396	5	87847273	T	A	0.42	0.91	0.01	2×10^{-11}	*MEF2C*
rs28718037	18	50572697	A	G	0.33	0.92	0.01	8.7×10^{-9}	*DCC*
rs6035830	20	21265728	C	T	0.28	1.10	0.01	1.5×10^{-9}	*XRN2*
Late-diagnosed ADHD									
rs1229758	7	114229139	G	A	0.43	0.90	0.01	2.1×10^{-8}	*FOXP2*

The variant ID (SNP), chromosome position (CHR), base position in hg19 (BP), effect allele (A1), other allele (A2), minor allele frequency (MAF) of A1, OR with respect to A1, s.e., association *P* value from logistic regression and nearest gene are given.

According to the ICD-10 diagnostic criteria, ADHD is a childhood-onset disorder, and the behavioral symptoms should be present before 7 years of age (before 12 years of age according to DSM-5 diagnostic criteria) with a duration of at least 6 months. However, the disorder is often diagnosed in adolescence and can also be diagnosed during adult life, which, according to the current diagnostic criteria, is termed late-diagnosed ADHD. Currently, there are insufficient data to clarify whether ADHD diagnosed in adulthood has the same underlying causes as childhood ADHD, or whether late-diagnosed ADHD is a disorder with a different etiology from that of childhood ADHD, resulting in a later diagnosis or even adult-onset ADHD[17].

ADHD symptoms such as hyperactivity and inattention are believed to have continuous distributions in the population, with diagnosed ADHD representing the extreme end; this is supported by genetic findings[4]. The symptoms have an impairing impact on an individual's life when the accumulation of environmental and genetic risk factors exceeds a threshold. Age at first diagnosis might therefore differ depending on when this threshold is passed. Environmental factors influencing this could be age-related, such as increased educational demands in college or university, resulting in ADHD diagnosed later in life. Age at first diagnosis might also be influenced by genetic factors affecting symptom heterogeneity and/or severity, and a recent study found that individuals with late-diagnosed ADHD had a burden of common ADHD risk alleles comparable to that of individuals without ADHD[18] by analyzing polygenic scores (PGSs) based on variant weights from the latest ADHD genome-wide association study (GWAS) meta-analysis[4]. The sample size was very small ($n=98$ for late-diagnosed individuals), and further investigation is needed to elucidate the impact of genetics on age at first diagnosis and to determine whether individuals diagnosed with ADHD in adulthood differ genetically from individuals diagnosed as children.

We previously performed a large GWAS to evaluate the genetic architecture of childhood and persistent ADHD, including a total of 17,149 ADHD cases and 32,411 controls[19]. The genetic correlation (r_g) between the two groups was high ($r_g=0.81$), suggesting that childhood and persistent ADHD to a large extent have the same underlying genetic architecture. However, we noticed that the genetic correlation was significantly different from 1 ($P=0.02$), suggesting that further dissection of the genetic architecture might reveal genetic differences. Moreover, in that study, all adult individuals with ADHD were grouped together, meaning that the persistent group consisted of both individuals diagnosed in childhood with symptoms persisting into adulthood and individuals diagnosed as adults (that is, late-diagnosed ADHD). Further subgrouping of individuals with ADHD depending on age at first diagnosis could therefore reveal further information about the genetic architecture underlying the disorder and its comorbidities.

Here, we perform in-depth characterization of the polygenic architecture of childhood, persistent and late-diagnosed ADHD in a large Danish population-based case–cohort sample of ADHD cases and controls generated by iPSYCH[20]. We identify differences among the groups with respect to common ADHD risk variants and rare protein-truncating variants (rPTVs). We also report several significant differences in genetic overlap of ADHD subgroups with other phenotypes, including an increased load of autism risk variants in individuals with childhood compared with late-diagnosed ADHD and a larger genetic overlap of persistent and late-diagnosed ADHD with depression compared with childhood ADHD.

Results

Sample characteristics. Individuals diagnosed with ADHD were identified in the large nationwide population-based case–cohort sample established by iPSYCH[20] consisting of 133,296 genotyped individuals (iPSYCH1+2; Methods). ADHD cases were divided into three groups depending on age at first diagnosis (see Methods for a detailed definition of the groups): (1) childhood ADHD ($n=14,878$), defined as individuals diagnosed with ADHD in childhood; (2) persistent ADHD ($n=1,473$), defined as individuals who received an ADHD diagnosis as a child and again as adults; and (3) late-diagnosed ADHD ($n=6,961$), defined as individuals receiving their first ADHD diagnosis as adults. Controls were individuals not diagnosed with ADHD, randomly selected from the same nationwide birth cohort ($n=38,303$).

The sex distribution was different among the three groups. Females composed 23% of childhood ADHD cases, 36% of persistent ADHD cases and 41% of late-diagnosed cases, and the male/female ratio was significantly different among all three groups (Supplementary Table 1). Moreover, comorbidity patterns were different in the three groups. Autism spectrum disorder was very frequent in childhood ADHD (23% comorbid) and persistent ADHD (18% comorbid) compared with late-diagnosed ADHD (6.2% comorbid) (Supplementary Table 2). The adolescence- and/or adulthood-onset disorders, namely schizophrenia, bipolar disorder and major depressive disorder, were more frequent among individuals with persistent and late-diagnosed ADHD. As many as 27% of individuals with late-diagnosed ADHD had comorbid major depressive disorder (Supplementary Table 2).

Genome-wide association analyses of ADHD subgroups. We conducted a GWAS for each of the three ADHD subgroups. The GWAS for childhood ADHD revealed four genome-wide significant loci on chromosomes 1, 5, 18 and 20 (Table 1 and Supplementary Figs. 1a and 2a–d). Two were new ADHD risk loci (on chromosomes 18 and 20) and two were known risk loci (on chromosomes 1 and 5) identified in our previous GWAS meta-analysis of ADHD[4], which included an earlier and smaller iPSYCH sample than that analyzed

here. One genome-wide significant locus was identified for late-diagnosed ADHD on chromosome 7, located in *FOXP2* (Table 1 and Supplementary Figs. 1b and 2e). The effect size was significantly higher in late-diagnosed ADHD (odds ratio (OR) = 1.11, s.e. = 0.01) compared with childhood ADHD (OR = 1.05, s.e. = 0.02) ($P = 0.012$). No genome-wide significant loci were found for persistent ADHD, which was expected owing to the low number of cases.

SNP heritability and genetic correlations. We estimated the SNP heritability (h^2_{SNP}) using best-guess genotypes and GCTA[21] assuming a prevalence of 5% for childhood ADHD and 3% for persistent and late-diagnosed ADHD. We found the highest h^2_{SNP} in the persistent ADHD group ($h^2_{SNP} = 0.29$), followed by the late-diagnosed ADHD ($h^2_{SNP} = 0.27$) and childhood ADHD ($h^2_{SNP} = 0.24$) groups (Supplementary Table 3a). The estimates did not significantly differ from each other (Supplementary Table 3). As the population prevalence of ADHD subtypes is not known precisely, we also estimated the heritability over a range of prevalence values. At all points, persistent ADHD demonstrated the highest h^2_{SNP}, followed by late-diagnosed ADHD (Supplementary Fig. 3 and Supplementary Table 3b).

Pairwise genetic correlations between ADHD subgroups revealed a high genetic correlation between childhood and persistent ADHD ($r_g = 0.82$, s.e. = 0.08) and between persistent and late-diagnosed ADHD ($r_g = 0.77$, s.e. = 0.08), whereas the genetic correlation between childhood ADHD and late-diagnosed ADHD was moderate ($r_g = 0.65$, s.e. = 0.04) (Supplementary Table 4).

ADHD polygenic risk load in ADHD subgroups. The polygenic risk load of variants associated with general liability to ADHD in the three ADHD subgroups was evaluated by PGS analyses. All groups demonstrated a highly significantly increased ADHD-PGS load compared with controls (Supplementary Table 5). The highest mean ADHD-PGS was found for persistent ADHD (mean = 0.41, s.d. = 0.95), followed by late-diagnosed ADHD (mean = 0.27, s.d. = 0.98) and then childhood ADHD (mean = 0.26, s.d. = 0.96) (Supplementary Table 5). The ADHD-PGS load in persistent ADHD was significantly higher than that in childhood ADHD ($P = 3.0 \times 10^{-4}$) and nominally significantly higher than that in late-diagnosed ADHD ($P = 0.02$). The results did not change in a sensitivity analysis where the childhood group was split into those younger than 18 years and those older than 18 years of age by the end of follow-up (Supplementary Fig. 4).

In an attempt to replicate the findings, we performed PGS analysis in a Spanish sample consisting of 453 individuals with childhood ADHD, 270 with persistent ADHD and 889 with late-diagnosed ADHD, as well as 3,440 controls. We did not replicate the findings, with trends in the opposite direction when comparing with controls (ADHD-PGS childhood ADHD: $\beta = 0.27$, s.e. = 0.05; persistent ADHD: $\beta = 0.21$, s.e. = 0.06; late-diagnosed ADHD: $\beta = 0.19$, s.e. = 0.04). However, the differences were not significant, and we could not draw any strong conclusions based on this small replication sample.

Genetic overlap with ADHD symptoms in the general population. Genetic overlap with ADHD symptoms in the general population was estimated using results from GWAS of ADHD subgroups and from GWAS meta-analyses of measures of inattention and hyperactivity ($n = 43{,}117$) in the general population[22]. Inattention and hyperactivity were highly correlated with both childhood ADHD ($r_{g\_inattention} = 0.86$, s.e. = 0.08; $r_{g\_hyperactivity} = 0.95$, s.e. = 0.08) and persistent ADHD ($r_{g\_inattention} = 0.87$, s.e. = 0.14; $r_{g\_hyperactivity} \approx 1$, s.e. = 0.15) (Supplementary Table 6) but showed a considerably lower correlation with late-diagnosed ADHD ($r_{g\_inattention} = 0.57$, s.e. = 0.08; $r_{g\_hyperactivity} = 0.59$, s.e. = 0.07). The genetic correlation of hyperactivity with late-diagnosed ADHD was significantly lower than that observed for childhood ADHD ($P = 0.004$). In addition, the genetic correlations of ADHD symptoms with late-diagnosed ADHD were significantly less than 1 ($P_{diff\_1\_inattention} = 7.66 \times 10^{-8}$; $P_{diff\_1\_hyperactivity} = 4.71 \times 10^{-9}$; Supplementary Table 6).

PGS analyses to test for enrichment in the three ADHD subgroups of variants associated with inattention and hyperactivity identified a nominally significantly lower PGS for hyperactivity in late-diagnosed ADHD compared with childhood ADHD ($P = 0.04$; Supplementary Table 7).

Genetic overlap with psychiatric disorders and other traits. The observed differences in patterns of comorbidity with other psychiatric disorders could reflect age differences among the groups but could also be influenced by differences in genetic architecture. To evaluate this, we performed genetic correlation and PGS analyses for major psychiatric disorders (schizophrenia[23], bipolar disorder[24], major depressive disorder[25], autism spectrum disorder[26], anorexia[27], obsessive–compulsive disorder (OCD)[28], cannabis use disorder[29] and alcohol use disorder[30]). We found positive genetic correlations of ADHD subgroups with autism, schizophrenia, bipolar disorder, major depressive disorder, alcohol use disorder and cannabis use disorder, and negative genetic correlations with OCD and anorexia (Fig. 1 and Supplementary Table 8), in line with previous findings[4]. We identified a significantly higher genetic correlation of childhood ADHD with autism ($r_g = 0.48$, s.e. = 0.05) compared with late-diagnosed ADHD ($r_g = 0.27$, s.e. = 0.06), and significantly higher genetic correlations of depression and alcohol use disorder with late-diagnosed ADHD ($r_{g\_depression} = 0.69$, s.e. = 0.04; $r_{g\_alcohol\_use\_disorder} = 0.82$, s.e. = 0.2) compared with childhood ADHD ($r_{g\_depression} = 0.45$, s.e. = 0.04; $r_{g\_alcohol\_use\_disorder} = 0.39$, s.e. = 0.09) ($P_{diff\_depression} = 8.7 \times 10^{-7}$; $P_{diff\_alcohol\_use\_disorder} = 3.8 \times 10^{-5}$) (Fig. 1 and Supplementary Table 8). The PGS results demonstrated the same pattern, with a significantly increased autism-PGS in childhood ADHD compared with late-diagnosed ADHD, a significantly higher PGS in persistent and late-diagnosed ADHD than in childhood ADHD for depression and cannabis use disorder, and a significantly increased PGS in late-diagnosed ADHD compared with childhood ADHD for schizophrenia and bipolar disorder (Fig. 2 and Supplementary Table 9).

We also performed genetic correlation and PGS analyses for phenotypes representing domains that had previously[4] demonstrated high genetic correlations with ADHD: cognition (educational years[31]), overweight (body mass index (BMI)[32]), reproduction (age at first birth[33]), mortality (maternal age of death[34]) and sleep (insomnia[35]). We identified stronger negative genetic correlations of late-diagnosed ADHD compared with childhood ADHD for educational years ($P_{difference} = 1.7 \times 10^{-5}$; $r_{g\_late\text{-}diagnosed} = -0.61$, s.e. = 0.03; $r_{g\_childhood} = -0.46$, s.e. = 0.03), increased age at first birth ($P_{difference} = 8.9 \times 10^{-5}$; $r_{g\_late\text{-}diagnosed} = -0.73$, s.e. = 0.04; $r_{g\_childhood} = -0.54$, s.e. = 0.04) and increased mother's age at death ($P_{difference} = 2.6 \times 10^{-4}$; $r_{g\_late\text{-}diagnosed} = -0.79$, s.e. = 0.10; $r_{g\_childhood} = -0.48$, s.e. = 0.08) (Fig. 1 and Supplementary Table 8). Furthermore, we identified a significantly less negative PGS in childhood ADHD compared with persistent and late-diagnosed ADHD for number of educational years ($P_{childhood\_vs\_persistent} = 8.02 \times 10^{-8}$; $P_{childhood\_vs\_late\text{-}diagnosed} = 4.35 \times 10^{-14}$) and a less negative PGS for age at first birth for childhood ADHD compared with late-diagnosed ADHD (Fig. 2 and Supplementary Table 9).

Except from autism and OCD, the highest PGS load was observed for persistent ADHD (Supplementary Table 10a); however, owing to the small sample size of this group, we had limited power to detect pairwise PGS differences for this group compared with the other two groups.

Finally, we performed two PGS sensitivity analyses. First, we evaluated PGS for autism, schizophrenia, bipolar disorder and depression in the three ADHD groups, excluding individuals with

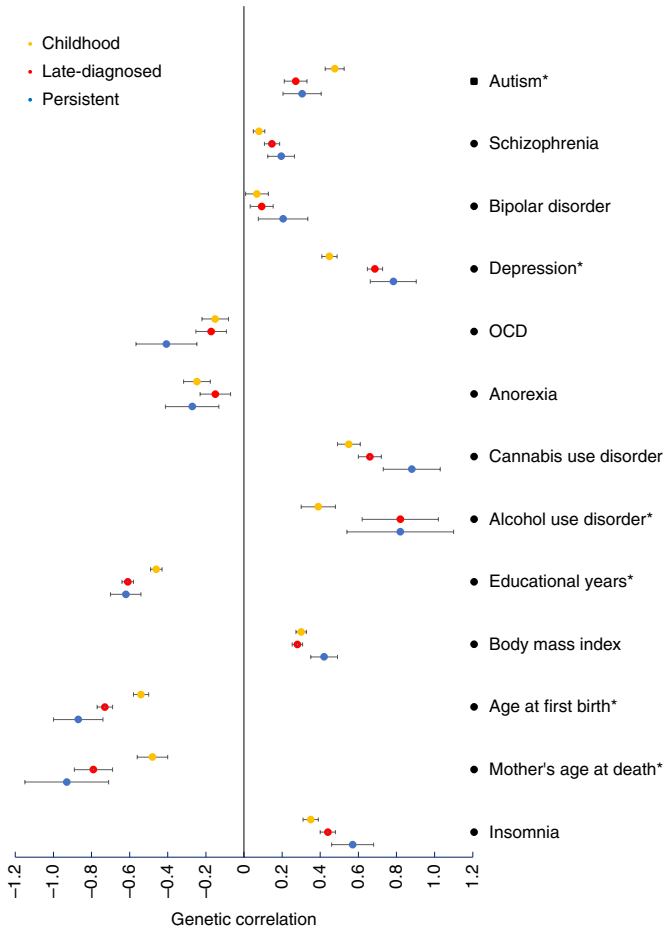


Fig. 1 | Genetic correlations of ADHD subgroups with major psychiatric disorders and other phenotypes. Results were used from genome-wide association analyses of ADHD subgroups including childhood (n = 14,878 individuals), late-diagnosed (n = 6,961 individuals) and persistent (n = 1,473 individuals) ADHD against 38,303 control individuals. Error bars (horizontal lines) represent s.e. Asterisk indicates a significant difference (after Bonferroni correction) with a two-sided P value less than 0.0013 in the genetic correlation observed for childhood ADHD compared with late-diagnosed ADHD.

the disorder corresponding to the PGS being analyzed. The results revealed the same patterns seen in the full sample, but PGS-autism was only nominally significantly higher in childhood ADHD compared with late-diagnosed ADHD (P = 0.02) (Supplementary Fig. 5 and Supplementary Tables 10b,c).

Second, we repeated the PGS analyses but this time with the childhood group split into two subgroups based on whether individuals were younger than 18 years or older than 18 years of age by the end of follow-up. The PGS loads in the two childhood ADHD groups were generally similar, except those for schizophrenia and bipolar disorder, for which the load was higher in individuals older than 18 years of age (Supplementary Fig. 6 and Supplementary Table 11).

Burden of rare variants in ADHD subgroups. We have previously demonstrated an enrichment of rPTVs in highly constrained genes in ADHD cases[36]. We explored this further by evaluating the load of rPTVs in the three ADHD subgroups using whole exome-sequencing (WES) data available for a subset of the iPSYCH cohort (childhood ADHD, n = 4,987; persistent ADHD, n = 748; late-diagnosed ADHD, n = 1,915; controls, n = 8,649). The burden of rPTVs and rare synonymous variants (rSYNs) in the three ADHD subgroups was tested in three gene sets: (1) highly constrained genes that are evolutionarily intolerant to loss-of-function mutations with a probability of being loss-of-function intolerant (pLI) score >0.9 (ref. [37]) (3,488 genes); (2) de novo constrained genes, the subset of highly constrained genes that overlap with another gene set of 285 genes found to be enriched with de novo mutations in individuals with neurodevelopmental disorders[38] (241 genes); and (3) low constrained genes that are relatively tolerant to loss-of-function mutations with a pLI score <0.1 (9,662 genes). When compared with controls, the load of rPTVs in highly constrained genes was significantly increased in childhood and persistent ADHD (childhood ADHD $\beta = 0.13$, s.e. = 0.02, $P = 2.41 \times 10^{-11}$; persistent ADHD $\beta = 0.12$, s.e. = 0.04, $P = 1.90 \times 10^{-3}$) but was lower and not significantly enriched in late-diagnosed ADHD ($\beta = 0.06$, s.e. = 0.03, $P = 0.02$). The same pattern was observed for de novo highly constrained genes (Fig. 3 and Supplementary Table 12). No pairwise comparisons among ADHD subgroups showed significant differences, but there was a tendency towards a higher burden of rPTVs in childhood compared with late-diagnosed ADHD in de novo highly constrained genes (P = 0.096). By comparison, we did not find enrichment of any rSYNs in the gene sets or enrichment of rPTVs in low constrained genes (Fig. 3 and Supplementary Table 12).

Discussion

We identified differences in the genetic architecture of childhood, persistent and late-diagnosed ADHD based on unique data from a large population-based Danish case–cohort sample. We identified the first four genome-wide significant loci associated with childhood ADHD, two of which were new ADHD risk loci located on chromosomes 18 and 20. The chromosome 18 index variant was located in *DCC*, a gene recently linked to general liability to psychiatric disorders[39]; thus, it does not seem specific to ADHD. The chromosome 20 locus was intergenic, and the index variant has previously been shown to have genome-wide significant associations with weight-related phenotypes[34]. We also identified a genome-wide significant locus associated with late-diagnosed ADHD in *FOXP2*. This locus received considerable attention when it was first reported to be a risk locus for ADHD[4], owing to the role of *FOXP2* in cognition, language and speech development[40–42]; recently, we also found *FOXP2* to be a risk gene for cannabis use disorder[29]. The effect size was significantly higher in late-diagnosed ADHD compared with childhood ADHD, suggesting that the association of *FOXP2* with ADHD is driven to a greater extent by late-diagnosed ADHD than by childhood ADHD.

When assessing the polygenic architecture, we identified the highest SNP heritability for persistent ADHD. In concordance with this, we observed the highest polygenic risk load for general liability to ADHD in individuals with persistent ADHD. This observation is consistent with the hypothesis that individuals with a higher genetic risk load for ADHD are those who will continue to have ADHD symptoms as adults. This finding is also in line with a previous study reporting an association of ADHD-PGS with persistence of ADHD symptoms in the general population[16] and with a recent smaller study reporting higher ADHD-PGS in persistent ADHD compared with late-diagnosed ADHD (although the difference was not significant)[18]. The ADHD-PGS represents the general liability to diagnosed ADHD because the scores are derived from data representing all individuals with ADHD in the Danish population born between 1981 and 2008 (Methods). In relation to this, it should be noted that the training data included a higher proportion of childhood ADHD cases than persistent and late-diagnosed ADHD cases, which could potentially result in better prediction of childhood ADHD. Despite this, we observed the opposite; there was a higher ADHD-PGS in the adult groups compared with childhood ADHD

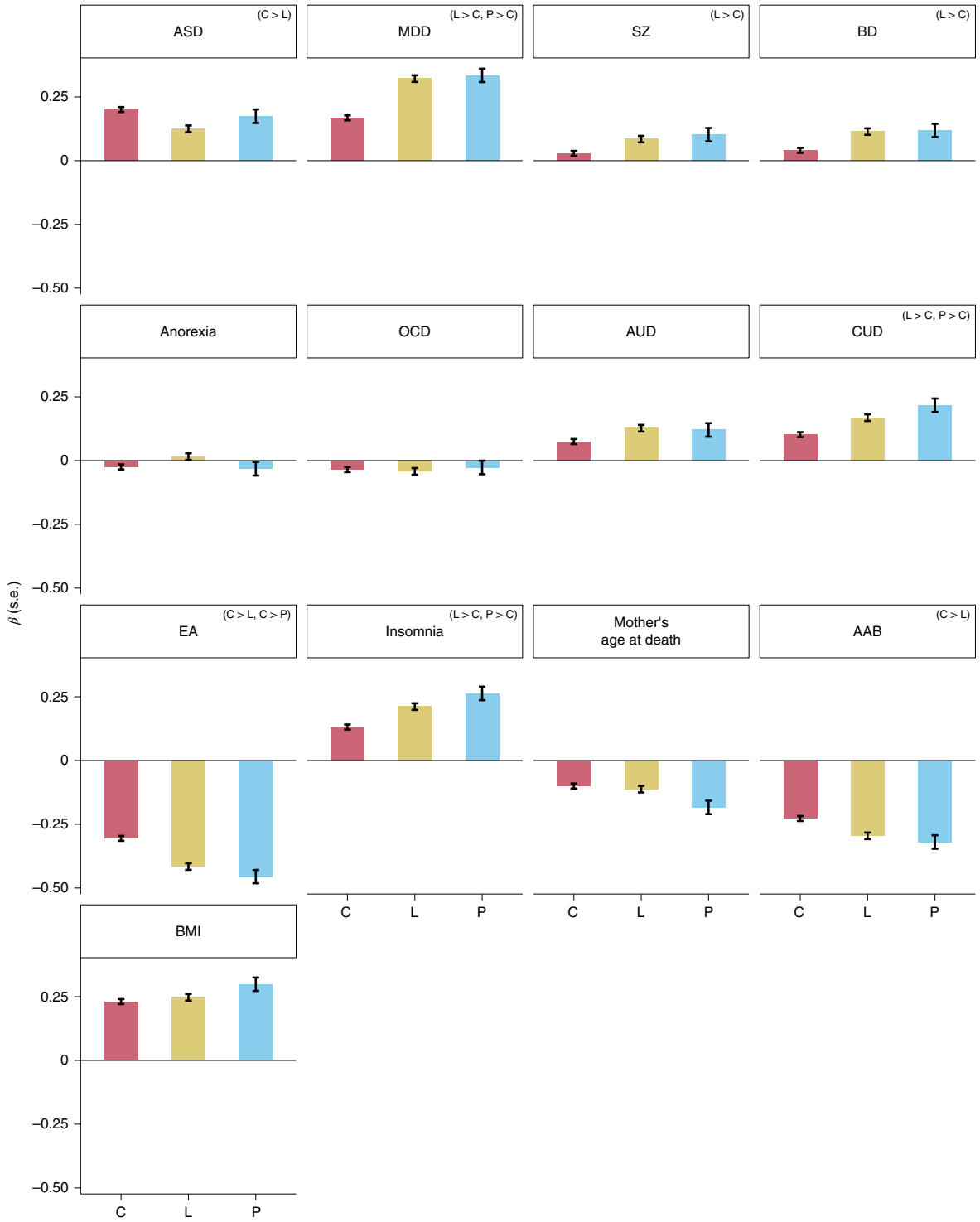


Fig. 2 | Associations of PGS with childhood, persistent and late-diagnosed ADHD. Results are shown for childhood (n = 14,878 individuals), persistent (n = 1,473 individuals) and late-diagnosed (n = 6,961 individuals) ADHD. PGS analysis was performed for psychiatric disorders: autism spectrum disorder (ASD), depression (MDD), schizophrenia (SZ), bipolar disorder (BD), anorexia, OCD, alcohol use disorder (AUD) and cannabis use disorder (CUD). PGSs for five phenotypes represent domains highly correlated with ADHD: educational attainment (EA), insomnia, mother's age at death and age at first birth (AOB). On the y axis is the beta from multiple regression against controls (n = 38,303 individuals); error bars (vertical lines) represent s.e. (Supplementary Table 10). Significant pairwise differences (after Bonferroni correction) with a two-sided P value less than 0.0013 are given in the right corner of the header: 'C' indicates childhood ADHD, 'L' indicates late-diagnosed ADHD, 'P' indicates persistent ADHD, and the direction of the difference in the betas is given by '>' (Supplementary Table 9).

(with significantly increased ADHD-PGS in persistent ADHD and a slight increase in late-diagnosed ADHD), reinforcing the validity of the results.

The findings could not be replicated in the Spanish cohort. This could have been because of the relatively small replication cohort or differences in ascertainment. The iPSYCH cohort reflects the

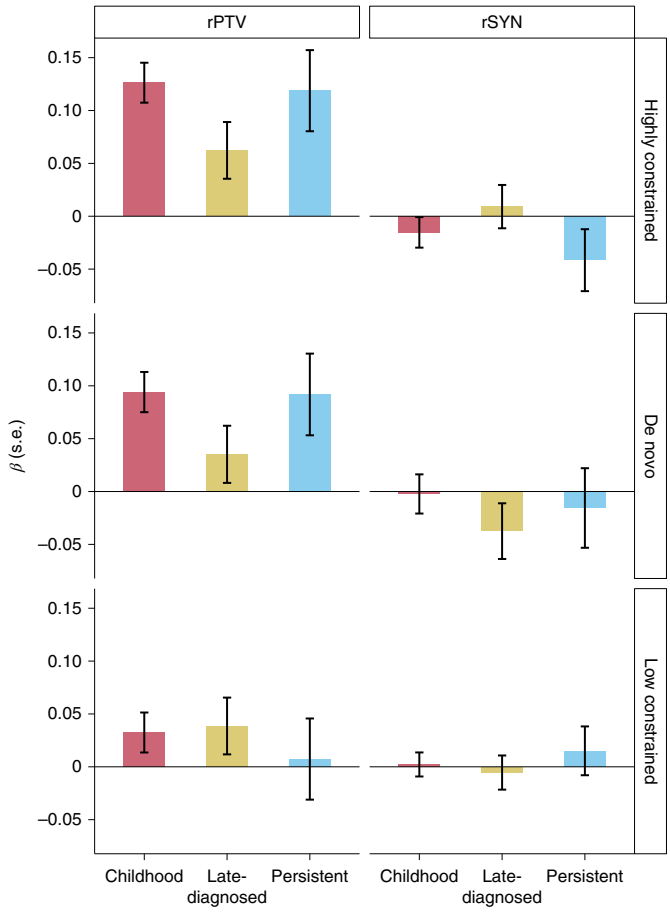


Fig. 3 | The load of rPTVs and rSYNs in ADHD subgroups. Highly constrained genes intolerant to loss-of-function mutations (pLI > 0.9); de novo highly constrained genes, which in another study have been found to be enriched with de novo mutations in individuals with neurodevelopmental disorders; and low constrained genes tolerant to loss-of-function mutations (pLI < 0.1). The *y* axis represents the beta from multiple regression of childhood ($n = 4{,}987$ individuals), persistent ($n = 748$ individuals) and late-diagnosed ($n = 1{,}915$ individuals) ADHD compared with controls ($n = 8{,}649$ individuals); error bars (vertical lines) represent s.e. For pairwise comparisons of ADHD subgroups, see Supplementary Table 12.

genetic architecture across all ADHD cases in the Danish population, whereas the Spanish cohort is a smaller clinical data set that could be influenced by unknown ascertainment biases.

The genetic correlation of childhood ADHD with persistent ADHD was high ($r_g = 0.82$) and at the same level as that reported previously ($r_g = 0.81$)[19], whereas the genetic correlation of childhood ADHD with late-diagnosed ADHD was lower ($r_g = 0.65$), suggesting some differences in the polygenic architecture of childhood and late-diagnosed ADHD. This could be due in part to a lower load of variants associated with hyperactivity and inattention in individuals with late-diagnosed ADHD; we observed a higher genetic correlation of ADHD symptoms with childhood and persistent ADHD compared with late-diagnosed ADHD. Likewise, PGS analyses suggested a lower burden of ADHD-symptom-associated variants in late-diagnosed ADHD than in the other groups. We cannot exclude the possibility that differences in age distributions in the GWAS meta-analyses of ADHD symptoms influenced the results. However, the age distribution was similar in the persistent and late-diagnosed groups (in which all individuals were older than 18 years of age), indicating that the decreased genetic overlap with late-diagnosed ADHD is not caused by age differences. Later diagnosis of ADHD could therefore be explained in part by genetic factors, with late-diagnosed individuals being less genetically predisposed to be inattentive and hyperactive, leaving their ADHD unnoticed until later in life.

The comorbidity pattern in the three groups differed, with a higher comorbidity of autism spectrum disorder in childhood (23%) and persistent ADHD (18%) compared with late-diagnosed ADHD (6.2%), in line with previous reports concerning comorbid autism in children with ADHD[43]. The observed comorbidity patterns were reflected in the genetic analyses, in which we found a significantly higher genetic correlation of autism with childhood ADHD compared with late-diagnosed ADHD, and higher PGS-autism in childhood ADHD compared with late-diagnosed ADHD. Therefore, childhood ADHD seems to be genetically more closely related to autism than late-diagnosed ADHD.

Comorbidities of psychiatric disorders with onset in adolescence and/or adulthood were more frequent in persistent and late-diagnosed ADHD (Supplementary Table 2). This is probably due in part to the age difference, because many individuals in the childhood group would be too young to develop these disorders. However, our results suggest that age alone cannot explain the comorbidity patterns. Genetics may play a part; in general we observed a higher genetic correlation or PGS for several of the disorders (schizophrenia, bipolar disorder, alcohol use disorder, cannabis use disorder and depression) in persistent and late-diagnosed ADHD compared with childhood ADHD (Figs. 1 and 2 and Supplementary Tables 8 and 9). This was particularly striking for depression, with a significantly higher PGS in individuals with persistent and late-diagnosed ADHD compared with those with childhood ADHD. The high comorbidity of depression with ADHD in adults is well known, but the causes are not. ADHD itself could be a risk factor[44,45], but genetic risk factors are also considered to exist, owing to the high genetic correlation of ADHD with depression[4]. Our results suggest genetic heterogeneity among ADHD cases: individuals with persistent and late-diagnosed ADHD are at higher risk of comorbid depression owing to the underlying genetic architecture of the disorder in these groups.

In analyses of five selected phenotypes (educational years, insomnia, mother's age at death, age at first birth and BMI) representing domains highly genetically correlated with ADHD[4], we observed the highest genetic correlations and the highest PGS load in persistent ADHD, followed closely by late-diagnosed ADHD, and the lowest in childhood ADHD (except for BMI), suggesting a similar polygenic architecture of persistent and late-diagnosed ADHD for these phenotypes (Figs. 1 and 2; see Supplementary Information, Note 1, regarding mother's age at death). These results also support the idea that the negative outcomes associated with persistent ADHD, such as decreased school performance[46] and sleep problems[47], are influenced by genetics to a greater extent than in childhood ADHD.

We found an increased burden of rPTVs in highly constrained genes in persistent and childhood ADHD compared with controls, but not in late-diagnosed ADHD. There was also a tendency towards a significantly higher burden of rPTVs in de novo highly constrained genes in childhood ADHD compared with late-diagnosed ADHD. These findings suggest that with respect to rPTVs, which are variants expected to have greater impact on the disorder than common variants, the genome of individuals with late-diagnosed ADHD is less burdened. When considering both common and rare variants, the emerging picture suggests that childhood ADHD is genetically more similar to autism (high genetic correlation with autism and increased rPTV burden), whereas late-diagnosed ADHD genetically is more similar to depression (high genetic correlation with depression and no significant increase in rPTVs in highly constrained genes compared with controls).

We could not rule out ascertainment differences among children and adults. However, the genetic correlations of late-diagnosed and

persistent ADHD (which requires an ADHD diagnosis in childhood) with depression and alcohol use disorder were very similar, suggesting that any ascertainment bias between the two groups was limited. Likewise, in the PGS sensitivity analyses, where individuals with the disorder associated with the PGS being analyzed were excluded, the genetic differences among the groups demonstrated the same patterns observed in the full sample. This further reinforces the conclusion that comorbid psychiatric disorders did not have a strong influence on the observed genetic differences among the ADHD subgroups.

In summary, our results are population-based and thus reflect the genetic architecture of ADHD and comorbidity patterns across ADHD subgroups in the Danish population. Persistent ADHD demonstrated the highest load of ADHD risk variants, whereas late-diagnosed ADHD was less enriched for variants associated with hyperactivity and inattention and did not, unlike childhood and persistent ADHD, demonstrate an increased burden of rPTVs compared with controls. This suggests that genetic factors might explain in part why some individuals are diagnosed late as adults. The comorbidity of depression and alcohol use disorder was highest in the late-diagnosed group. If this was only due to age differences among groups, we would not expect the genetic correlations to differ, but we found a higher genetic overlap of these disorders with late-diagnosed ADHD compared with childhood ADHD. This suggests that the higher comorbidity among individuals with late-diagnosed ADHD is not only due to those individuals being older but also due to a higher genetic risk. Conversely, the childhood ADHD group demonstrated a higher genetic overlap with autism and a higher burden of rPTVs in highly constrained genes than the other two groups. Overall, we have identified genetic heterogeneity among ADHD subgroups, and our findings suggest that genetic factors influence the time of first ADHD diagnosis, persistence of ADHD into adulthood and comorbidity patterns.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01143-7.

Received: 2 August 2021; Accepted: 22 June 2022;
Published online: 4 August 2022

References

1. Faraone, S. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers* **1**, 15020 (2015).
2. Franke, B. et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol. Psychiatry* **17**, 960–987 (2012).
3. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* **24**, 562–575 (2019).
4. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
5. Faraone, S. V., Biederman, J. & Mick, E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychol. Med.* **36**, 159–165 (2006).
6. Fatseas, M. et al. Addiction severity pattern associated with adult and childhood attention deficit hyperactivity disorder (ADHD) in patients with addictions. *Psychiatry Res* **246**, 656–662 (2016).
7. Agnew-Blais, J. C. et al. Young adult mental health and functional outcomes among individuals with remitted, persistent and late-onset ADHD. *Br. J. Psychiatry* **213**, 526–534 (2018).
8. Ilbegi, S. et al. Substance use and nicotine dependence in persistent, remittent, and late-onset ADHD: a 10-year longitudinal study from childhood to young adulthood. *J. Neurodev. Disord.* **10**, 42 (2018).
9. Yoshimasu, K. et al. Adults with persistent ADHD: gender and psychiatric comorbidities - a population-based longitudinal study. *J. Atten. Disord.* **22**, 535–546 (2018).
10. Kim, K. M. et al. Psychopathological, temperamental, and characteristic factors in adults with remaining childhood attention-deficit hyperactivity symptoms. *Int J. Psychiatry Clin. Pract.* **21**, 236–241 (2017).
11. Tandon, M., Tillman, R., Agrawal, A. & Luby, J. Trajectories of ADHD severity over 10 years from childhood into adulthood. *Atten. Defic. Hyperact Disord.* **8**, 121–130 (2016).
12. Kan, K. J. et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *J. Am. Acad. Child Adolesc. Psychiatry* **52**, 12–25 (2013).
13. Larsson, H. et al. Genetic and environmental influences on adult attention deficit hyperactivity disorder symptoms: a large Swedish population-based study of twins. *Psychol. Med.* **43**, 197–207 (2013).
14. Brikell, I., Kuja-Halkola, R. & Larsson, H. Heritability of attention-deficit hyperactivity disorder in adults. *Am. J. Med. Genet. B Neuropsychiatr. Genet.***168**, 406–413 (2015).
15. Pingault, J. B. et al. Genetic and environmental influences on the developmental course of attention-deficit/hyperactivity disorder symptoms from childhood to adolescence. *JAMA Psychiatry* **72**, 651–658 (2015).
16. Riglin, L. et al. Association of genetic risk variants with attention-deficit/hyperactivity disorder trajectories in the general population. *JAMA Psychiatry* **73**, 1285–1292 (2016).
17. Asherson, P. & Agnew-Blais, J. Annual research review: does late-onset attention-deficit/hyperactivity disorder exist? *J. Child Psychol. Psychiatry* **60**, 333–352 (2019).
18. Agnew-Blais, J. C. et al. Polygenic risk and the course of attention-deficit/hyperactivity disorder from childhood to young adulthood: findings from a nationally-representative cohort. *J. Am. Acad. Child Adolesc. Psychiatry* **60**, 1147–1156 (2021).
19. Rovira, P. et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **45**, 1617–1626 (2020).
20. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
21. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
22. Zayats, T. QIMR Group, SickKids Group, TEDS Group, CATSS Group, MoBa Group, NTR Group, ALSPAC Group, ATGU Group. Genome-wide association meta-analysis of ADHD symptomatology: inattention and hyperactivity/impulsivity. *Eur. Neuropsychopharmacol.* **29**, S1190 (2019).
23. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
24. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
25. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
26. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
27. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
28. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) & OCD Collaborative Genetics Association Studies (OCGAS) Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
29. Johnson, E. C. et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
30. Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
31. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
32. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
33. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
34. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
35. Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
36. Satterstrom, F. K. et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961–1965 (2019).

37. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
38. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).
39. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482.e11 (2019).
40. Enard, W. et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
41. Schreiweis, C. et al. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl Acad. Sci. USA* **111**, 14253–14258 (2014).
42. Vernes, S. C. et al. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* **7**, e1002145 (2011).
43. Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A. & Buitelaar, J. K. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Eur. Child Adolesc. Psychiatry* **19**, 281–295 (2010).
44. Skirrow, C. & Asherson, P. Emotional lability, comorbidity and impairment in adults with attention-deficit hyperactivity disorder. *J. Affect. Disord.* **147**, 80–86 (2013).
45. Chang, Z., D'Onofrio, B. M., Quinn, P. D., Lichtenstein, P. & Larsson, H. Medication for attention-deficit/hyperactivity disorder and risk for depression: a nationwide longitudinal cohort study. *Biol. Psychiatry* **80**, 916–922 (2016).
46. Biederman, J., Petty, C. R., Evans, M., Small, J. & Faraone, S. V. How persistent is ADHD? A controlled 10-year follow-up study of boys with ADHD. *Psychiatry Res.* **177**, 299–304 (2010).
47. Van Veen, M. M., Kooij, J. J., Boonstra, A. M., Gordijn, M. C. & Van Someren, E. J. Delayed circadian rhythm in adults with attention-deficit/hyperactivity disorder and chronic sleep-onset insomnia. *Biol. Psychiatry* **67**, 1091–1096 (2010).

Methods

Sample characteristics. Individuals included in the study were identified from a nationwide population-based case–cohort sample established by iPSYCH[20] comprising 133,296 genotyped individuals, of whom 91,378 had been diagnosed with at least one of six mental disorders (schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, ADHD and anorexia) and the remainder were population-based controls. Samples were selected from a baseline birth cohort comprising all singletons born in Denmark between 1 May 1981 and 31 December 2008 who had a known mother (99.9% of all individuals born in Denmark since 1970 have a known mother[48]) and were resident in Denmark on their first birthday (n = 1,472,762). We included all individuals in the cohort diagnosed with ADHD by psychiatrists according to the ICD-10 criteria (F90.0 diagnosis code) identified in the Danish Psychiatric Central Research Register[49]. See Supplementary Information, Note 2, for information on cases potentially missed by the diagnostic system.

The ICD-10 diagnosis code F90.0 describes a disorder characterized by early onset, usually in the first 5 years of life, with hyperactivity and decreased attention. According to the current diagnostic criteria, individuals diagnosed with ADHD as adults should be able to describe ADHD symptoms in childhood retrospectively. Diagnoses were given in 2016 or earlier for individuals at least 1 year old. Individuals were divided into three groups depending on age at first diagnosis: (1) childhood ADHD, defined as individuals diagnosed with ADHD and less than 18 years of age in 2016, or individuals older than 18 years by the end of follow-up (2016) who did not receive another ADHD diagnosis after the age of 18 years (n = 15,338 before quality control (QC)); (2) persistent ADHD, defined as individuals diagnosed with ADHD in childhood (before 18 years of age) and again as adults (after 18 years of age) (n = 1,709 before QC); and (3) late-diagnosed ADHD, defined as individuals diagnosed with ADHD as adults (after 18 years of age) (n = 7,815 before QC). Controls were randomly selected from the same nationwide birth cohort and were individuals not diagnosed with ADHD (n = 45,398 before QC).

The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark.

Differences in the female/male ratio between ADHD subgroups were tested by chi-squared test. Information about comorbidity with other major psychiatric disorders was obtained from the Danish Psychiatric Central Research Register[49]: autism spectrum disorder (ICD-10 diagnosis code F84), schizophrenia (ICD-10 diagnosis code F20), bipolar disorder (ICD-10 diagnosis codes F30–F31), major depressive disorder (ICD-10 diagnosis codes F32–F33), OCD (ICD-10 diagnosis code F42), anorexia (ICD-10 diagnosis codes F50), alcohol use disorder (ICD-10 diagnosis code F10.1–F10.9) and cannabis use disorder (ICD-10 diagnosis code F12.1–F12.9).

Genotyping and QC. The study subjects were linked to their biological samples (dried blood spots) stored in the Danish Newborn Screening Biobank[50] through the personal identification number[51] assigned to all individuals with residence in Denmark. DNA was extracted from the dried blood spots and whole-genome amplified in triplicate[52,53]. Genotyping was done in two rounds. In round one (iPSYCH1), 79,492 individuals were genotyped by using Illumina Bead Arrays (PsychChip). In round two (iPSYCYH2), 53,804 individuals were genotyped using the Illumina Global Screening array. iPSYCH1 genotypes were a result of merging call sets from two different calling algorithms, GenCall (v.1.6.2.2)[54] and Birdseed (v.1.6)[55], which were used to call genotypes with a minor allele frequency (MAF) > 0.01. iPSYCH2 genotypes were called by using GenTrain v.3.

All downstream analyses were performed on our secure server (GenomeDK; http://genome.au.dk). Stringent QC was applied to the full iPSYCH sample. Only individuals with a high call rate (>0.95) were included, and only genotypes with a high call rate (>0.98), no strong deviation from Hardy–Weinberg equilibrium (HWE) ($P > 1 \times 10^{-6}$ for controls or $P > 1 \times 10^{-10}$ for cases) and low heterozygosity rates ($|F_{het}| < 0.2$) were included. Genotypes were phased and imputed using the Haplotype Reference Consortium[56] data as a reference panel, and prephasing used Eagle v.2.3.5 (ref. [57]) and imputation with Minimac3 (ref. [58]).

Relatedness and population stratification were evaluated for ADHD cases and controls using merged data from iPSYCH1 and iPSYCH2 and a set of high-quality markers (best-guess genotypes with MAF > 0.05, HWE $P > 1 \times 10^{-9}$, SNP call rate >0.99, imputation info score > 0.9), which were pruned for linkage disequilibrium (LD) ($r^2 < 0.1$), resulting in a set of 37,986 pruned variants (variants located in long-range LD regions defined by Price et al.[59] were excluded). Genetic relatedness was estimated using PLINK v.1.9 (refs. [60,61]) to identify first- and second-degree relatives ($\hat{\pi} > 0.2$), and one individual was excluded from each related pair (cases kept preferably over controls). Genetic outliers were excluded based on principal component analyses (PCA) using EIGENSOFT v.6.1.3 (refs. [62,63]). After the first PCA, the principal components (PCs) from a set of individuals born in Denmark for three generations were used as a reference to generate an ellipsoid based on information from the first six PCs and their standard deviations (eight standard deviations were used). Those who fell outside this ellipsoid were removed. The PCA was repeated, and the new PCs were used as covariates to correct for any remaining population substructure in subsequent analyses. After QC, the numbers of included individuals were: (1) childhood ADHD, n = 14,878; (2) persistent ADHD, n = 1,473; and (3) late-diagnosed ADHD, n = 7,188. The control group contained 38,3030 individuals.

GWAS. A flow chart of the genetic analyses performed in this study can be found in Supplementary Fig. 7. We performed GWAS for each ADHD subgroup against a common set of controls (n = 38,3030). We used merged iPSYCH1 and iPSYCH2 high-quality best-guess genotypes (MAF > 0.01, imputation info score > 0.80, missing rate <1%; n = 5,826,893 variants) and tested for association using logistic regression in PLINK v.1.9 (ref. [60]) with the following covariates: ten PCs from PCA, sex and a covariate for genotyping round (iPSYCH1 or iPSYCH2).

We tested whether the effect size of the genome-wide significant locus in late-diagnosed ADHD was significantly higher than the effect size observed for the other groups using a z test and effect sizes from association analyses based on nonoverlapping samples. The numbers of nonoverlapping controls were: childhood controls, 24,443; persistent controls, 2,289; and late-diagnosed controls, 11,571.

SNP heritability and genetic correlations of ADHD subgroups. The h^2_{SNP} value was estimated for each group against the same controls (n = 38,3030) using univariate GREML implemented in GCTA[21] (and the same covariates as used in the GWAS) and a population prevalence of 0.05 for childhood ADHD[1,64], 0.03 for persistent ADHD[3] and 0.03 for late-diagnosed ADHD. To test for differences in h^2_{SNP} estimates among groups, we also derived estimates by using nonshared controls (control numbers: childhood controls, 24,443; persistent controls, 2,289; and late-diagnosed controls, 11,571). Difference in h^2_{SNP} were determined by z test. In addition, h^2_{SNP} values in the subgroups were estimated over a range of population prevalence values ranging from 1% to 15%.

Genetic correlations between ADHD subgroups were calculated using bivariate GREML analysis in GCTA and nonshared controls.

PGS analyses of ADHD and other phenotypes. The PGSs for ADHD were generated by a fivefold cross-validation approach, similar to that described previously[4]. In short, the sample was split into five groups, aiming for equal numbers of ADHD cases and controls within each group. We then conducted GWAS using four of the five groups to derive effect sizes with respect to ADHD risk. These effect sizes were then used to estimate the PGS for the remaining target group. Thus, the training data were independent of the target data. This procedure was repeated five times until PGSs had been estimated in all target groups. Indels and variants in the extended major histocompatibility complex region (chromosome 6: 25–34 Mb) were also removed. Only independent variants were used to generate the score, and clumping of the training data was performed on the summary statistics by employing PLINK and the flags -clump-p1 1, -clump-p2 1, -clump-r2 0.1 and -clump-kb 500. PGS was estimated for each individual in the target sample by using a range of P value thresholds in the training data (5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.05, 0.1, 0.2, 0.5 and 1.0), multiplying the natural logarithm of the OR of each variant by the allele count of each variant. The whole-genome PGS was obtained by summing values over variants for each individual. The PGSs were standardized for each of the five target sample groups (subtracting the mean and dividing by the s.d.). The scores from the five target groups were then pooled at each threshold and tested for association with general ADHD (that is, all cases versus controls), and the P value threshold for the scores explaining the maximum variance (estimated by Nagelkerke's R^2) in the target data of general ADHD (that is, all ADHD cases versus controls) was used to test for differences in ADHD-PGS load across ADHD subgroups ($P_{threshold} < 0.1$). As our ADHD cohort was population-based, including all individuals with ADHD born in Denmark between 1981 and 2008 and diagnosed before or in 2016, the generated PGS represented the general liability to diagnosed ADHD (because the cross-validation approach was based on all population-based cases). We stress that the PGS only reflects the general liability with respect to diagnosed ADHD as some cases potentially might be missed by the diagnostic system (Supplementary Information, Note 2).

PGSs for ADHD symptoms and 13 other phenotypes (schizophrenia, autism, bipolar disorder, alcohol use disorder, cannabis use disorder, OCD, anorexia, depression, educational years, mother's age at death, BMI, age at first birth and insomnia) were generated using summary statistics from large GWAS of the phenotypes (see Supplementary Table 9 for references) and the approach described above (without fivefold cross-validation). The data on ADHD symptoms (inattention and hyperactivity/impulsivity) came from a genome-wide association meta-analysis on up to 43,117 children and adolescents[22]. In the PGS analyses, P value thresholds in the training GWAS that captured most variance (estimated by Nagelkerke's R^2) in the target data were used as thresholds for analyses of the PGS load in the subgroups (threshold information is provided in Supplementary Table 9).

We tested for differences in PGS load among ADHD subgroups by multiple regression in R v.3.6.0 and the package 'multcomp', with ADHD coded as four factors: controls, childhood, adulthood and persistent ADHD; we also included covariates to correct for genotyping round (iPSYCH1 or iPSYCH2), sex and ten ancestry PCs. Correction for multiple testing was done separately for the following three analyses: (1) PGS-ADHD load among subgroups correcting for three pairwise comparisons; (2) PGS load for ADHD symptoms (inattention and

hyperactivity) correcting for six pairwise comparisons; and (3) PGS load for 13 other phenotypes correcting for 39 pairwise comparisons.

We also performed two sensitivity PGS analyses. (1) We evaluated whether the differences in the PGS load of four major psychiatric disorders (depression, schizophrenia, bipolar disorder and autism) could be caused primarily by the presence of individuals with comorbidity. We excluded all individuals in the target sample with the diagnosis being analyzed; that is, all depression cases were excluded in the analysis of depression-PGS (sample sizes are given in Supplementary Table 10b). (2) We evaluated the potential genetic heterogeneity in the childhood group caused by age. This was done by splitting the childhood group into two groups, those younger than 18 years of age ($n=8,664$) and those older than 18 years of age ($n=6,214$) by the end of follow-up. We then repeated the PGS analyses including the two childhood ADHD groups (persistent and late-diagnosed ADHD).

PGS analysis in the Spanish cohort[19] consisting of 453 individuals with childhood ADHD, 270 with persistent ADHD, 889 with late-diagnosed ADHD and 3,440 controls was followed the approach described above (see Supplementary Information, Note 3, for exclusion criteria).

Genetic correlations with ADHD symptoms and other phenotypes. Genetic correlations (r_g) of the three ADHD subgroups with ADHD symptoms[22] and the 13 phenotypes listed above were calculated using summary statistics from GWAS and LD score regression[65]. No sample overlap and no population stratification were assumed when calculating r_g with ADHD symptoms; therefore, the intercept was restricted by setting the single-trait intercepts to 1 and cross-trait intercepts to 0.

Statistical differences between two r_g estimates were calculated by the block jackknife method implemented in the LD score regression software[65,66]. The genome was divided into 200 blocks, and jackknife deleted values were calculated by excluding one block at a time. The jackknife deleted values were then used to calculate corresponding jackknife pseudovalues. Based on the mean and variance of the jackknife pseudovalues, z scores and corresponding P values were computed, testing the null hypothesis that the difference between the genetic correlations was equal to zero. A z test was used to determine whether the genetic correlation differed from 1.

Correction for multiple pairwise comparisons was done separately for the following two evaluations: (1) differences in r_g among ADHD subgroups with ADHD symptoms correcting for six pairwise comparisons; and (2) r_g differences among ADHD subgroups with 13 other phenotypes correcting for 39 pairwise comparisons.

Burden of rare variants in ADHD subgroups. WES data were available for a subset of the iPSYCH samples. It has previously been shown that WES of DNA from dried blood spots results in high-quality data[67]. DNA was extracted from dried blood spot samples of the study subjects and whole-genome-amplified in triplicate[52,53]. The coding regions of the genome were extracted using an Illumina Nextera capture kit, and sequencing was performed in multiple waves (pilot 1, wave 1, wave 2 and wave 3) with an Illumina HiSeq platform at the Genomics Platform of the Broad Institute.

Part of the data (pilot 1, wave 1 and wave 2) were also included in a recent study by Satterstrom et al.[36], in which the authors examined the overall burden of rPTVs in ADHD; the same QC procedure was used in this study, including all data (pilot 1, wave 1, wave 2 and wave 3). In short, the raw sequencing data were aligned to the reference genome Hg19 using BWA[68]. Calling of genotypes followed the best practice recommended by the Genome Analysis Toolkit[69] (GATK) v.3.4. Most QC steps were performed with Hail 0.1 (Hail Team, https://github.com/hail-is/hail). All variants annotated to American College of Medical Genetics[70] genes were removed owing to Danish legislation. Samples were removed if they lacked complete phenotype information, if there were inconsistencies of the imputed sex with the reported sex, if they were duplicates or genetic outliers identified by PCA, if they had an estimated level of contamination greater than 5% or if they had an estimated level of chimeric reads greater than 5%.

Only autosomal genotypes were included in our analyses. Genotypes were removed if they did not pass GATK variant quality score recalibration (VQSR) or had a read depth <10 or >1,000. Homozygous alleles were removed if they had reference calls with genotype quality less than 25, homozygous alternate alleles with PL(HomRef) (that is, the phred-scaled likelihood of being a homozygous reference) <25 or <90% reads supporting an alternate allele. Heterozygous alleles were removed if they had PL(HomRef) <25 or <25% reads supporting the alternate allele, <90% informative reads (that is, the number of reads supporting the reference allele plus the number of reads supporting an alternate allele <90% of the read depth) or a probability of the allele balance calculated from a binomial distribution centered on 0.5 of less than 1×10^{-9}. After the application of these basic genotype filters, variants with a call rate <90% were removed, then samples with a call rate <95% and variants with a call rate <95% were removed. Between the sample call rate filter and the final variant call rate filter, one of each pair of related samples was removed, defining relatedness as individuals with a pairwise $\hat{\pi} \geq 0.2$. After QC, the numbers of individuals were: childhood ADHD, $n=4,987$; persistent ADHD, $n=748$; late-diagnosed ADHD, $n=1,915$; and controls, $n=8,649$.

Following QC, variants were annotated using SnpEff[71] v.4.3t. The variants were also annotated with the gnomAD[72] exomes r2.1.1 database using SnpSift[71] v.4.3t. Variants were only included if they were located in consensus high-confidence coding regions with a high read depth in both the iPSYCH data and the gnomAD data set (80% of the samples in both data sets had at least 10× sequencing coverage in the region). Variants were defined as rPTVs if they were annotated as having large effects on gene function (nonsense variant, frameshift, splice site) and were rare in the sample, defined as having an allele count ≤5 across the combined counts in iPSYCH ($n=16,299$) and non-Finnish Europeans in the nonpsychiatric gnomAD exome database ($n=44,779$).

The burden of rPTVs and rSYNs in ADHD subgroups and controls was tested in (1) highly constrained genes ($n=3,488$), defined as genes highly intolerant to loss-of-function mutations with pLI score >0.9 (ref. [37]), and (2) de novo highly constrained genes ($n=241$), defined as highly constrained genes that overlap with another set of genes ($n=285$) previously found to be enriched with de novo mutations in individuals with neurodevelopmental disorders[38]; (3) for comparison, we also tested a set of 9,662 evolutionarily less constrained genes with pLI score <0.1. Enrichment in rPTVs and rSYNs variants was tested by multiple regression with the three ADHD groups and controls included in the same regression model (using R v.3.6.0 and the R packages foreign, nnet, ggplot2 and reshape2). The outcome (dependent) variable was rPTV count, and the predictor (independent) variables were ADHD given as categorical variables with multiple factors (childhood, persistent and late-diagnosed ADHD and controls, with controls as the reference factor) and relevant covariates (with the regression model coded in R as follows: rPTV counts ~ ADHD (controls|childhood|persistent|late-diagnosed) + covariate 1 + covariate 2 + … + covariate n). ADHD was coded as the independent variable rather than the dependent variable to enable pairwise comparisons between ADHD subtypes in the same analysis. The covariates were birth year, sex, the first ten PCs from ancestry PCA, the number of rSYNs, the percentage of target with coverage >20x, the mean read depth at sites within the exome target passing variant quality score recalibration, the total number of variants and the sequencing wave. Testing for enrichment of rPTVs in ADHD subgroups compared with controls was corrected for nine tests (three groups × three gene sets, that is, a new threshold of $P=0.006$), and testing for differences between groups was corrected for nine pairwise comparisons (three gene sets × three pairwise comparisons for each set).

Statistics and reproducibility. GWAS of ADHD subgroups was performed by logistic regression using an additive model of imputed dosage to estimate the association of the effect allele with childhood, persistent and late-diagnosed ADHD. Differences in the PGS load among ADHD subgroups were tested by multiple regression. Genetic correlations were calculated using LD score regression, and statistical differences between two r_g estimates were calculated using the block jackknife method. The burden of rare variants in the three ADHD subgroups was analyzed using multiple logistic regression, with the three ADHD groups and controls included in the same regression model. All analyses were corrected using relevant covariates, and Bonferroni correction was applied when appropriate (see Methods section for details). No statistical method was used to determine sample size. The sample size was fixed, as we initially (that is, before QC) included all individuals born in Denmark between 1981 and 2008 who had been diagnosed with ADHD in 2016 or before.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Summary statistics from GWAS of childhood, persistent and late-diagnosed ADHD are available at the iPSYCH website (https://ipsych.dk/en/research/downloads/). All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secure Danish server (GenomeDK; https://genome.au.dk) as the data are protected by Danish legislation. For data access, please contact: D.D. or A.D.B. (anders@biomed.au.dk). Correspondence and requests for materials should be addressed to D.D. (ditte@biomed.au.dk).

Code availability

No previously unreported custom computer code or algorithm were used to generate results, all software used in the study are publicly available from the Internet as described in Methods and Reporting Summary.

References

48. Pedersen, C. B., Gotzsche, H., Moller, J. O. & Mortensen, P. B. The Danish civil registration system. A cohort of eight million persons. *Dan. Med. Bull.* **53**, 441–449 (2006).
49. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand. J. Public Health* **39**, 54–57 (2011).
50. Nørgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish Newborn Screening Biobank. *J. Inherit. Metab. Dis.* **30**, 530–536 (2007).

51. Pedersen, C. B. The Danish civil registration system. *Scand. J. Public Health* **39**, 22–25 (2011).
52. Borglum, A. D. et al. Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol. Psychiatry* **19**, 325–333 (2014).
53. Hollegaard, M. V. et al. Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet.* **12**, 58 (2011).
54. *Illumina GenCall Data Analysis Software* Illumina Tech Note (Illumina, 2005).
55. Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
56. Iglesias, A. I. et al. Haplotype reference consortium panel: practical implications of imputations with large reference panels. *Hum. Mutat.* **38**, 1025–1032 (2017).
57. Loh, P. R. et al. Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
58. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
59. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008); author reply, 135–139.
60. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
61. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
62. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
63. Galinsky, K. J. et al. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
64. Dalsgaard, S. et al. Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence. *JAMA Psychiatry* **77**, 155–164 (2020).
65. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
66. Hubel, C. et al. Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *Am. J. Med. Genet. B Neuropsychiatr. Genet.***180**, 428–438 (2019).
67. Poulsen, J. B. et al. High-quality exome sequencing of whole-genome amplified neonatal dried blood spot DNA. *PLoS ONE* **11**, e0153253 (2016).
68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
69. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
70. Kalia, S. S. et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* **19**, 249–255 (2017).
71. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
72. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

Acknowledgements

D.D. was supported by the Novo Nordisk Foundation (NNF20OC0065561) and the Lundbeck Foundation (R344-2020-1060). The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724 and R248-2017-2003), the EU FP7 Program (grant number 602805, 'Aggressotype') and H2020 Program (grant number 667302, 'CoCA'), National Institute of Mental Health (1U01MH109514-01 to A.D.B.) and the universities and university hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data at the GenomeDK high-performance computer facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). M.S.A. is a recipient of a Juan de la Cierva Incorporación contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I). The research leading to these results has received funding from the Instituto de Salud Carlos III (PI17/00289, PI18/01788, P19/01224 and PI20/00041) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461) and was cofinanced by the European Regional Development Fund. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the families in Norway who have taken part in this ongoing cohort study. L.V.R. is a recipient of a predoctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285). M.R. was a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023). T.Z. is funded by R37MH107649-07S1 and by the Research Council of Norway (grant number 288083).

Author contributions

V.M.R, J.D. and L.V.-R. carried out the analysis. J.G., T. Z., J.A.R.-Q., F.K.S., M.S.A., J.B.-G., M.B.-H., T.D.A., A.R., M.J.D., B.M.N., M.N., T.W., O.M., D.M.H. and P.B.M. performed sample and/or data provision and processing. D.D. and V.M.R. wrote the manuscript. D.D., V.M.R., F.K.S., A.D.B. and M.R. revised the manuscript. D.D. and V.M.R. were responsible for the study design. All authors contributed with critical revisions of the manuscript.

Competing interests

D.D. has received a speaker fee from Takeda. J.A.R.-Q. has been on the speaker's bureau and/or has acted as a consultant for Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice and Rubiand Raffo in the past 3 years. He has also received travel awards (air tickets and hotel) for taking part in psychiatric meetings from Janssen-Cilag, RubiShire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him has received unrestricted educational and research support from the following companies in the past 3 years: Janssen-Cilag, Shire, Oryzon, Roche, Psious and Rubió. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01143-7.

Correspondence and requests for materials should be addressed to Ditte Demontis.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Archival Report

Biological Psychiatry

Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors

Niamh Mullins, JooEun Kang, Adrian I. Campos, Jonathan R.I. Coleman, Alexis C. Edwards, Hanga Galfalvy, Daniel F. Levey, Adriana Lori, Andrey Shabalin, Anna Starnawska, Mei-Hsin Su, Hunna J. Watson, Mark Adams, Swapnil Awasthi, Michael Gandal, Jonathan D. Hafferty, Akitoyo Hishimoto, Minsoo Kim, Satoshi Okazaki, Ikuo Otsuka, Stephan Ripke, Erin B. Ware, Andrew W. Bergen, Wade H. Berrettini, Martin Bohus, Harry Brandt, Xiao Chang, Wei J. Chen, Hsi-Chung Chen, Steven Crawford, Scott Crow, Emily DiBlasi, Philibert Duriez, Fernando Fernández-Aranda, Manfred M. Fichter, Steven Gallinger, Stephen J. Glatt, Philip Gorwood, Yiran Guo, Hakon Hakonarson, Katherine A. Halmi, Hai-Gwo Hwu, Sonia Jain, Stéphane Jamain, Susana Jiménez-Murcia, Craig Johnson, Allan S. Kaplan, Walter H. Kaye, Pamela K. Keel, James L. Kennedy, Kelly L. Klump, Dong Li, Shih-Cheng Liao, Klaus Lieb, Lisa Lilenfeld, Chih-Min Liu, Pierre J. Magistretti, Christian R. Marshall, James E. Mitchell, Eric T. Monson, Richard M. Myers, Dalila Pinto, Abigail Powers, Nicolas Ramoz, Stefan Roepke, Vsevolod Rozanov, Stephen W. Scherer, Christian Schmahl, Marcus Sokolowski, Michael Strober, Laura M. Thornton, Janet Treasure, Ming T. Tsuang, Stephanie H. Witt, D. Blake Woodside, Zeynep Yilmaz, Lea Zillich, Rolf Adolfsson, Ingrid Agartz, Tracy M. Air, Martin Alda, Lars Alfredsson, Ole A. Andreassen, Adebayo Anjorin, Vivek Appadurai, María Soler Artigas, Sandra Van der Auwera, M. Helena Azevedo, Nicholas Bass, Claiton H.D. Bau, Bernhard T. Baune, Frank Bellivier, Klaus Berger, Joanna M. Biernacka, Tim B. Bigdeli, Elisabeth B. Binder, Michael Boehnke, Marco P. Boks, Rosa Bosch, David L. Braff, Richard Bryant, Monika Budde, Enda M. Byrne, Wiepke Cahn, Miguel Casas, Enrique Castelao, Jorge A. Cervilla, Boris Chaumette, Sven Cichon, Aiden Corvin, Nicholas Craddock, David Craig, Franziska Degenhardt, Srdjan Djurovic, Howard J. Edenberg, Ayman H. Fanous, Jerome C. Foo, Andreas J. Forstner, Mark Frye, Janice M. Fullerton, Justine M. Gatt, Pablo V. Gejman, Ina Giegling, Hans J. Grabe, Melissa J. Green, Eugenio H. Grevet, Maria Grigoroiu-Serbanescu, Blanca Gutierrez, Jose Guzman-Parra, Steven P. Hamilton, Marian L. Hamshere, Annette Hartmann, Joanna Hauser, Stefanie Heilmann-Heimbach, Per Hoffmann, Marcus Ising, Ian Jones, Lisa A. Jones, Lina Jonsson, René S. Kahn, John R. Kelsoe, Kenneth S. Kendler, Stefan Kloiber, Karestan C. Koenen, Manolis Kogevinas, Bettina Konte, Marie-Odile Krebs, Mikael Landén, Jacob Lawrence, Marion Leboyer, Phil H. Lee, Douglas F. Levinson, Calwing Liao, Jolanta Lissowska, Susanne Lucae, Fermin Mayoral, Susan L. McElroy, Patrick McGrath, Peter McGuffin, Andrew McQuillin, Sarah E. Medland, Divya Mehta, Ingrid Melle, Yuri Milaneschi, Philip B. Mitchell, Esther Molina, Gunnar Morken, Preben Bo Mortensen, Bertram Müller-Myhsok, Caroline Nievergelt, Vishwajit Nimgaonkar, Markus M. Nöthen, Michael C. O'Donovan, Roel A. Ophoff, Michael J. Owen, Carlos Pato, Michele T. Pato, Brenda W.J.H. Penninx, Jonathan Pimm, Giorgio Pistis, James B. Potash, Robert A. Power, Martin Preisig, Digby Quested, Josep Antoni Ramos-Quiroga, Andreas Reif, Marta Ribasés, Vanesa Richarte, Marcella Rietschel, Margarita Rivera, Andrea Roberts, Gloria Roberts, Guy A. Rouleau, Diego L. Rovaris, Dan Rujescu, Cristina Sánchez-Mora, Alan R. Sanders, Peter R. Schofield, Thomas G. Schulze, Laura J. Scott, Alessandro Serretti, Jianxin Shi, Stanley I. Shyn, Lea Sirignano, Pamela Sklar, Olav B. Smeland, Jordan W. Ṣmoller, Edmund J.S. Sonuga-Barke, Gianfranco Spalletta, John S. Strauss, Beata Świątkowska, Maciej Trzaskowski, Gustavo Turecki, Laura Vilar-Ribó, John B. Vincent, Henry Völzke,

ISSN: 0006-3223

Biological Psychiatry February 1, 2022; 91:313–327 www.sobp.org/journal

James T.R. Walters, Cynthia Shannon Weickert, Thomas W. Weickert, Myrna M. Weissman, Leanne M. Williams, Naomi R. Wray, Clement C. Zai, Allison E. Ashley-Koch, Jean C. Beckham, Elizabeth R. Hauser, Michael A. Hauser, Nathan A. Kimbrel, Jennifer H. Lindquist, Benjamin McMahon, David W. Oslin, Xuejun Qin, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Eating Disorders Working Group of the Psychiatric Genomics Consortium, German Borderline Genomics Consortium, MVP Suicide Exemplar Workgroup, VA Million Veteran Program, Esben Agerbo, Anders D. Børglum, Gerome Breen, Annette Erlangsen, Tõnu Esko, Joel Gelernter, David M. Hougaard, Ronald C. Kessler, Henry R. Kranzler, Qingqin S. Li, Nicholas G. Martin, Andrew M. McIntosh, Ole Mors, Merete Nordentoft, Catherine M. Olsen, David Porteous, Robert J. Ursano, Danuta Wasserman, Thomas Werge, David C. Whiteman, Cynthia M. Bulik, Hilary Coon, Ditte Demontis, Anna R. Docherty, Po-Hsiu Kuo, Cathryn M. Lewis, J. John Mann, Miguel E. Rentería, Daniel J. Smith, Eli A. Stahl, Murray B. Stein, Fabian Streit, Virginia Willour, and Douglas M. Ruderfer

ABSTRACT

BACKGROUND: Suicide is a leading cause of death worldwide, and nonfatal suicide attempts, which occur far more frequently, are a major source of disability and social and economic burden. Both have substantial genetic etiology, which is partially shared and partially distinct from that of related psychiatric disorders.

METHODS: We conducted a genome-wide association study (GWAS) of 29,782 suicide attempt (SA) cases and 519,961 controls in the International Suicide Genetics Consortium (ISGC). The GWAS of SA was conditioned on psychiatric disorders using GWAS summary statistics via multitrait-based conditional and joint analysis, to remove genetic effects on SA mediated by psychiatric disorders. We investigated the shared and divergent genetic architectures of SA, psychiatric disorders, and other known risk factors.

RESULTS: Two loci reached genome-wide significance for SA: the major histocompatibility complex and an intergenic locus on chromosome 7, the latter of which remained associated with SA after conditioning on psychiatric disorders and replicated in an independent cohort from the Million Veteran Program. This locus has been implicated in risk-taking behavior, smoking, and insomnia. SA showed strong genetic correlation with psychiatric disorders, particularly major depression, and also with smoking, pain, risk-taking behavior, sleep disturbances, lower educational attainment, reproductive traits, lower socioeconomic status, and poorer general health. After conditioning on psychiatric disorders, the genetic correlations between SA and psychiatric disorders decreased, whereas those with nonpsychiatric traits remained largely unchanged.

CONCLUSIONS: Our results identify a risk locus that contributes more strongly to SA than other phenotypes and suggest a shared underlying biology between SA and known risk factors that is not mediated by psychiatric disorders.

https://doi.org/10.1016/j.biopsych.2021.05.029

Suicide is a worldwide public health problem, accounting for almost 800,000 deaths per year (1). Nonfatal suicide attempt (SA), defined as self-injurious behavior with the intent to die, has been estimated to occur over 20 times more frequently and is a major source of disability, reduced quality of life, and social and economic burden (1,2). The lifetime prevalence of SA in adults ranges from 0.5% to 5% worldwide (3). There are several well-established comorbidities and risk factors for SA, with psychiatric illness having the strongest effect on lifetime suicide rates (4,5). However, the vast majority of patients with psychiatric disorders never attempt suicide (6–8). Other major risk factors for SA include prior self-injurious thoughts and behaviors (9), physical illness or disability (10,11), sleep disorders (12–15), family history of psychiatric disorders (16), substance abuse (17), smoking (18–20), impulsivity (21) and social factors including childhood maltreatment (21), isolation (22), and stressful life events (23).

Both suicide and SA are heritable, with estimates from genetic epidemiology studies ranging from 17% to 55% (24–26). Several genome-wide association studies (GWASs) of SA have reported significant single nucleotide polymorphism (SNP)–heritability estimates of ~4%, indicating an underlying polygenic architecture (27–31). Using polygenic risk scoring or genetic correlation analyses, these studies have also demonstrated shared genetic etiology between SA and psychiatric disorders, with major depressive disorder (MDD) showing the largest genetic overlap (28,29,31). This genetic overlap, along with the high prevalence of MDD in the population (32), make it

a particularly salient risk factor. Importantly, genetic epidemiology studies have consistently indicated a genetic component of SA that is partially distinct from that of psychiatric disorders (25). Consistent with this, one GWAS of SA that covaried for cases' psychiatric diagnoses estimated a SNP-heritability of 1.9% (27).

With few genetic samples collected specifically for SA, studies often rely on individuals ascertained for psychiatric disorders. For example, a large GWAS of SA included over 6500 cases from clinical cohorts of MDD, bipolar disorder (BIP), and schizophrenia (SCZ) cases, within the Psychiatric Genomics Consortium (PGC) (31). In an "SA within psychiatric diagnosis" study design, SA cases were compared with cases of the same psychiatric disorder without SA, in order to disentangle the genetic etiology of SA and psychiatric disorders. While GWAS of SA have found genome-wide significant associations (27–31), thus far none of these loci have replicated, possibly owing to limited statistical power or different study designs that may probe varying components of the genetic etiology of SA. Depending on the method of ascertainment, the prevalence of psychiatric disorders may be much higher in SA cases than in controls in these studies, which may confound the genetics of SA. Well-powered and carefully designed studies are necessary to dissect the contribution of genetic variation to SA versus psychiatric disorders and advance our understanding of the genetics of SA.

Here, we present the first GWAS meta-analysis of SA from the International Suicide Genetics Consortium (ISGC), including over 29,000 SA or suicide cases from 18 cohorts worldwide. We identify novel loci implicated in SA, disentangle the genetic etiology of SA from that of MDD and psychiatric disorders, and characterize the genetic relationship among SA, psychiatric disorders, and a range of other risk factors.

METHODS AND MATERIALS

Cohorts and Case Definition

The primary SA meta-analysis comprised 18 cohorts (Table S1 in Supplement 2; Supplement 1) ascertained for psychiatric disorders, including substance use (12 cohorts), studies of suicide or SA (4 cohorts), and population-based biobanks (2 cohorts). Cases were individuals who made a nonfatal SA (16 cohorts) or died by suicide (2 cohorts). A nonfatal SA was defined as a lifetime act of deliberate self-harm with intent to die. Information on SA was ascertained using structured clinical interviews for 10 cohorts, self-report questionnaires for 4 cohorts, and hospital records or International Classification of Diseases codes for 2 cohorts. Cases of death by suicide were ascertained from the Utah State Office of the Medical Examiner or the Medical Examiner's Office of the Hyogo Prefecture and the Division of Legal Medicine, at the Kobe University Graduate School of Medicine in Japan. A proportion of cases in the iPSYCH and Columbia University cohorts had died by suicide, determined using the Cause of Death Register in Denmark and the Columbia Classification Algorithm for Suicide Assessment, respectively (33). Individuals only endorsing suicidal ideation or nonsuicidal self-injurious behavior were not included as cases. There were 14 cohorts of European (EUR) ancestries, 2 of admixed African American (AA) ancestries, and 2 of East Asian (EAS) ancestries. All individual studies received institutional and ethical approval from their local institutional review board. Detailed cohort information is in Supplement 1 and Table S1 in Supplement 2.

Control Definition

All controls ascertained on psychiatric disorders were screened for the absence of lifetime SA. Controls from general population cohorts were screened for the absence of SA, if possible; however, because the prevalence of SA in the general population is low (3), some cohorts included unscreened controls. No controls in this study were screened for suicidal ideation or nonsuicidal self-injurious behavior. The primary SA GWAS included 29,782 cases and 519,961 controls from 18 cohorts (Table 1). Genome-wide significant associations with SA were tested in an independent replication cohort of 14,089 SA cases and 395,359 controls from Million Veteran Program (details in Supplement 1).

Genotyping, Quality Control, and Imputation

Cohorts were required to have at least 200 cases prior to quality control for inclusion. Samples underwent standard genotyping, quality control, and imputation, performed by the collaborating research teams using comparable procedures (details per cohort available in Supplement 1). Briefly, samples were genotyped on microarrays, with the exception of the China, Oxford and Virginia Commonwealth University Experimental Research on Genetic Epidemiology (CONVERGE) study, which used low-coverage sequencing. Standard parameters were used to retain individuals and SNPs after quality control for missingness, relatedness, and Hardy-Weinberg equilibrium. Imputation was performed using the appropriate ancestry reference panels, resulting in >7.7 million SNPs that were well-represented across cohorts. Identical individuals between the PGC and UK Biobank cohorts were detected using genotype-based checksums (https://personal.broadinstitute.org/sripke/share_links/zpXkV8INxUg9bayDpLToG4g58TMtjN_PGC_SCZ_w3.0718d.76) and removed from PGC cohorts. There was no other known overlap of controls between any of the 18 cohorts.

GWASs and Meta-analysis

GWASs were performed in each cohort separately, and procedures are outlined in Supplement 1. GWASs were conducted within ancestry group, covarying for ancestry-informative principal components, genomic relatedness matrices, or factors capturing site of recruitment or genotyping batch, as required. The linkage disequilibrium score regression (LDSC) intercept was calculated for all GWAS results to estimate potential confounding from cryptic relatedness or population stratification (34). Studies with significant LDSC intercepts ($p < .05$) were corrected for confounding by multiplying the standard error per SNP by the square root of the intercept (34). A transancestry meta-analysis was conducted using an inverse variance-weighted fixed-effects model in METAL (35), implemented using the Rapid Imputation for COnsortias PIpeLIne (36). A EUR-only meta-analysis was also conducted (SA-EUR) (26,590 cases and 492,022 controls). The weighted mean allele frequency and imputation INFO score per SNP was calculated, weighted by the effective sample size per

Table 1. Numbers of Cases and Controls for 18 Cohorts in the International Suicide Genetics Consortium

Cohort (Ancestry)	SA Cases	Controls
Psychiatric Genomics Consortium MDD (EUR)	1528	16,626
Psychiatric Genomics Consortium BIP (EUR)	3214	17,642
Psychiatric Genomics Consortium SCZ (EUR)	1640	7112
Psychiatric Genomics Consortium ED (EUR)	170	5070
Army STARRS (EUR)	670	10,637
German Borderline Genomics Consortium (EUR)	481	1653
UK Biobank (EUR)	2433	334,766
iPSYCH (EUR)	7003	52,227
Janssen (EUR)	255	1684
Yale-Penn (EUR)	475	1817
GISS Ukraine (EUR)	660	660
Columbia University (EUR)	577	1233
Australian Genetics of Depression Study and QSkin Study (EUR)	2792	20,193
University of Utah (EUR)	4692	20,702
Japan (EAS)	746	14,049
CONVERGE (EAS)	1148	6515
Grady Trauma Project (Admixed AA)	669	4473
Yale-Penn (Admixed AA)	629	2902
Total	29,782	519,961

AA, African American; Army STARRS, Army Study to Assess Risk and Resilience in Servicemembers; BIP, bipolar disorder; EAS, East Asian; ED, eating disorder; EUR, European; GISS, Genetic Investigation of Suicide and Suicide Attempt; MDD, major depressive disorder; SA, suicide attempt; SCZ, schizophrenia.

cohort. SNPs with a weighted minor allele frequency of <1%, weighted INFO score <0.6, or SNPs present in <80% of total effective sample size were removed from the meta-analysis results. A genome-wide significant locus was defined as the region around a SNP with $p < 5.0 \times 10^{-8}$ with linkage disequilibrium (LD) $r^2 > 0.1$, within a 3000 kb window, based on the LD structure of the Haplotype Reference Consortium European ancestries reference panel (version 1.0) (37).

Statistical Conditioning on Psychiatric Disorders

The results of the SA-EUR meta-analysis were conditioned on the genetics of MDD using multitrait-based conditional and joint analysis using GWAS summary data (mtCOJO) (38), implemented in the GCTA software package (39). mtCOJO (38) estimates the effect size of a SNP on an outcome trait conditioned on exposure trait(s). Genome-wide significant SNPs for the exposure are used as instruments to estimate the effect of the exposure on the outcome, and this effect is used to perform genome-wide conditioning, yielding conditioned effect sizes and *p* values for the outcome trait. We conditioned SA (outcome) on MDD (exposure), because MDD is the most prevalent psychiatric disorder among individuals who die by suicide (40) and has the highest genetic correlation with SA (28). The SA-EUR GWAS summary statistics were used as the outcome trait, because mtCOJO requires an ancestry-matched LD reference panel and GWAS summary statistics for the exposure trait. The PGC MDD GWAS results (excluding 23andMe) (41) were used as the exposure, and the results yielded GWAS summary statistics for SA conditioned on MDD (SA-EUR|MDD). mtCOJO is robust to sample overlap between the GWAS of the exposure and outcome. To select SNPs as instruments, independence was defined as SNPs more than 1 megabase apart or with LD $r^2 < 0.05$ based on the 1000 Genomes Project Phase 3 EUR reference panel (42). To obtain at least 10 independent instruments for MDD, the genome-wide significance threshold was adjusted to $p < 5.0 \times 10^{-7}$, leading to 15 SNPs used. In a further sensitivity analysis, GWAS summary statistics for BIP (43) and SCZ (44) were additionally included as exposure traits.

LD Score Regression

LDSC (34) was used to estimate the phenotypic variance in SA explained by common SNPs (SNP-heritability, h^2_{SNP}) from GWAS summary statistics. h^2_{SNP} was calculated on the liability scale assuming a lifetime prevalence of SA in the general population of 2% (middle of the range reported worldwide) (3). The bivariate genetic correlation attributable to genome-wide SNPs (r_g) was estimated between the SA-EUR and SA-EUR|MDD GWAS and a range of psychiatric disorders, self-harm ideation, and propensity toward risk-taking behavior, using the largest available GWAS summary statistics (Bonferroni-corrected significance threshold $p < .0042$, adjusting for 12 traits tested). Differences in r_g with SA-EUR versus SA-EUR|MDD were tested for deviation from 0, using the block jackknife method, implemented in LDSC software (45). The r_gs of SA-EUR and SA-EUR|MDD with 768 other nonoverlapping human diseases and traits were calculated on LD Hub (46) (Bonferroni-corrected significance threshold $p < 6.51 \times 10^{-5}$). Traits were precategorized manually into 15 risk factor groups previously ascribed to SA (4,5,10): autoimmune disease, neurologic disease, heart disease, hypertension, diabetes, kidney disease, cancer, alcohol use, smoking, pain, psychiatric, sleep, life stressors, socioeconomic, and education/cognition. There were 259 traits belonging to these categories, and a second reviewer validated the categories assigned to traits and their relevance to SA risk.

Polygenic Risk Scoring

Polygenic risk scores (PRSs) for SA were tested for association with SA or death by suicide versus controls in 7 target cohorts: PGC MDD, BIP and SCZ, CONVERGE (EAS ancestries), the University of Utah (suicide death cohort), Yale-Penn (AA ancestries), and Grady Trauma Project (AA ancestries). The primary SA GWAS meta-analysis was repeated excluding each cohort in turn, to create independent discovery datasets. PRSs were generated using PRS-CS (47), which uses a Bayesian regression framework to place continuous shrinkage priors on effect sizes of SNPs in the PRS, adaptive to the strength of their association signal in the discovery GWAS, and the LD structure from an external reference panel (47). The 1000 Genomes EUR, EAS, or African reference panels (42) were used to estimate LD between SNPs, as appropriate for each target cohort. PLINK 1.9 (48) was used to weight SNPs by their effect sizes calculated using PRS-CS and sum all SNPs into PRS for each individual in the target cohorts. PRSs were tested for association with case versus control status in the target cohort using a logistic regression model including covariates as per

the GWAS. The amount of phenotypic variance explained by the PRS (R^2) was calculated on the liability scale, assuming a lifetime prevalence of SA in the general population of 2% (3). Analyses in the PGC cohorts were repeated using PRSs generated from the SA-EUR|MDD GWAS results, excluding each PGC cohort in turn. Analyses performed are summarized in Table S2 in Supplement 2 (Bonferroni-corrected significance threshold $p < 3.12 \times 10^{-3}$, adjusting for 16 tests).

RESULTS

SA Shows Significant SNP-Heritability and PRS Associations

The primary SA GWAS included 29,782 cases and 519,961 controls from 18 cohorts (Table 1). Cases were predominantly of EUR ancestries (90%), with 6% of EAS ancestries and 4% of admixed AA ancestries. Case definition was lifetime SA, with ~20% of cases having died by suicide. The SNP-heritability (h^2_{SNP}) SA was 6.8% (SE = 0.005, $p = 2.00 \times 10^{-42}$) on the liability scale. The LDSC intercept was 1.04 (SE = 0.01, $p = 2.84 \times 10^{-4}$), and the attenuation ratio was 0.14 (SE = 0.04), indicating that the majority of GWAS test statistic inflation was due to polygenicity (Figure S1 in Supplement 1). PRSs for SA were tested in 7 target cohorts (Table S2 in Supplement 2). SA PRSs were significantly associated with SA in the PGC MDD, BIP, and SCZ cohorts, with a phenotypic explained variance (R^2) of 0.69% ($p = 7.17 \times 10^{-15}$), 0.68% ($p = 8.11 \times 10^{-28}$), and 0.88% ($p = 1.24 \times 10^{-17}$), respectively (liability scale). PRSs for SA were also associated with death by suicide in the University of Utah cohort, explaining slightly more phenotypic variance ($R^2 = 1.08\%$, $p = 9.79 \times 10^{-81}$). The r_g between the University of Utah suicide death GWAS and a meta-analysis of the nonfatal SA cohorts in our study was 0.77 (SE = 0.08, $p = 3.08 \times 10^{-20}$). Examining the performance of SA PRSs across ancestries showed a significant association with SA in the CONVERGE EAS cohort, although with a lower explained variance ($R^2 = 0.25\%$, $p = 3.06 \times 10^{-3}$). Analyses in 2 admixed AA cohorts showed variable results ($R^2 = 0.21\%$, $p = 5.28 \times 10^{-1}$ and $R^2 = 0.58\%$, $p = 3.44 \times 10^{-3}$, respectively) (Table S2 in Supplement 2).

GWAS of SA Identifies Locus With Stronger Effect on SA Than Psychiatric Disorders

The primary SA GWAS identified 2 genome-wide significant loci ($p < 5 \times 10^{-8}$) (Table S3 in Supplement 2). The most strongly associated locus was in an intergenic region on chromosome 7 (index SNP rs62474683, odds ratio A allele = 1.06 [1.04–1.08], $p = 1.91 \times 10^{-10}$, frequency in cases = 0.52, frequency in controls = 0.50, I^2 heterogeneity index = 0%) (forest plot Figure S2 in Supplement 1). The second genome-wide significant locus was in the major histocompatibility complex (MHC) (index SNP rs71557378, odds ratio T allele = 1.10 [1.06–1.13], $p = 1.97 \times 10^{-8}$, frequency in cases = 0.91, frequency in controls = 0.90, I^2 heterogeneity index = 46%) (forest plot Figure S3 in Supplement 1). Both loci were also genome-wide significant in the SA-EUR meta-analysis, with the same effect sizes (Table S4 in Supplement 2). In order to identify SA genetic effects not mediated by MDD, we conditioned the SA-EUR GWAS on the genetic effects of MDD via mtCOJO. After conditioning, only the chromosome 7 locus remained genome-wide significant (index SNP = rs62474683, odds ratio A allele = 1.06 [1.04–1.08], $p = 1.33 \times 10^{-8}$) (Figure 1A). Figures S4 and S5 in Supplement 1 show regional association plots of the loci before and after conditioning. The association of the chromosome 7 index SNP with SA was further replicated in the independent Million Veteran Program cohort (rs62474683, odds ratio A allele = 1.03 [1.01–1.07], $p = 3.27 \times 10^{-3}$), while the index SNP in the MHC was not associated with SA in this cohort (Table S4 in Supplement 2).

Examination of the chromosome 7 locus in published GWAS results using the Open Targets Genetics web portal (49) indicated smaller and nonsignificant effects on all psychiatric disorders (Figure 1B). Additionally, the SA-index SNP has been implicated at genome-wide significance in lifetime smoking index (50) (accounts for duration and amount of smoking) and propensity toward risk-taking behavior (51), although again with smaller effect sizes than on SA (Figure 1B; Tables S5 and S6 in Supplement 2). Pairwise GWAS analysis (see Supplement 1) of the genomic region containing the chromosome 7 locus suggested the existence of a single putative causal variant shared between SA and these phenotypes (lifetime smoking index: posterior probability = 0.99, risk-taking behavior: posterior probability = 1) (Table S7 in Supplement 2). Furthermore, a variant in high LD with the chromosome 7 index SNP (rs12666306, LD $r^2 = 0.94$) has a positive genome-wide significant effect on insomnia (reported in GWAS catalog, full summary statistics not available) (Figure 1B; Tables S5 and S6 in Supplement 2). The SA-index SNP has also been implicated in self-harm ideation (52), although not at genome-wide significance, and with a smaller effect size than on SA (Figure 1B).

MAGMA (53) enrichment analyses performed on the primary SA GWAS (see Supplement 1) showed significant enrichment of SA associations in 7 genes (Table S8 in Supplement 2), including *BTN2A1*, which is a brain-expressed gene (54) located within the MHC, that encodes a plasma-membrane protein. There was no enrichment of SA association signal in any of the biological gene sets tested (Table S9 in Supplement 2) or in the set of genes expressed in any of the 54 tissues from the Genotype-Tissue Expression project (Table S10 in Supplement 2). Examining individual genes, a transcriptome-wide association study (see Supplement 1) found 5 genes for which SA risk alleles were significantly associated with brain gene expression: *ERC2*, *RP11−266A24.1*, *TIAF1*, *BACE2*, and *NUFIP2* ($p < 4.28 \times 10^{-6}$) (Table S11 in Supplement 2). None of these genes were within genome-wide significant loci.

Evidence for Substantial Proportion of SNP-Heritability of SA Not Mediated by Psychiatric Disorders

h^2_{SNP} based on the SA-EUR GWAS was 7.5% (SE = 0.006, $p = 3.02 \times 10^{-40}$) on the liability scale (Table S12 in Supplement 2). Conditioning SA-EUR on MDD resulted in a 45% decrease in the h^2_{SNP} of SA to 4.1% (SE = 0.005, $p = 1.20 \times 10^{-16}$) on the liability scale (Table S12 in Supplement 2). Conditioning on BIP and SCZ in addition to MDD did not further change the h^2_{SNP} estimate ($h^2_{SNP} = 4.1\%$, SE = 0.005, $p = 1.20 \times 10^{-16}$). The SA-EUR|MDD results showed comparable h^2_{SNP} and complete r_g with a direct GWAS of SA within psychiatric diagnosis

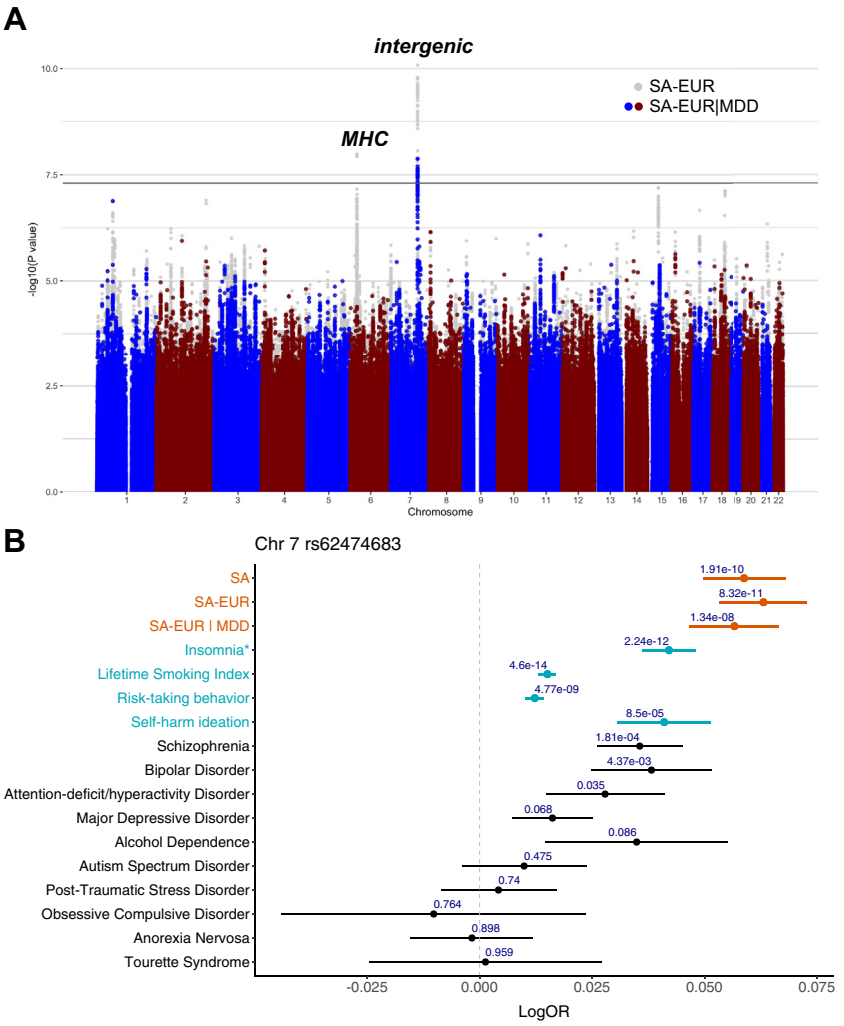


Figure 1. Genome-wide significant locus contributes to SA more strongly than psychiatric disorders and other traits. **(A)** Manhattan plot: the x-axis shows genomic position, and the y-axis shows statistical significance as $-\log_{10}(p \text{ value})$. The gray points in the background depict the results of SA-EUR, and the colored points in the foreground depict the results after conditioning these results on MDD (SA-EUR|MDD). The horizontal line shows the genome-wide significance threshold ($p < 5.0 \times 10^{-8}$). **(B)** Forest plot: the points indicate the log odds ratio of the A allele at rs62474683 (SA-index single nucleotide polymorphism on chromosome 7) on each phenotype, and the error bars show the standard error. The p value of association with each phenotype is shown above the error bars. *For insomnia, the effect size of a variant in high linkage disequilibrium with the index single nucleotide polymorphism is shown instead (rs12666306 A allele, linkage disequilibrium $r^2 = 0.94$ with rs62474683 A allele). MDD, major depressive disorder; MHC, major histocompatibility complex; OR, odds ratio; SA, suicide attempt; SA-EUR, European-only suicide attempt meta-analysis; SA-EUR|MDD, SA-EUR results after conditioning on MDD.

(Supplement 1), confirming the validity of the statistical conditioning approach to control for the genetic effects of psychiatric disorders.

Significant Genetic Overlap Between SA and Psychiatric Traits or Disorders

Genetic correlations were calculated to explore the genetic overlap between SA and 12 psychiatric traits or disorders before and after conditioning on MDD. The SA-EUR GWAS showed significant r_g with 11 traits or disorders tested, most strongly with self-harm ideation ($r_g = 0.82$, SE = 0.07, $p = 3.57 \times 10^{-36}$), MDD ($r_g = 0.78$, SE = 0.04, $p = 4.11 \times 10^{-106}$), and posttraumatic stress disorder ($r_g = 0.74$, SE = 0.09, $p = 5.29 \times 10^{-17}$) (Figure 2; Table S13 in Supplement 2). Moderate genetic correlations were also observed between SA and SCZ, attention-deficit/hyperactivity disorder, BIP, posttraumatic stress disorder, and alcohol dependence (r_gs 0.45–0.74) (Figure 2; Table S13 in Supplement 2).

To investigate whether these genetic correlations were mediated by MDD, we estimated r_g with the same traits and disorders using the SA-EUR|MDD results. Most genetic correlations with psychiatric disorders remained significant after conditioning, except for autism spectrum disorder and Tourette syndrome (Figure 2; Table S13 in Supplement 2). As expected, the r_g with MDD significantly decreased after conditioning ($p = 8.4 \times 10^{-22}$ block jackknife), as did the r_gs with self-harm ideation, posttraumatic stress disorder, and autism spectrum disorder (Figure 2; Table S13 in Supplement 2). The remaining psychiatric disorders did not show Bonferroni corrected significant differences in r_g after conditioning on MDD. Because conditional analysis only removes SNP effects on SA mediated by MDD, the remaining r_g between SA-EUR|MDD and MDD ($r_g = 0.53$, SE = 0.06, $p = 8.9 \times 10^{-19}$) indicates pleiotropic SNP effects.

Substantial Shared Genetic Architecture of SA and Nonpsychiatric Risk Factors Not Mediated by MDD

To assess the shared genetic architecture of SA, psychiatric, and nonpsychiatric phenotypes, we calculated genetic

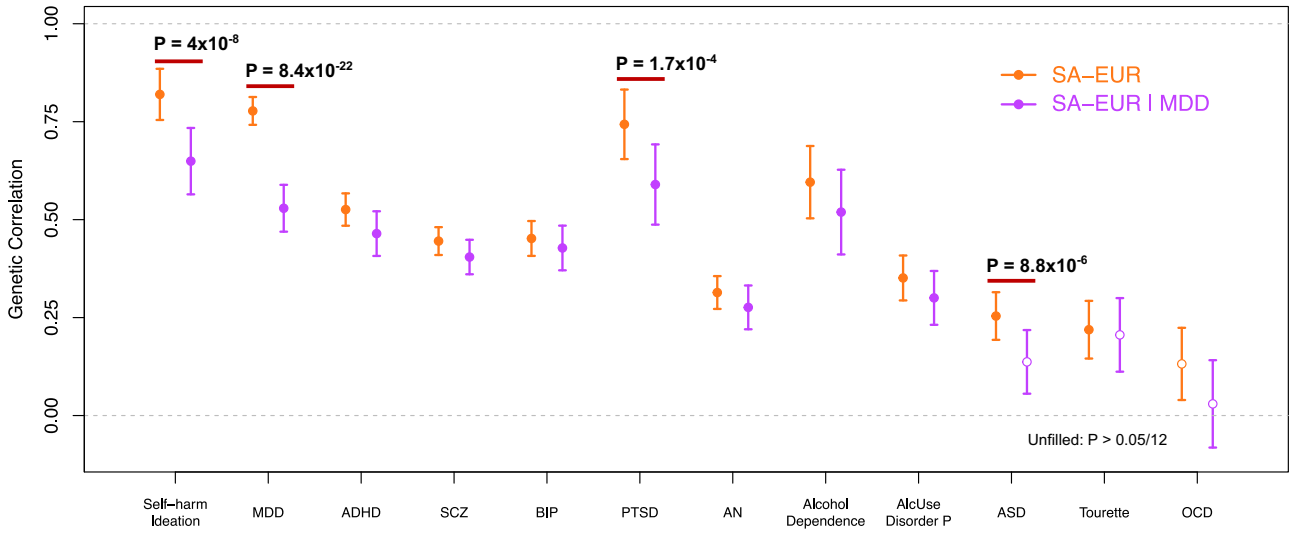


Figure 2. Substantial genetic correlation of SA with psychiatric traits or disorders before and after conditioning SA on MDD. Unfilled points indicate genetic correlations that did not pass the Bonferroni-corrected significance threshold ($p < 4.17 \times 10^{-3}$). Error bars represent the standard error. *p* values indicate significant differences in genetic correlation after conditioning that pass Bonferroni correction. ADHD, attention-deficit/hyperactivity disorder; AlcUse Disorder P, Alcohol Use Disorders Identification Test-P (measure of problematic consequences of drinking); AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, posttraumatic stress disorder; SA, suicide attempt; SA-EUR, European-only suicide attempt meta-analysis; SA-EUR|MDD, SA-EUR results after conditioning on MDD; SCZ, schizophrenia.

correlations of SA with 768 nonoverlapping phenotypes (46). There were 198 phenotypes that showed a significant r_g with SA-EUR, 133 of which were in one of the predefined SA risk categories (Figure 3A; Table S14 in Supplement 2). The most significant genetic correlations were predominantly with traits related to depressive symptoms, smoking, and socioeconomic status. On examining phenotypes in the risk categories after conditioning on MDD, 110 phenotypes retained a significant r_g with SA-EUR|MDD (Table S14 in Supplement 2). Within the psychiatric risk category, there was a 38% average decrease in the magnitude of genetic correlations with SA-EUR after conditioning, whereas the r_g values in other risk categories were much less affected by conditioning (smoking: 4.6% decrease, education/cognition: 3% decrease, alcohol: 14.5% decrease, and socioeconomic: 9.3% decrease) (Figure 3B).

DISCUSSION

We present a GWAS of SA in over 29,000 cases, identifying 2 genome-wide significant loci, including one more strongly associated with SA than psychiatric disorders or related traits. We demonstrate that a substantial proportion of the SNP-heritability of SA is independent of psychiatric diagnosis. Finally, we show that the genetic liability to SA not mediated by psychiatric disorders is shared with the genetic architecture of nonpsychiatric risk factors.

The locus most strongly associated with SA was in an intergenic region on chromosome 7. The index SNP had a larger effect on SA than on any common psychiatric disorder, remained genome-wide significant after conditioning on MDD, and replicated in an independent cohort from the Million Veteran Program. Taken together, these results suggest that the genetic association with SA at this locus is not mediated through risk for psychiatric disorders. Functional genomic data do not clearly link this variant to any gene, with the nearest gene being a long noncoding RNA (*LINC01392*) 149 kb away. The index SNP (rs62474683) is a methylation quantitative trait locus, with the SA risk allele associated with decreased methylation of a nearby DNA methylation site (probe cg04544267) in blood (55). However, this methylation site has not been linked to any gene transcript. Intriguingly, SA risk alleles at this locus have been implicated at genome-wide significance in risk-taking behavior (51), smoking (50), and insomnia (56). While variants in the MHC also reached genome-wide significance for SA, this effect did not remain after conditioning on MDD, suggesting that this association may be a byproduct of psychiatric diagnosis. Indeed, variants in the MHC have previously been associated with risk for a range of psychiatric disorders, including MDD (57).

Our GWAS results provide robust evidence of the h^2_{SNP} of SA, with an estimate of 6.8% on the liability scale (7.5% based on SA-EUR). Importantly, conditioning on MDD resulted in a smaller but significant h^2_{SNP} estimate (4.1%), corroborating previous reports (25,27) of the independent genetic contribution to SA, and illustrating the importance of accounting for potential confounding from the genetics of psychiatric disorders. Traditionally, GWASs have sought to dissect the specific genetic component of SA by studying SA within psychiatric diagnosis or covarying for cases' psychiatric diagnoses (27). Here, we demonstrate that statistical conditioning is an appropriate and easily applicable approach to control for the genetic effects of psychiatric disorders, producing equivalent results to a direct GWAS of SA within psychiatric diagnosis (Supplement 1).

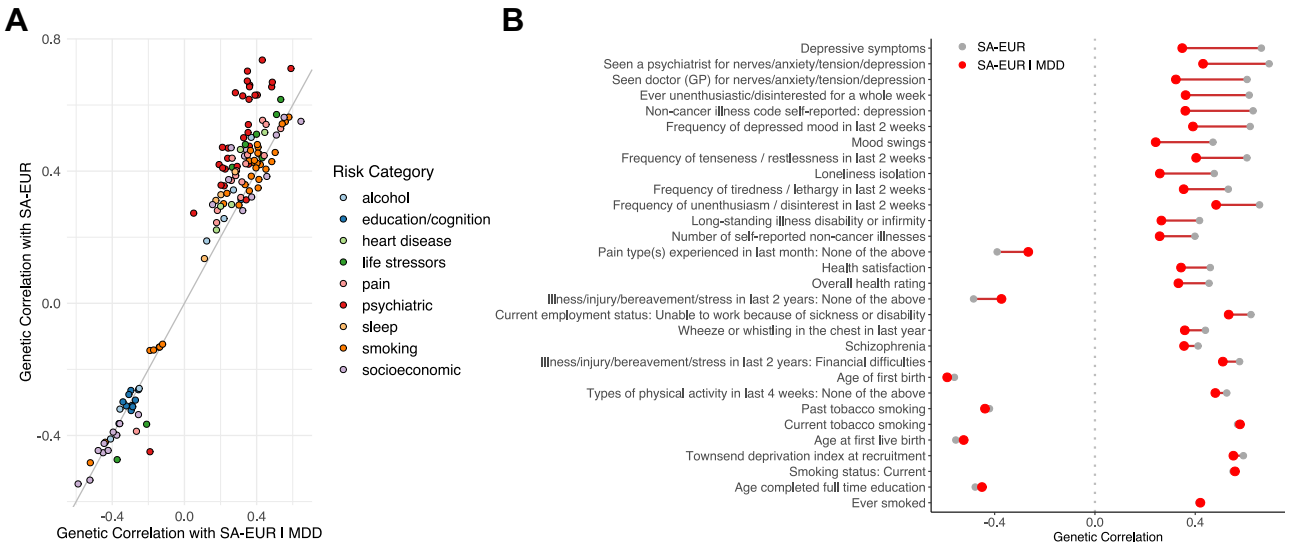


Figure 3. Conditioning SA on MDD reduces genetic correlation with psychiatric phenotypes but has limited effect on other traits. **(A)** Comparison of significant genetic correlations with the SA-EUR vs. genetic correlations with SA-EUR|MDD. Data include 198 significant genetic correlations after Bonferroni correction ($p < .05/768 = 6.51 \times 10^{-5}$) annotated by risk category. **(B)** Top 30 phenotypes with the most significant genetic correlations with SA-EUR before (gray) and after (red) conditioning on MDD (SA-EUR|MDD). Full genetic correlation results, including standard errors, are provided in Table S14 in Supplement 2. GP, general practitioner; MDD, major depressive disorder; SA, suicide attempt; SA-EUR, European-only suicide attempt meta-analysis; SA-EUR|MDD, SA-EUR results after conditioning on major depressive disorder.

SA showed substantial positive genetic correlation with many psychiatric disorders, the highest being with MDD (r_g = 0.78, SE = 0.03), consistent with previous reports (28,29,31). Genetic overlap was also particularly strong with posttraumatic stress disorder, attention-deficit/hyperactivity disorder, SCZ, and BIP (r_g = 0.44 − 0.74). After conditioning on MDD, there was a modest decrease in the genetic correlation of SA with most psychiatric disorders. Notably, SA remained strongly genetically correlated with MDD (r_g = 0.53, SE = 0.06, $p = 8.85 \times 10^{-19}$), representing pleiotropic effects between them. This genetic correlation would only be eliminated if all SNP effects on SA were mediated by MDD. Pleiotropy between psychiatric disorders is widespread (58,59), and accordingly, genetic overlap between SA and related disorders is anticipated. Our findings suggest that many pleiotropic genetic variants increase the risk for SA directly, independent of their effects on psychiatric disorders.

Significant genetic overlap was found between SA and many nonpsychiatric traits, including smoking, lower socioeconomic status, pain, lower educational attainment, reproductive traits, risk-taking behavior, sleep disturbances, and poorer overall general health. While conditioning SA on MDD reduced genetic correlations with psychiatric disorders, the genetic correlation of SA with most nonpsychiatric traits remained unchanged. This suggests a shared genetic architecture between SA and these risk factors that is not mediated by psychiatric illness. There is substantial epidemiological literature on the relationship between sleep disorders (12–15), smoking (18–20), and socioeconomic factors (60–62) and risk for SA but less on genetic overlap between them. We have not examined potential causal relationships between these risk factors and SA, but future Mendelian randomization studies that will become possible with further increases in the power of SA GWAS may highlight modifiable risk factors.

Several limitations of our study must be noted. Cases were defined using a variety of diagnostic interviews, self-report, or hospital records, which may result in phenotypic heterogeneity. However, suicidal intent was central to all phenotype definitions, and a previous study found 98% concordance between self-report of lifetime SA and face-to-face clinician interview (63). Our GWAS included both nonfatal SA and suicide death cases, and these phenotypes were highly but imperfectly genetically correlated (r_g = 0.77). Genetic correlations between SA and psychiatric disorders were examined using publicly available GWAS summary statistics; however, the prevalence of SA among the cases in these studies is unknown. Finally, population, demographic, and environmental factors are always present in genetic analyses, and while our sample is large and diverse, we did not have sufficient data to assess their possible contribution or confounding effects.

This first collaborative SA GWAS by the ISGC is almost 5-fold larger than previous studies, substantially improving statistical power. We identify a robustly associated SA risk locus and demonstrate genetic liability to SA that is not mediated through psychiatric disorders but is shared with known risk factors. We emphasize that genetic risk does not currently have meaningful predictive utility for SA, and its premature use in clinical or direct-to-consumer settings could be harmful. Future larger studies dissecting the genetic etiology of SA, psychiatric disorders, and other risk factors will provide further insights into the biological mechanisms of risk and assess potential clinical utility.

ACKNOWLEDGMENTS AND DISCLOSURES

Statistical analyses were carried out on the NL Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara and the Mount Sinai high performance computing cluster (http://hpc.mssm.edu), which is supported by the Office of Research Infrastructure of the National Institutes of Health (Grant Nos. S10OD018522 and S10OD026880). This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. This work was funded by the National Institutes of Health (Grant Nos. R01MH116269 and R01MH121455 [to DMR]), NIGMS of the National Institutes of Health (Grant No. T32GM007347 [to JK]), and the Brain & Behavior Research Foundation (NARSAD Young Investigator Award No. 29551 [to NM]).

The content is solely the responsibility of the authors and does not necessarily represent the official views of any funding body.

We thank the participants who donated their time, life experiences, and DNA to this research, and the clinical and scientific teams that worked with them.

The International Suicide Genetics Consortium has made genome-wide summary results from this study available online (https://tinyurl.com/ISGC2021). This study included some publicly available datasets accessed through dbGaP (Psychiatric Genomics Consortium [PGC] bundle phs001254) and the Haplotype Reference Consortium reference panel v.1.0 (http://www.haplotype-reference-consortium.org/home). Databases used: Open Targets Genetics web portal (https://genetics.opentargets.org), LDHub (http://ldsc.broadinstitute.org), and FUMA (https://fuma.ctglab.nl).

In the past 3 years, RCK was a consultant for Datastat, Inc., Sage Pharmaceuticals, and Takeda. HRK and JG are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. HRK is a member of an advisory board for Dicerna Pharmaceuticals and of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Dicerna, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. DL is an employee of Janssen Research & Development, LLC, and shareholder in Johnson & Johnson, the parent company of the Janssen companies. DL declares that, except for income received from her primary employer, no financial support or compensation has been received from any individual or corporate entity over the past 3 years for research or professional service, and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest. MBS has in the past 3 years been a consultant for Actelion, Acadia Pharmaceuticals, Aptinyx, Bionomics, BioXcel Therapeutics, Epivario, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, and Oxeia Biopharmaceuticals. MBS has stock options in Oxeia Biopharmaceuticals and Epivario. HJG has received travel grants and speaker honoraria from Fresenius Medical Care, Neuraxpharm, Servier, and Janssen-Cilag as well as research funding from Fresenius Medical Care. OAA is a consultant for HealthLytix and received speaker's honorarium from Lundbeck and Sunovion. RAP is employed by and holds shares in BioMarin Pharmaceuticals. MCO and MJO are supported by a collaborative research grant from Takeda Pharmaceuticals. That support did not contribute to the work described in this manuscript. EHG has served in the speakers' bureau and the advisory board of Takeda (former Shire do Brasil) Pharmaceutical. JAR-Q was on the speakers' bureau and/or acted as consultant for Eli Lilly, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogui, Lundbeck, Almirall, Braingaze, Sincrolab, Medice, and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogui, Bial, Medice, and Eli Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli Lilly, Lundbeck, Janssen-Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió. VR was on the speakers' bureau and/or acted as consultant for Takeda and Rubió in the last 5 years. She also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Rubió, Shire, Takeda, and Lundbeck. MC was on the speakers' bureau and/or acted as consultant for Janssen-Cilag in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag. All other authors report no biomedical financial interests or potential conflicts of interest.

Full acknowledgments are available in Supplement 1.

ARTICLE INFORMATION

From the Department of Genetics and Genomic Sciences (NM, DPi, PS, EAS), Department of Neuroscience (PS), and Department of Psychiatry (NM, DPi, RSK, PS, EAS), Icahn School of Medicine at Mount Sinai; Department of Biostatistics (HG), Departments of Psychiatry and Radiology (JJM), and Department of Psychiatry (HG, IO), Columbia University; Department of Psychiatry (KAH), Weill Cornell Medical College; Department of Psychiatry and Behavioral Sciences (TBB, AHF), State University of New York Downstate Medical Center; Department of Psychiatry (PMr), Columbia University College of Physicians and Surgeons; Columbia University College of Physicians and Surgeons (MMW); and Division of Translational Epidemiology (MMW), New York State Psychiatric Institute, New York, New York; Division of Genetic Medicine (JK, DMR), Department of Medicine, Vanderbilt Genetics Institute, and Department of Biomedical Informatics (DMR) and Department of Psychiatry and Behavioral Sciences (DMR), Vanderbilt University Medical Center, Nashville, Tennessee; Department of Genetics and Computational Biology (AIC, MER) and Genetics and Computational Biology (SEM), QIMR Berghofer Medical Research Institute; School of Biomedical Sciences (AIC, MER), Faculty of Medicine, Institute for Molecular Bioscience (EMB, MT, NRW), Queensland Brain Institute (DM, NRW), and Centre for Children's Health Research (EMB), The University of Queensland; School of Psychology and Counseling (DM), Queensland University of Technology, Brisbane, Queensland, Australia; National Institute for Health Research Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust (JRIC, GB); Institute of Psychiatry, Psychology and Neuroscience (JT), Department of Psychological Medicine; Institute of Psychology, Psychiatry & Neuroscience (EJSS-B); Department of Medical & Molecular Genetics (CML); and Social Genetic and Developmental Psychiatry Centre (JRIC, PMu, RAP, MRy, GB, CML), King's College London; National Institute for Health Research Biomedical Research Centre (JT), King's College London and South London and Maudsley National Health Service Foundation Trust; Division of Psychiatry (NB, AM, JP), University College London; and Genetics (RAP), BioMarin Pharmaceuticals; London, United Kingdom; Department of Psychiatry (ACE, ARD, TBB, AHF, KSK), Virginia Commonwealth University, Richmond, Virginia; Department of Psychiatry (DFLy, JG), Veterans Affairs Connecticut Healthcare Center, West Haven, Connecticut; Division of Human Genetics (DFLy, JG), Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut; Department of Psychiatry and Behavioral Sciences (AL, AP, EBB), Emory University School of Medicine, Atlanta, Georgia; Department of Psychiatry (ASh, EDB, ETM, HC, ARD) and Biomedical Informatics (HC), University of Utah School of Medicine, Salt Lake City, Utah; Centre for Genomics and Personalized Medicine (ASt, ADB, DD), Centre for Integrative Sequencing (ASt, PBMo, ADB, DD), Department of Biomedicine (ASt, ADB, DD), National Centre for Register-Based Research (ZY, PBMo, EA), Centre for Integrated Register-based Research (PBMo, EA), Psychosis Research Unit (OM), and Lundbeck Foundation Initiative for Integrative Psychiatric Research (ASt, ADB, AE, DD), iPSYCH, Aarhus University; and Lundbeck Foundation Initiative for Integrative Psychiatric Research (EA, DMH, OM, MN, TW), Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark; Institute of Epidemiology and Preventive Medicine (M-HS, WJC, P-HK), College of Public Health, National Taiwan University; Department of Psychiatry (WJC, H-CC, S-CL, C-ML, P-HK), National Taiwan University Hospital; and Department of Psychiatry (H-GH), National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; School of Psychology (HJW), Curtin University; and Division of Paediatrics (HJW), The University of Western Australia, Perth, Western Australia, Australia; Department of Psychiatry (HJW, LMT, ZY, CMB), Department of Genetics (ZY), and Department of Nutrition (CMB), University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Division of Psychiatry (MAd, JDH, AMM) and Institute for Genetics and Molecular Medicine (DPo), University of Edinburgh, Edinburgh, United Kingdom; Department of Psychiatry and Psychotherapy (SA, SRi), Universitätsmedizin Berlin; and Department of Psychiatry (SRo), Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Campus Benjamin Franklin, Berlin, Germany; Department of Psychiatry and Biobehavioral Sciences (MG, MKi), Semel Institute, David Geffen School of Medicine, University of California Los Angeles; David Geffen School of Medicine (MSt), University of California Los Angeles; Department of

Psychiatry and Biobehavioral Sciences (MSt), Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles; and Jane and Terry Semel Institute for Neuroscience and Human Behavior (RAO), Los Angeles; Biostatistics Research Center (SJai), Herbert Wertheim School of Public Health and Human Longevity Science; Department of Psychiatry (DLB, JRK, CN), Institute for Genomic Medicine (JRK) and Department of Psychiatry and School of Public Health (MBS), University of California San Diego, La Jolla; and Department of Psychiatry (WHK) and Center for Behavioral Genomics (MTT), Department of Psychiatry, University of California San Diego, San Diego, California; Department of Psychiatry (AHi), Yokohama City University Graduate School of Medicine, Yokohama, Japan; Department of Psychiatry (SO, IO), Kobe University Graduate School of Medicine, Kobe, Japan; Stanley Center for Psychiatric Research (SRi, PHL, JWS, CCZ), Stanley Center for Psychiatric Research (KCK), and Program in Medical and Population Genetics (TE, EAS), Broad Institute, Cambridge, Massachusetts; Analytical and Translational Genetics Unit (SRi, PHL), Department of Psychiatry (KCK, JWS), Psychiatric and Neurodevelopmental Genetics Unit (PHL), and Psychiatric and Neurodevelopmental Genetics Unit (JWS), Massachusetts General Hospital; Department of Epidemiology (KCK, CCZ) and Department of Environmental Health (ARo), Harvard TH Chan School of Public Health; and Department of Health Care Policy (RCK), Harvard Medical School, Boston, Massachusetts; Population Studies Center (EBW), Center for Statistical Genetics and Department of Biostatistics (MBe, LJS), and Survery Research Center (EBW), Institute for Social Research, University of Michigan, Ann Arbor, Michigan; BioRealm, LLC (AWB), Walnut, California; Oregon Research Institute (AWB), Eugene, Oregon; Department of Psychiatry (WHB), Center for Neurobiology and Behavior, and Department of Psychiatry (DWO, HRK), Perelman School of Medicine, University of Pennsylvania; Perelman School of Medicine (HH), University of Pennsylvania; Center for Applied Genomics (XC, YG, HH, DL), Children's Hospital of Philadelphia; and VISN 4 Mental Illness Research, Education, and Clinical Center (DWO, HRK), Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania; Department of Psychosomatic Medicine and Psychotherapy (MBh, CS) and Department of Genetic Epidemiology in Psychiatry (SHW, LZ, FS, JCF, TGS, LS, MRe), Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; Center for Eating Disorders at Sheppard Pratt (HB, SCa); and Department of Psychiatry and Behavioral Sciences (TGS) and Department of Mental Health (AE), Johns Hopkins University School of Medicine, Baltimore, Maryland; Center for Neuropsychiatric Research (WJC), National Health Research Institutes, Miaoli County, Taiwan; Department of Psychiatry (SCo), University of Minnesota, Minneapolis, Minnesota; Inserm U955, Institut Mondor de recherches Biomédicales, Laboratoire, Neuro-Psychiatrie Translationnelle, and Fédération Hospitalo-Universitaire de Précision Médecine en Addictologie et Psychiatrie (SJam), and Faculté de Médecine (MLe), University Paris-Est-Créteil, Créteil; Hôpital Sainte-Anne (PD, PG), GHU Paris Psychiatrie et Neurosciences; Institute of Psychiatry and Neuroscience of Paris (PD, PG, NR), INSERM U1266, Université de Paris; Department of Psychiatry and Addiction Medicine (FB, MLe), Assistance Publique Hôpitaux de Paris; Paris Bipolar and TRD Expert Centres (FB), FondaMental Foundation; UMR-S1144 Team 1: Biomarkers of relapse and therapeutic response in addiction and mood disorders (FB), INSERM; Psychiatry (FB), Université Paris Diderot; Institut de Psychiatrie (BC, M-OK), CNRS GDR 3557; Department of Evaluation, Prevention and Therapeutic innovation (BC, M-OK), GHU Paris Psychiatrie et Neurosciences; Team Pathophysiology of psychiatric diseases (BC, M-OK), Université de Paris, Institute of Psychiatry and Neuroscience of Paris, INSERM U1266; and INSERM (MLe), Paris, France; Department of Psychiatry (FF-A, SJ-M), University Hospital Bellvitge-IDIBELL and CIBEROBN; Department of Psychiatry (MSA, RB, MC, JAR-Q, MRb, VRi, CS-M, LV-R), Hospital Universitari Vall d'Hebrón; Department of Genetics, Microbiology & Statistics (MSA, MRb, CS-M), University of Barcelona; Psychiatric Genetics Unit (MSA, MC, JAR-Q, MRb, CS-M, LV-R), Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona; Department of Psychiatry and Legal Medicine (RB, MC, JAR-Q, VRi), Universitat Autònoma de Barcelona; and Center for Research in Environmental Epidemiology (MKo), Barcelona, Spain; Department of Psychiatry and Psychotherapy (MMF), Ludwig-Maximilians-University; Schön Klinik Roseneck affiliated with the Medical Faculty of the University of Munich (MMF); Department of Translational Research in Psychiatry (EBB, BM-M), Max Planck Institute of Psychiatry; Institute of Psychiatric Phenomics and Genomics (MBu, TGS), University Hospital, Ludwig-Maximilians-University; Department of Psychiatry (IG), University of Munich; Max Planck Institute of Psychiatry (MI, SK, SL); and Munich Cluster for Systems Neurology (BM-M), Munich, Germany; Department of Surgery (SG), Faculty of Medicine, Department of Psychiatry (ASK, JLK, DBW, SK, JSS, CCZ), Laboratory Medicine and Pathobiology (CCZ), and Institute of Medical Sciences (CCZ, ASK, JLK, DBW), University of Toronto; Centre for Addiction and Mental Health (ASK, JLK, SK, JSS); Department of Paediatric Laboratory Medicine (CRM) and Department of Genetics and Genomic Biology (SWS), The Hospital for Sick Children; Centre for Mental Health (DBW) and Program for Eating Disorders (DBW), University Health Network; Molecular Brain Science (JBV), Centre for Addiction and Mental Health; and Molecular Brain Science (CCZ), Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; Department of Psychiatry and Behavioral Sciences (SJG), SUNY Upstate Medical University, Syracuse, New York; Eating Recovery Center (CJ), Denver, Colorado; Department of Psychology (PKK), Florida State University, Tallahassee, Florida; Department of Psychology (KLK), Michigan State University, Lansing, Michigan; Department of Psychiatry and Psychotherapy (KL), University Medical Center, Mainz, Germany; Department of Clinical Psychology (LL), The Chicago School of Professional Psychology, Washington, DC; Behavioral Health Services (SIS), Kaiser Permanente Washington, Seattle, Washington, DC; BESE Division (PJM), King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; Department of Psychiatry (PJM, EC, GP, MP), Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; Department of Psychiatry and Behavioral Science (JEM), University of North Dakota School of Medicine and Health Sciences, Fargo, North Dakota; HudsonAlpha Institute for Biotechnology (RMM), Huntsville, Alabama; Department of Psychology (VRo), St. Petersburg State University; and Department of Borderline Disorders and Psychotherapy (VRo), V.M. Bekhterev National Medical Research Center for Psychiatry and Neurology; Saint Petersburg, Russian Federation; National Centre for Suicide Research and Prevention of Mental Ill-Health (MSo, DW), LIME, Karolinska Institutet; Department of Clinical Neuroscience (IA), Centre for Psychiatry Research, Karolinska Institutet; Department of Clinical Neuroscience (LA), Institut of Environmental Medicine (LA), and Department of Medical Epidemiology and Biostatistics (MLa, CMB), Karolinska Institutet, Stockholm, Sweden; Department of Clinical Sciences (RA), Psychiatry, Umeå University Medical Faculty, Umeå, Sweden; Department of Psychiatric Research (IA), Diakonhjemmet Hospital; NORMENT (IA), Institute of Clinical Medicine, University of Oslo; Division of Mental Health and Addiction (OAA, IM, OBS) and Department of Medical Genetics (SD), Oslo University Hospital; NORMENT (OAA, OBS), University of Oslo; and Division of Mental Health and Addiction (IM), University of Oslo, Institute of Clinical Medicine, Oslo, Norway; Discipline of Psychiatry (TMA), University of Adelaide, Adelaide, South Australia, Australia; Department of Psychiatry (MAl), Dalhousie University, Halifax, Nova Scotia, Canada; National Institute of Mental Health (MAl), Klecany, Czech Republic; Psychiatry (AA), Berkshire Healthcare NHS Foundation Trust, Bracknell, United Kingdom; Institute of Biological Psychiatry (VA, TW), Copenhagen Mental Health Services, Copenhagen University Hospital; Lundbeck Foundation Initiative for Integrative Psychiatric Research (VA, PBMo), iPSYCH; Danish Research Institute for Suicide Prevention (AE), Mental Health Centre Copenhagen; Center for Neonatal Screening (DMH), Department for Congenital Disorders, Statens Serum Institut; Mental Health Center Copenhagen (MN), Copenhagen University Hospital; and Department of Clinical Medicine (TW), and Lundbeck Foundation GeoGenetics Centre (TW), GLOBE Institute, University of Copenhagen, Copenhagen, Denmark; Biomedical Network Research Centre on Mental Health (MSA, RB, MC, JAR-Q, MRb, VRi, CS-M), Instituto de Salud Carlos III, Madrid, Spain; Department of Psychiatry and Psychotherapy (SVA, HJG) and Institute for Community Medicine (HV), University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, Germany; Department of Psychiatry (MHA), University of Coimbra, Coimbra, Portugal; Laboratory of Developmental Psychiatry (CHDB) and ADHD Outpatient Program (EHG), Adult Division, Hospital de Clínicas de Porto Alegre; Department of Genetics (CHDB) and Department of Psychiatry (EHG), Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul,

Brazil; Department of Psychiatry (BTB), Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia; Department of Psychiatry (BTB) and Institute of Epidemiology and Social Medicine (KB), University of Münster, Münster, Nordrhein-Westfalen, Germany; Health Sciences Research (JMB) and Department of Psychiatry & Psychology (MF), Mayo Clinic, Rochester, Minnesota; Department of Psychiatry (MPB, WC, RSK), UMC Utrecht Hersencentrum, Utrecht, the Netherlands; School of Psychology (RB, JMG), School of Psychiatry (MJG, PBMi, GR, CSW, TWW), and School of Medical Sciences (JMF, PRS), University of New South Wales; and Neuroscience Research Australia (JMF, JMG, MJG, PRS, CSW, TWW), Sydney, New South Wales, Australia; Mental Health Unit (JAC), Department of Psychiatry, Faculty of Medicine, Granada University Hospital Complex, and Department of Psychiatry (BG), Department of Biochemistry and Molecular Biology II and Institute of Neurosciences (MRv), and Department of Nursing (EM), Faculty of Medicine and Biomedical Research Centre, University of Granada, Granada, Spain; Institute of Neuroscience and Medicine (SCi, AJF), Research Centre Jülich, Jülich, Germany; Institute of Medical Genetics and Pathology (SCi, PH) and Department of Biomedicine (SCi, PH), University Hospital Basel, Basel, Switzerland; Institute of Human Genetics (SCi, FD, AJF, SH-H, PH, MMN), University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany; Neuropsychiatric Genetics Research Group (AC), Department of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland; Medical Research Council Centre for Neuropsychiatric Genetics and Genomics (NC, MLH, IJ, MCO, MJO, JTRW), Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, United Kingdom; Department of Translational Genomics (DC), University of Southern California, Pasadena, California; NORMENT (SD), KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway; Department of Medical & Molecular Genetics (HJE), Indiana University; and Biochemistry and Molecular Biology (HJE), Indiana University School of Medicine, Indianapolis, Indiana; Centre for Human Genetics (AJF), University of Marburg, Marburg, Germany; Department of Psychiatry and Behavioral Sciences (PVG, ARS), NorthShore University HealthSystem, Evanston, Illinois; Department of Psychiatry and Behavioral Neuroscience (PVG, ARS), University of Chicago, Chicago, Illinois; Department of Psychiatry, Psychotherapy and Psychosomatics (IG, AHa, BK, DR), Martin-Luther-University Halle-Wittenberg, Halle, Germany; Biometric Psychiatric Genetics Research Unit (MG-S), Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania; Mental Health Department (JG-P, FM), University Regional Hospital, Biomedicine Institute, Málaga, Spain; Psychiatry (SPH), Kaiser Permanente Northern California, San Francisco, California; Department of Psychiatry (JH), Laboratory of Psychiatric Genetics, Poznan University of Medical Sciences, Poznań, Poland; Department of Psychological Medicine (LAJ), University of Worcester, Worcester, United Kingdom; Department of Psychiatry and Neuroscience (LJ, MLa), University of Gothenburg, Gothenburg, Sweden; Psychiatry (JLa), North East London NHS Foundation Trust, Ilford, Cheshire, United Kingdom; Department of Psychiatry and Behavioral Sciences (DFLn, LMW), Stanford University, Stanford, California; Department of Human Genetics (CL), Department of Psychiatry (GT), and Department of Neurology and Neurosurgery (GAR), Faculty of Medicine, McGill University; and Montreal Neurological Institute and Hospital (CL, GAR), Montreal, Québec, Canada; Cancer Epidemiology and Prevention (JLi), Sklodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland; Research Institute (SLM), Lindner Center of HOPE, Mason, Ohio; Department of Psychiatry (YM, BWJHP), Amsterdam UMC, Vrije Universiteit and GGZ inGeest, Amsterdam, Netherlands; Mental Health (GM), Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, NTNU; and Psychiatry (GM), St. Olavs University Hospital, Trondheim, Norway; University of Liverpool (BM-M), Liverpool, United Kingdom; Psychiatry and Human Genetics (VN), University of Pittsburgh, Pittsburgh, Pennsylvania; Psychiatry (RAO), Erasmus University Medical Center, Rotterdam, the Netherlands; College of Medicine Institute for Genomic Health (CP) and Institute for Genomic Health (CP, MTP), SUNY Downstate Medical Center College of Medicine, Brooklyn, New York; Department of Psychiatry (JBP, VW), University of Iowa, Iowa City, Iowa; Department of Psychiatry (DQ), University of Oxford; and University of Oxford (RAP), St. Edmund Hall, Oxford, United Kingdom; Department of Psychiatry (ARe), Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt, Germany; Department of Physiology and Biophysics (DLR), Instituto de Ciencias Biomedicas Universidade de Sao Paulo, São Paulo, São Paulo, Brazil; Human Genetics Branch (TGS), Intramural Research Program, National Institute of Mental Health; Department of Psychiatry (RJU), Uniformed University of the Health Sciences; and Division of Cancer Epidemiology and Genetics (JS), National Cancer Institute, Bethesda, Maryland; Department of Psychiatry and Psychotherapy (TGS), University Medical Center Göttingen, Göttingen, Germany; Department of Biomedical and NeuroMotor Sciences (ASe), University of Bologna, Bologna, Italy; Menninger Department of Psychiatry and Behavioral Sciences (GS), Baylor College of Medicine, Houston, Texas; Laboratory of Neuropsychiatry (GS), IRCCS Santa Lucia Foundation, Rome, Italy; Department of Environmental Epidemiology (BŚ), Nofer Institute of Occupational Medicine, Lódz, Poland; Duke Molecular Physiology Institute (AEA-K, ERH, MAH, XQ) and Department of Psychiatry and Behavioral Sciences (JCB, NAK), Duke University Medical Center; VISN 6 Mid-Atlantic Mental Illness Research (JCB, NAK) and Cooperative Studies Program Epidemiology Center, Education, and Clinical Center (ERH), and VA Health Services Research and Development Center of Innovation to Accelerate Discovery and Practice Transformation (JHL), Durham Veterans Affairs Health Care System, Durham, North Carolina; Theoretical Division (BM), Los Alamos National Laboratory, Los Alamos National Laboratory, Los Alamos, New Mexico; Center of Mental Health Research (AE), Australian National University, Canberra, Australia; Estonian Genome Center (TE), Institute of Genomics, University of Tartu, Tartu, Estonia; Neuroscience (QSL), Janssen Research & Development, LLC, Titusville, New Jersey; Department of Genetics and Computational Biology (NGM) and Department of Population Health (CMO, DCW), QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland, Australia; and Institute of Health and Wellbeing (DJS), University of Glasgow, Glasgow, United Kingdom.

Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Naomi R. Wray[*,1,2], Stephan Ripke[*,3,4,5], Manuel Mattheisen[*,6,7,8], Maciej Trzaskowski[1], Enda M. Byrne[1], Abdel Abdellaoui[9], Mark J. Adams[10], Esben Agerbo[11,12,13], Tracy M. Air[14], Till F.M. Andlauer[15,16], Silviu-Alin Bacanu[17], Marie Bækvad-Hansen[13,18], Aartjan T.F. Beekman[19], Tim B. Bigdeli[17,20], Elisabeth B. Binder[15,21], Julien Bryois[22], Henriette N. Buttenschøn[13,23,24], Jonas Bybjerg-Grauholm[13,18], Na Cai[25,26], Enrique Castelao[27], Jane Hvarregaard Christensen[8,13,24], Toni-Kim Clarke[10], Jonathan R.I. Coleman[28], Lucía Colodro-Conde[29], Baptiste Couvy-Duchesne[2,30], Nick Craddock[31], Gregory E. Crawford[32,33], Gail Davies[34], Franziska Degenhardt[35], Eske M. Derks[29], Nese Direk[36,37], Conor V. Dolan[9], Erin C. Dunn[38,39,40], Thalia C. Eley[28], Valentina Escott-Price[41], Farnush Farhadi Hassan Kiadeh[42], Hilary K. Finucane[43,44], Jerome C. Foo[45], Andreas J. Forstner[35,46,47,48], Josef Frank[45], Héléna A. Gaspar[28], Michael Gill[49], Fernando S. Goes[50], Scott D. Gordon[29], Shantel Marie Weinsheimer[13,54], Jürgen Wellmann[101], Gonneke Willemsen[9], Stephanie H. Witt[45], Yang Wu[1], Hualin S. Xi[112], Jian Yang[2,113], Futao Zhang[1], Volker Arolt[114], Bernhard T. Baune[114,115,116], Klaus Berger[101], Dorret I. Boomsma[9], Sven Cichon[35,47,117,118], Udo Dannlowski[114], E.J.C. de Geus[9,119], J. Raymond Depaulo[50], Enrico Domenici[120], Katharina Domschke[121,122], Tõnu Esko[5,78], Hans J. Grabe[109], Steven P. Hamilton[123], Jakob Grove[8,13,24,51], Lynsey S. Hall[10,52], Christine Søholm Hansen[13,18], Thomas F. Hansen[53,54,55], Stefan Herms[35,47], Ian B. Hickie[56], Per Hoffmann[35,47], Georg Homuth[57], Carsten Horn[58], Jouke-Jan Hottenga[9], David M. Hougaard[13,18], David M. Howard[10,28], Marcus Ising[59], Rick Jansen[19], Ian Jones[60], Lisa A. Jones[61], Eric Jorgenson[62], James A. Knowles[63], Isaac S. Kohane[64,65,66], Julia Kraft[4], Warren W. Kretzschmar[67], Zoltán Kutalik[68,69], Yihan Li[67], Penelope A. Lind[29], Donald J. MacIntyre[70,71], Dean F. MacKinnon[50], Robert M. Maier[2], Wolfgang Maier[72], Jonathan Marchini[73], Hamdi Mbarek[9], Patrick McGrath[74], Peter McGuffin[28], Sarah E. Medland[29], Divya Mehta[2,75], Christel M. Middeldorp[9,76,77], Evelin Mihailov[78], Yuri Milaneschi[19], Lili Milani[78], Francis M. Mondimore[50], Grant W. Montgomery[1], Sara Mostafavi[79, 80], Niamh Mullins[28], Matthias Nauck[81,82], Bernard Ng[80], Michel G. Nivard[9], Dale R. Nyholt[83], Paul F. O'Reilly[28], Hogni Oskarsson[84], Caroline Hayward[124], Andrew C. Heath[89], Kenneth S. Kendler[17], Stefan Kloiber[59,125,126], Glyn Lewis[127], Qingqin S. Li[128], Susanne Lucae[59], Pamela A.F. Madden[89], Patrik K. Magnusson[22], Nicholas G. Martin[29], Andrew M. McIntosh[10,34], Andres Metspalu[78,129], Ole Mors[13,130], Preben Bo Mortensen[11,12,13,24], Bertram Müller-Myhsok[15,131,132], Merete Nordentoft[13,133],

Markus M. Nöthen[35], Michael C. O'Donovan[60], Sara A. Paciga[134], Nancy L. Pedersen[22], Michael J. Owen[60], Jodie N. Painter[29], Carsten Bøcker Pedersen[11,12,13], Marianne Giørtz Pedersen[11,12,13], Roseann E. Peterson[17,85], Wouter J. Peyrot[19], Giorgio Pistis[27], Danielle Posthuma[86,87], Jorge A. Quiroz[88], Per Qvist[8,13,24], John P. Rice[89], Brien P. Riley[17], Margarita Rivera[28,90], Saira Saeed Mirza[36], Robert Schoevers[91], Eva C. Schulte[92,93], Ling Shen[62], Jianxin Shi[94], Stanley I. Shyn[95], Engilbert Sigurdsson[96], Grant C.B. Sinnamon[97], Johannes H. Smit[19], Daniel J. Smith[98], Hreinn Stefansson[99], Stacy Steinberg[99], Fabian Streit[45], Jana Strohmaier[45], Katherine E. Tansey[100], Henning Teismann[101], Alexander Teumer[102], Wesley Thompson[13,54,103,104], Pippa A. Thomson[105], Thorgeir E. Thorgeirsson[99], Matthew Traylor[106], Jens Treutlein[45], Vassily Trubetskoy[4], André G. Uitterlinden[107], Daniel Umbricht[108], Sandra Van der Auwera[109], Albert M. van Hemert[110], Alexander Viktorin[22], Peter M. Visscher[1,2], Yunpeng Wang[13,54,104], Bradley T. Webb[111], Brenda W.J.H. Penninx[19], Roy H. Perlis[38,135], David J. Porteous[105], James B. Potash[136], Martin Preisig[27], Marcella Rietschel[45], Catherine Schaefer[62], Thomas G. Schulze[45,93,137,138,139], Jordan W. Smoller[38,39,40], Kari Stefansson[99,140], Henning Tiemeier[36,141,142], Rudolf Uher[143], Henry Völzke[102], Myrna M. Weissman[74,144], Thomas Werge[13,54,145], Cathryn M. Lewis*[,28,146], Douglas F. Levinson*[,147], Gerome Breen*[,28,148], Anders D. Børglum*[,8,13,24], and Patrick F. Sullivan*[,22,149,150]

See Supplement 1 for all affiliations of consortium members.

Bipolar Disorder Working Group of the Psychiatric Genomics Consortium

Niamh Mullins[1,2], Andreas J. Forstner[3,4,5], Kevin S. O'Connell[6,7], Brandon Coombes[8], Jonathan R.I. Coleman[9,10], Zhen Qiao[11], Thomas D. Als[12,13,14], Tim B. Bigdeli[15,16], Sigrid Børte[17,18,19], Julien Bryois[20], Alexander W. Charney[2], Ole Kristian Drange[21,22], Michael J. Gandal[23], Saskia P. Hagenaars[9,10], Masashi Ikeda[24], Nolan Kamitaki[25,26], Minsoo Kim[23], Kristi Krebs[27], Georgia Panagiotaropoulou[28], Brian M. Schilder[1,29,30,31], Laura G. Sloofman[1], Stacy Steinberg[32], Vassily Trubetskoy[28], Bendik S. Winsvold[19,33], Hong-Hee Won[34], Liliya Abramova[35], Kristina Adorjan[36,37], Esben Agerbo[14,38,39], Mariam Al Eissa[40], Diego Albani[41], Ney Alliey-Rodriguez[42,43], Adebayo Anjorin[44], Verneri Antilla[45], Anastasia Antoniou[46], Swapnil Awasthi[28], Ji Hyun Baek[47], Marie Bækvad-Hansen[14,48], Nicholas Bass[40], Michael Bauer[49], Eva C. Beins[3], Sarah E. Bergen[20], Armin Birner[50], Carsten Bøcker Pedersen[14,38,39], Erlend Bøen[51], Marco P. Boks[52], Rosa Bosch[53,54,55,56], Murielle Brum[57], Ben M. Brumpton[19], Nathalie Brunkhorst-Kanaan[57], Monika Budde[36], Jonas Bybjerg-Grauholm[14,48], William Byerley[58], Murray Cairns[59], Miquel Casas[53,54,55,56], Pablo Cervantes[60], Toni-Kim Clarke[61], Cristiana Cruceanu[60,62], Alfredo Cuellar-Barboza[63,64], Julie Cunningham[65], David Curtis[66,67], Piotr M. Czerski[68], Anders M. Dale[69], Nina Dalkner[50], Friederike S. David[3], Franziska Degenhardt[3,70], Srdjan Djurovic[71,72], Amanda L. Dobbyn[1,2], Athanassios Douzenis[46], Torbjørn Elvsåshagen[18,73,74], Valentina Escott-Price[75], I. Nicol Ferrier[76], Alessia Fiorentino[40], Tatiana M. Foroud[77], Liz Forty[75], Josef Frank[78], Oleksandr Frei[6,18], Nelson B. Freimer[23,79], Louise Frisén[80], Katrin Gade[36,81], Julie Garnham[82], Joel Gelernter[83,84,85], Marianne Giørtz Pedersen[14,38,39], Ian R. Gizer[86], Scott D. Gordon[87], Katherine Gordon-Smith[88], Tiffany A. Greenwood[89], Jakob Grove[12,13,14,90], José Guzman-Parra[91], Kyooseob Ha[92], Magnus Haraldsson[93], Martin Hautzinger[94], Urs Heilbronner[36], Dennis Hellgren[20], Stefan Herms[3,95,96], Per Hoffmann[3,95,96], Peter A. Holmans[75], Laura Huckins[1,2], Stéphane Jamain[97,98], Jessica S. Johnson[1,2], Janos L. Kalman[36,37,99], Yoichiro Kamatani[100,101], James L. Kennedy[102,103,104,105], Sarah Kittel-Schneider[57,106], James A. Knowles[107,108], Manolis Kogevinas[109], Maria Koromina[110], Thorsten M. Kranz[57], Henry R. Kranzler[111,112], Michiaki Kubo[113], Ralph Kupka[114,115,116], Steven A. Kushner[117], Catharina Lavebratt[118,119], Jacob Lawrence[120], Markus Leber[121], Heon-Jeong Lee[122], Phil H. Lee[123], Shawn E. Levy[124], Catrin Lewis[75], Calwing Liao[125,126], Susanne Lucae[62], Martin Lundberg[118,119], Donald J. MacIntyre[127], Sigurdur H. Magnusson[32], Wolfgang Maier[128], Adam Maihofer[89], Dolores Malaspina[1,2], Eirini Maratou[129], Lina Martinsson[80], Manuel Mattheisen[12,13,14,106,130], Nathaniel W. McGregor[131], Peter McGuffin[9], James D. McKay[132], Helena Medeiros[108], Sarah E Medland[87], Vincent Millischer[118,119], Grant W. Montgomery[11], Jennifer L. Moran[25,133], Derek W. Morris[134], Thomas W. Mühleisen[4,95], Niamh O'Brien[40], Claire O'Donovan[82], Loes M. Olde Loohuis[23,79], Lilijana Oruc[135], Sergi Papiol[36,37], Antonio F. Pardiñas[75], Amy Perry[88], Andrea Pfennig[49], Evgenia Porichi[46], James B. Potash[136], Digby Quested[137,138], Towfique Raj[1,29,30,31], Mark H. Rapaport[139], J. Raymond DePaulo[136], Eline J. Regeer[140], John P. Rice[141], Fabio Rivas[91], Margarita Rivera[142,143], Julian Roth[106], Panos Roussos[1,2,29], Douglas M. Ruderfer[144], Cristina Sánchez-Mora[53,54,56,145], Eva C. Schulte[36,37], Fanny Senner[36,37], Sally Sharp[40], Paul D. Shilling[89], Engilbert Sigurdsson[93,146], Lea Sirignano[78], Claire Slaney[82], Olav B. Smeland[6,7], Daniel J. Smith[147], Janet L. Sobell[148], Christine Søholm Hansen[14,48], Maria Soler Artigas[53,54,56,145], Anne T. Spijker[149], Dan J. Stein[150], John S. Strauss[102], Beata Świątkowska[151], Chikashi Terao[101], Thorgeir E. Thorgeirsson[32], Claudio Toma[152,153,154], Paul Tooney[59], Evangelia-Eirini Tsermpini[110], Marquis P. Vawter[155], Helmut Vedder[156], James T.R. Walters[75], Stephanie H. Witt[78], Simon Xi[157], Wei Xu[158], Jessica Mei Kay Yang[75], Allan H. Young[159,160], Hannah Young[1], Peter P. Zandi[136], Hang Zhou[83,84], Lea Zillich[78], HUNT All-In Psychiatry[161], Rolf Adolfsson[162], Ingrid Agartz[51,130,163], Martin Alda[82,164], Lars Alfredsson[165], Gulja Babadjanova[166], Lena Backlund[118,119], Bernhard T. Baune[167,168,169], Frank Bellivier[170,171], Susanne Bengesser[50], Wade H. Berrettini[172], Douglas H.R. Blackwood[61], Michael Boehnke[173], Anders D. Børglum[14,174,175], Gerome Breen[9,10], Vaughan J. Carr[176], Stanley Catts[177], Aiden Corvin[178], Nicholas Craddock[75], Udo Dannlowski[167], Dimitris Dikeos[179], Tõnu Esko[26,27,180,181], Bruno Etain[170,171], Panagiotis Ferentinos[9,46], Mark Frye[64], Janice M. Fullerton[152,153], Micha Gawlik[106], Elliot S. Gershon[42,182], Fernando S. Goes[136], Melissa J. Green[152,176], Maria Grigoroiu-Serbanescu[183], Joanna Hauser[68], Frans Henskens[59], Jan Hillert[80], Kyung Sue Hong[47], David M. Hougaard[14,48], Christina M. Hultman[20], Kristian Hveem[19,184], Nakao Iwata[24], Assen V. Jablensky[185], Ian Jones[75], Lisa A Jones[88], René S. Kahn[2,52], John R. Kelsoe[89], George Kirov[75], Mikael Landén[20,186], Marion Leboyer[97,98,187], Cathryn M. Lewis[9,10,188], Qingqin S. Li[189], Jolanta Lissowska[190], Christine Lochner[191], Carmel Loughland[59], Nicholas G. Martin[87,192], Carol A. Mathews[193], Fermin Mayoral[91], Susan L. McElroy[194], Andrew M. McIntosh[127,195], Francis J. McMahon[196], Ingrid Melle[6,197], Patricia Michie[59], Lili Milani[27], Philip B. Mitchell[176], Gunnar Morken[21,198], Ole Mors[14,199], Preben Bo Mortensen[12,14,38,39], Bryan Mowry[177], Bertram Müller-Myhsok[62,200,201], Richard M. Myers[124], Benjamin M. Neale[25,45,180], Caroline M. Nievergelt[89,202], Merete Nordentoft[14,203], Markus M. Nöthen[3], Michael C. O'Donovan[75], Ketil J. Oedegaard[204,205], Tomas Olsson[206], Michael J. Owen[75], Sara A. Paciga[207], Chris Pantelis[208], Carlos Pato[108], Michele T. Pato[108], George P. Patrinos[110,209,210], Roy H. Perlis[211,212], Danielle Posthuma[213,214], Josep Antoni Ramos-Quiroga[53,54,55,56], Andreas Reif[57], Eva Z. Reininghaus[50], Marta Ribasés[53,54,56,145], Marcella Rietschel[78], Stephan Ripke[25,28,45], Guy A. Rouleau[126,215], Takeo Saito[24], Ulrich Schall[59], Martin Schalling[118,119], Peter R. Schofield[152,153], Thomas G. Schulze[36,78,81,136,216], Laura J. Scott[173], Rodney J. Scott[59], Alessandro Serretti[217], Cynthia Shannon Weickert[152,176,218], Jordan W. Smoller[25,133,219], Hreinn Stefansson[32], Kari Stefansson[32,220], Eystein Stordal[221,222], Fabian Streit[78], Patrick F. Sullivan[20,223,224], Gustavo Turecki[225], Arne E. Vaaler[226], Eduard Vieta[227], John B. Vincent[102], Irwin D. Waldman[228], Thomas W. Weickert[152,176,218], Thomas Werge[14,229,230,231], Naomi R. Wray[11,232], John-Anker Zwart[18,19,33], Joanna M. Biernacka[8,64], John I. Nurnberger[233], Sven Cichon[3,4,95,96], Howard J. Edenberg[77,234], Eli A. Stahl[1,2,180], Andrew McQuillin[40], Arianna Di Florio[75,224], Roel A. Ophoff[23,79,117,235], and Ole A. Andreassen[6,7]

See Supplement 1 for all affiliations of consortium members.

Eating Disorders Working Group of the Psychiatric Genomics Consortium

Roger A.H. Adan[1,2,3], Lars Alfredsson[4], Tetsuya Ando[5], Ole A. Andreassen[6], Harald Aschauer[7], Jessica H. Baker[8], Vladimir Bencko[9], Andrew W. Bergen[10,11], Wade H. Berrettini[12], Andreas Birgegård[13,14,15], Joseph M. Boden[16], Ilka Boehm[17], Claudette Boni[18], Vesna Boraska Perica[19,20], Harry Brandt[21], Gerome Breen[22,23], Julien Bryois[15], Katharina Buehren[24], Cynthia M. Bulik[8,15,25], Roland Burghardt[26], Laura Carlberg[27], Matteo Cassina[28], Sven Cichon[29,30,31], Maurizio Clementi[28], Jonathan R.I. Coleman[22,23], Roger D. Cone[32], Philippe Courtet[33], Steven Crawford[21], Scott Crow[34], James J. Crowley[13,35], Unna N. Danner[2], Oliver S.P. Davis[36,37], Martina de Zwaan[38], George Dedoussis[39], Daniela Degortes[40], Janiece E. DeSocio[41], Danielle M. Dick[42,43,44], Dimitris Dikeos[45], Christian Dina[46], Monika Dmitrzak-Weglarz[47], Elisa Docampo Martinez[48,49,50], Laramie E. Duncan[51], Karin Egberts[52], Christian R. Marshall[126], Nicholas G. Martin[72], Manuel Mattheisen[13,14,75,127], Morten Mattingsdal[6], Sara McDevitt[128,129], Peter McGuffin[22], Sarah E. Medland[72], Andres Metspalu[53,130], Ingrid Meulenbelt[131], Nadia Micali[132,133], James Mitchell[134], Karen Mitchell[135,136], Palmiero Monteleone[137], Alessio Maria Monteleone[124], Grant W.

Montgomery[72,86,138], Preben Bo Mortensen[76,114,115], Melissa A. Munn-Chernoff[8], Benedetta Nacmias[139], Marie Navratilova[63], Ioanna Ntalla[39], Catherine M. Olsen[140], Roel A. Ophoff[141,142], Julie K. O'Toole[143], Leonid Padyukov[110], Aarno Palotie[54,102,144], Jacques Pantel[18], Hana Papezova[97], Richard Parker[72], John F. Pearson[145], Nancy L. Pedersen[15], Stefan Ehrlich[17], Geòrgia Escaramís[48,49,50], Tõnu Esko[53,54], Thomas Espeseth[55], Xavier Estivill[48,49,50,56], Anne Farmer[22], Angela Favaro[40], Fernando Fernández-Aranda[57,58], Manfred M. Fichter[59,60], Krista Fischer[53], James A.B. Floyd[61], Manuel Föcker[62], Lenka Foretova[63], Andreas J. Forstner[30,64,65,66], Monica Forzan[28], Christopher S. Franklin[19], Steven Gallinger[67], Giovanni Gambaro[68], Héléna A. Gaspar[22,23], Ina Giegling[69], Johanna Giuranna[70], Paola Giusti-Rodríquez[35], Fragiskos Gonidakis[71], Scott Gordon[72], Philip Gorwood[73,74], Monica Gratacos Mayora[48,49,50], Jakob Grove[75,76,77,78], Sébastien Guillaume[33], Yiran Guo[79], Hakon Hakonarson[79,80], Katherine A. Halmi[81], Ken B. Hanscombe[82], Konstantinos Hatzikotoulas[19,83], Joanna Hauser[84], Johannes Hebebrand[70], Sietske G. Helder[22,85], Anjali K. Henders[86], Stefan Herms[29,30], Beate Herpertz-Dahlmann[24], Wolfgang Herzog[87], Anke Hinney[70], L. John Horwood[16], Christopher Hübel[15,22], Liselotte V. Petersen[76,114,115], Dalila Pinto[88], Kirstin L. Purves[22], Anu Raevuori[101], Nicolas Ramoz[18], Ted Reichborn-Kjennerud[112,146], Valdo Ricca[147], Samuli Ripatti[148], Stephan Ripke[149,150,151], Franziska Ritschel[17,152], Marion Roberts[22], Dan Rujescu[69], Filip Rybakowski[154], Paolo Santonastaso[155], André Scherag[156], Stephen W. Scherer[157,158], Ulrike Schmidt[22], Nicholas J. Schork[159], Alexandra Schosser[160], Jochen Seitz[24], Lenka Slachtova[161], P. Eline Slagboom[131], Margarita C.T. Slof-Op 't Landt[162,163], Agnieszka Slopien[164], Nicole Soranzo[19,165,166,167], Sandro Sorbi[139,168], Lorraine Southam[19], Vidar W. Steen[169,170], Michael Strober[171,172], Laura M. Huckins[88], James I. Hudson[89], Hartmut Imgart[90], Hidetoshi Inoko[91], Vladimir Janout[92], Susana Jiménez-Murcia[57,58], Craig Johnson[93], Jennifer Jordan[94,95], Antonio Julià[96], Gursharan Kalsi[22], Deborah Kaminská[97], Allan S. Kaplan[98,99,100], Jaakko Kaprio[101,102], Leila Karhunen[103], Andreas Karwautz[104], Martien J.H. Kas[1,105], Walter H. Kaye[106], James L. Kennedy[98,99,100], Martin A. Kennedy[107], Anna Keski-Rahkonen[101], Kirsty Kiezebrink[108], Youl-Ri Kim[109], Katherine M. Kirk[72], Lars Klareskog[110], Kelly L. Klump[111], Gun Peggy S. Knudsen[112], Mikael Landén[15,113], Janne T. Larsen[76,114,115], Stephanie Le Hellard[116,117,118], Virpi M. Leppä[15], Dong Li[79], Paul Lichtenstein[15], Lisa Lilenfeld[119], Bochao Danae Lin[1], Jolanta Lissowska[120], Astri Lundervold[121], Jurjen Luykx[1], Pierre J. Magistretti[122,123], Mario Maj[124], Katrin Mannik[53,125], Sara Marsal[96], Garret D. Stuber[8,173], Patrick F. Sullivan[8,15,35], Beata Świątkowska[174], Jin P. Szatkiewicz[35], Ioanna Tachmazidou[19], Elena Tenconi[40], Laura M. Thornton[8], Alfonso Tortorella[175,176], Federica Tozzi[177], Janet Treasure[22], Artemis Tsitsika[178], Marta Tyszkiewicz-Nwafor[164], Konstantinos Tziouvas[179], Annemarie A. van Elburg[2,180], Eric F. van Furth[162,163], Tracey D. Wade[181], Gudrun Wagner[104], Esther Walton[17], Hunna J. Watson[8,182,183], Thomas Werge[184], David C. Whiteman[140], H. Erich Wichmann[185], Elisabeth Widen[102], D. Blake Woodside[99,100,186,187], Shuyang Yao[15], Zeynep Yilmaz[8,35], Eleftheria Zeggini[19,83], Stephanie Zerwas[8], and Stephan Zipfel[188]

See Supplement 1 for all affiliations of consortium members.

German Borderline Genomics Consortium

Stephanie H. Witt[1], Fabian Streit[1], Martin Jungkunz[2,3], Josef Frank[1], Swapnil Awasthi[4], Jens Treutlein[1], Lydie Dietl[5], Cornelia E Schwarze[6], Norbert Dahmen[8], Björn H. Schott[4,9], Markus M. Nöthen[10,11], Stephan Ripke[4,12,13], Arian Mobascher[8], Dan Rujescu[7], Klaus Lieb[8], Stefan Roepke[5], Christian Schmahl[2], Martin Bohus[3], and Marcella Rietschel[1]

See Supplement 1 for all affiliations of consortium members.

MVP Suicide Exemplar Workgroup

Silvia Crivelli, Ph.D. (Lawrence Berkeley National Laboratory), Michelle F. Dennis, B.A. (Durham Veterans Affairs Health Care System & Duke University School of Medicine), Phillip D. Harvey, Ph.D. (University of Miami Miller School of Medicine, Miami, FL), Bruce W. Carter (VA Medical Center), Jennifer E. Huffman, Ph.D. (Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System), Daniel Jacobson, Ph.D. (Oak Ridge National Laboratory), Ravi Madduri, Ph.D. (Argonne National Laboratory), Maren K. Olsen, Ph.D. (Duke University School of Medicine), and John Pestian, Ph.D. (Oak Ridge National Laboratory).

VA Million Veteran Program (MVP)

J. Michael Gaziano, M.D., M.P.H. (co-chair, VA Boston Healthcare System), Sumitra Muralidhar, Ph.D. (co-chair, U.S. Department of Veterans Affairs), Rachel Ramoni, D.M.D., Sc.D. (U.S. Department of Veterans Affairs), Jean Beckham, Ph.D. (Durham VA Medical Center), Kyong-Mi Chang, M.D. (Philadelphia VA Medical Center), Christopher J. O'Donnell, M.D., M.P.H. (VA Boston Healthcare System), Philip S. Tsao, Ph.D. (VA Palo Alto Health Care System), James Breeling, M.D. (Ex-Officio, U.S. Department of Veterans Affairs), Grant Huang, Ph.D. (Ex-Officio, U.S. Department of Veterans Affairs), and J.P. Casas Romero, M.D., Ph.D. (Ex-Officio, VA Boston Healthcare System). MVP Program Office: Sumitra Muralidhar, Ph.D., and Jennifer Moser, Ph.D., both of U.S. Department of Veterans Affairs. MVP Recruitment/Enrollment: Recruitment/Enrollment Director/Deputy Director, Boston—Stacey B. Whitbourne, Ph.D., Jessica V. Brewer, M.P.H. (VA Boston Healthcare System). MVP Coordinating Centers: Clinical Epidemiology Research Center (CERC), West Haven—Mihaela Aslan, Ph.D. (West Haven VA Medical Center). Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque—Todd Connor, Pharm.D., Dean P. Argyres, B.S., M.S. (New Mexico VA Health Care System). Genomics Coordinating Center, Palo Alto—Philip S. Tsao, Ph.D. (VA Palo Alto Health Care System). MVP Boston Coordinating Center, Boston—J. Michael Gaziano, M.D., M.P.H. (VA Boston Healthcare System). MVP Information Center, Canandaigua—Brady Stephens, M.S. (Canandaigua VA Medical Center). VA Central Biorepository, Boston—Mary T. Brophy, M.D., M.P.H., Donald E. Humphries, Ph.D., Luis E. Selva, Ph.D. (VA Boston Healthcare System). MVP Informatics, Boston—Nhan Do, M.D., Shahpoor Shayan (VA Boston Healthcare System). MVP Data Operations/Analytics, Boston—Kelly Cho, Ph.D. (VA Boston Healthcare System). MVP Science: Science Operations—Christopher J. O'Donnell, M.D., M.P.H. (VA Boston Healthcare System). Genomics Core— Christopher J. O'Donnell, M.D., M.P.H., Saiju Pyarajan, Ph.D. (VA Boston Healthcare System), Philip S. Tsao, Ph.D. (VA Palo Alto Health Care System). Phenomics Core—Kelly Cho, M.P.H., Ph.D. (VA Boston Healthcare System). Data and Computational Sciences—Saiju Pyarajan, Ph.D. (VA Boston Healthcare System). Statistical Genetics—Elizabeth Hauser, Ph.D. (Durham VA Medical Center). Yan Sun, Ph.D. (Atlanta VA Medical Center). Hongyu Zhao, Ph.D. (West Haven VA Medical Center. Current MVP Local Site Investigators: Peter Wilson, M.D. (Atlanta VA Medical Center); Rachel McArdle, Ph.D. (Bay Pines VA Healthcare System); Louis Dellitalia, M.D. (Birmingham VA Medical Center); Kristin Mattocks, Ph.D., M.P.H. (Central Western Massachusetts Healthcare System); John Harley, M.D., Ph.D. (Cincinnati VA Medical Center); Clement J. Zablocki (VA Medical Center); Jeffrey Whittle, M.D., M.P.H.; Frank Jacono, M.D. (VA Northeast Ohio Healthcare System); Jean Beckham, Ph.D. (Durham VA Medical Center); Edith Nourse Rogers Memorial Veterans Hospital; Salvador Gutierrez, M.D. (Edward Hines, Jr. VA Medical Center); Gretchen Gibson, D.D.S., M.P.H. (Veterans Health Care System of the Ozarks); Kimberly Hammer, Ph.D. (Fargo VA Health Care System); Laurence Kaminsky, Ph.D. (VA Health Care Upstate New York); Gerardo Villareal, M.D. (New Mexico VA Health Care System); Scott Kinlay, M.B.B.S., Ph.D. (VA Boston Healthcare System); Junzhe Xu, M.D. (VA Western New York Healthcare System); Mark Hamner, M.D. (Ralph H. Johnson VA Medical Center); Roy Mathew, M.D. (Columbia VA Health Care System); Sujata Bhushan, M.D. (VA North Texas Health Care System); Pran Iruvanti, DO, Ph.D. (Hampton VA Medical Center); Michael Godschalk, M.D. (Richmond VA Medical Center); Zuhair Ballas, M.D. (Iowa City VA Health Care System); Douglas Ivins, M.D. (Eastern Oklahoma VA Health Care System); Stephen Mastorides, M.D. (James A. Haley Veterans' Hospital); Jonathan Moorman, M.D., Ph.D. (James H. Quillen VA Medical Center); Saib Gappy, M.D. (John D. Dingell VA Medical Center); Jon Klein, M.D., Ph.D. (Louisville VA Medical Center); Nora Ratcliffe, M.D. (Manchester VA Medical Center); Hermes Florez, M.D., Ph.D. (Miami VA Health Care System); Olaoluwa Okusaga, M.D. (Michael E. DeBakey VA Medical Center); Maureen Murdoch, M.D., M.P.H. (Minneapolis VA Health Care System); Peruvemba Sriram, M.D. (N FL/S GA Veterans Health System); Shing Shing Yeh, Ph.D., M.D. (Northport VA Medical Center); Neeraj Tandon, M.D. (Overton Brooks VA Medical Center); Darshana Jhala, M.D. (Philadelphia VA Medical Center); Samuel Aguayo, M.D. (Phoenix VA Health Care System); David Cohen, M.D. (Portland VA Medical Center); Satish Sharma, M.D. (Providence VA Medical Center); Suthat Liangpunsakul, M.D., M.P.H. (Richard Roudebush VA Medical Center); Kris Ann Oursler, M.D. (Salem VA

Medical Center); Mary Whooley, M.D. (San Francisco VA Health Care System); Sunil Ahuja, M.D. (South Texas Veterans Health Care System); Joseph Constans, Ph.D. (Southeast Louisiana Veterans Health Care System); Paul Meyer, M.D., Ph.D. (Southern Arizona VA Health Care System); Jennifer Greco, M.D. (Sioux Falls VA Health Care System); Michael Rauchman, M.D. (St. Louis VA Health Care System); Richard Servatius, Ph.D. (Syracuse VA Medical Center); Melinda Gaddy, Ph.D. (VA Eastern Kansas Health Care System); Agnes Wallbom, M.D., M.S. (VA Greater Los Angeles Health Care System); Timothy Morgan, M.D. (VA Long Beach Healthcare System); Todd Stapley, D.O. (VA Maine Healthcare System); Scott Sherman, M.D., M.P.H. (VA New York Harbor Healthcare System); George Ross, M.D. (VA Pacific Islands Health Care System); Philip Tsao, Ph.D. (VA Palo Alto Health Care System); Patrick Strollo Jr., M.D. (VA Pittsburgh Health Care System); Edward Boyko, M.D. (VA Puget Sound Health Care System); Laurence Meyer, M.D., Ph.D. (VA Salt Lake City Health Care System); Samir Gupta, M.D., M.S.C.S. (VA San Diego Healthcare System); Mostaqul Huq, Pharm.D., Ph.D. (VA Sierra Nevada Health Care System); Joseph Fayad, M.D. (VA Southern Nevada Healthcare System); Adriana Hung, M.D., M.P.H. (VA Tennessee Valley Healthcare System); Jack Lichy, M.D., Ph.D. (Washington, DC VA Medical Center); Robin Hurley, M.D. (W.G., Bill Hefner VA Medical Center); Brooks Robey, M.D. (White River Junction VA Medical Center); and Robert Striker, M.D., Ph.D. (William S. Middleton Memorial Veterans Hospital).

NM and JK contributed equally to this work.

Address correspondence to Niamh Mullins, Ph.D., at niamh.mullins@mssm.edu, or Douglas M. Ruderfer, Ph.D., at douglas.ruderfer@vanderbilt.edu.

Received Jan 5, 2021; revised May 7, 2021; accepted May 26, 2021.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2021.05.029.

REFERENCES

1. World Health Organization: Suicide data Available at: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. Accessed May 9, 2020.
2. Centers for Disease Control and Prevention: WISQARS™ — Web-based injury statistics query and reporting system Available at: https://www.cdc.gov/injury/wisqars/index.html. Accessed May 9, 2020.
3. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, *et al.* (2008): Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry 192:98–105.
4. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, *et al.* (2017): Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull 143:187–232.
5. Fazel S, Runeson B (2020): Suicide. N Engl J Med 382:266–274.
6. Arsenault-Lapierre G, Kim C, Turecki G (2004): Psychiatric diagnoses in 3275 suicides: A meta-analysis. BMC Psychiatry 4:37.
7. Cavanagh JTO, Carson AJ, Sharpe M, Lawrie SM (2003): Psychological autopsy studies of suicide: A systematic review. Psychol Med 33:395–405.
8. Qin P (2011): The impact of psychiatric illness on suicide: Differences by diagnosis of disorders and by sex and age of subjects. J Psychiatr Res 45:1445–1452.
9. Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, Nock MK (2016): Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. Psychol Med 46:225–236.
10. Ahmedani BK, Peterson EL, Hu Y, Rossom RC, Lynch F, Lu CY, *et al.* (2017): Major physical health conditions and risk of suicide. Am J Prev Med 53:308–315.
11. Khazem LR (2018): Physical disability and suicide: Recent advancements in understanding and future directions for consideration. Curr Opin Psychol 22:18–22.
12. Bernert RA, Nadorff MR (2015): Sleep disturbances and suicide risk. Sleep Med Clin 10:35–39.
13. Woznica AA, Carney CE, Kuo JR, Moss TG (2015): The insomnia and suicide link: Toward an enhanced understanding of this relationship. Sleep Med Rev 22:37–46.
14. Pigeon WR, Pinquart M, Conner K (2012): Meta-analysis of sleep disturbance and suicidal thoughts and behaviors. J Clin Psychiatry 73:e1160–e1167.
15. Bishop TM, Walsh PG, Ashrafioun L, Lavigne JE, Pigeon WR (2020): Sleep, suicide behaviors, and the protective role of sleep medicine. Sleep Med 66:264–270.
16. Qin P, Agerbo E, Mortensen PB (2002): Suicide risk in relation to family history of completed suicide and psychiatric disorders: A nested case-control study based on longitudinal registers. Lancet 360:1126–1130.
17. Chesney E, Goodwin GM, Fazel S (2014): Risks of all-cause and suicide mortality in mental disorders: A meta-review. World Psychiatry 13:153–160.
18. Poorolajal J, Darvishi N (2016): Smoking and suicide: A meta-analysis. PLoS One 11:e0156348.
19. Orri M, Séguin JR, Castellanos-Ryan N, Tremblay RE, Côté SM, Turecki G, Geoffroy MC (2021): A genetically informed study on the association of cannabis, alcohol, and tobacco smoking with suicide attempt. Mol Psychiatry 26:5061–5070.
20. Harrison R, Munafò MR, Davey Smith G, Wootton RE (2020): Examining the effect of smoking on suicidal ideation and attempts: Triangulation of epidemiological approaches. Br J Psychiatry 217:701–707.
21. McMahon K, Hoertel N, Olfson M, Wall M, Wang S, Blanco C (2018): Childhood maltreatment and impulsivity as predictors of interpersonal violence, self-injury and suicide attempts: A national study. Psychiatry Res 269:386–393.
22. Calati R, Ferrari C, Brittner M, Oasi O, Olié E, Carvalho AF, Courtet P (2019): Suicidal thoughts and behaviors and social isolation: A narrative review of the literature. J Affect Disord 245:653–667.
23. Liu RT, Miller I (2014): Life events and suicidal ideation and behavior: A systematic review. Clin Psychol Rev 34:181–192.
24. Voracek M, Loibl LM (2007): Genetics of suicide: A systematic review of twin studies. Wien Klin Wochenschr 119:463–475.
25. Brent DA, Melhem N (2008): Familial transmission of suicidal behavior. Psychiatr Clin North Am 31:157–177.
26. Fu Q, Heath AC, Bucholz KK, Nelson EC, Glowinski AL, Goldberg J, *et al.* (2002): A twin study of genetic and environmental influences on suicidality in men. Psychol Med 32:11–24.
27. Erlangsen A, Appadurai V, Wang Y, Turecki G, Mors O, Werge T, *et al.* (2020): Genetics of suicide attempts in individuals with and without mental disorders: A population-based genome-wide association study. Mol Psychiatry 25:2410–2421.
28. Ruderfer DM, Walsh CG, Aguirre MW, Tanigawa Y, Ribeiro JD, Franklin JC, Rivas MA (2020): Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. Mol Psychiatry 25:2422–2430.
29. Strawbridge RJ, Ward J, Ferguson A, Graham N, Shaw RJ, Cullen B, *et al.* (2019): Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide. EBioMedicine 41:517–525.
30. Docherty AR, Shabalin AA, DiBlasi E, Monson E, Mullins N, Adkins DE, *et al.* (2020): Genome-wide association study of suicide death and polygenic prediction of clinical antecedents. Am J Psychiatry 177:917–927.
31. Mullins N, Bigdeli TB, Børglum AD, Coleman JRI, Demontis D, Mehta D, *et al.* (2019): GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. Am J Psychiatry 176:651–660.
32. Kessler RC, Bromet EJ (2013): The epidemiology of depression across cultures. Annu Rev Public Health 34:119–138.
33. Posner K, Oquendo MA, Gould M, Stanley B, Davies M (2007): Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry 164:1035–1043.
34. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295.
35. Willer CJ, Li Y, Abecasis GR (2010): METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191.
36. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, *et al.* (2020): RICOPILI: Rapid imputation for COnsortias PIpeLIne. Bioinformatics 36:930–933.

37. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, *et al.* (2016): A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48:1279–1283.
38. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, *et al.* (2018): Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun 9:224.
39. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82.
40. Bachmann S (2018): Epidemiology of suicide and the psychiatric perspective. Int J Environ Res Public Health 15:1425.
41. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, *et al.* (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50:668–681.
42. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* (2015): A global reference for human genetic variation. Nature 526:68–74.
43. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, *et al.* (2019): Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 51:793–803.
44. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, *et al.* (2018): Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50:381–389.
45. Hübel C, Gaspar HA, Coleman JRI, Finucane H, Purves KL, Hanscombe KB, *et al.* (2019): Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. Am J Med Genet B Neuropsychiatr Genet 180:428–438.
46. Zheng J, Mesut Erzurumluoglu A, Elsworth BL, Kemp JP, Howe L, Haycock PC, *et al.* (2017): LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33:272–279.
47. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW (2019): Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 10:1776.
48. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4:7.
49. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, *et al.* (2021): Open Targets Genetics: An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nature Genetics 53:1527–1533.
50. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, *et al.* (2020): Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. Psychol Med 50:2435–2443.
51. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, *et al.* (2019): Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 51:245–257.
52. Campos AI, Verweij KJH, Statham DJ, Madden PAF, Maciejewski DF, Davis KAS, *et al.* (2020): Genetic aetiology of self-harm ideation and behaviour. Sci Rep 10:9713.
53. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol 11: e1004219.
54. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, *et al.* (2017): Genetic effects on gene expression across human tissues [published correction appears in Nature 2018; 553:530]. Nature 550:204–213.
55. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, *et al.* (2021): Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. Nature Genetics 53:1311–1321.
56. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, *et al.* (2019): Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet 51:394–403.
57. Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, *et al.* (2012): Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry 17:880–886.
58. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. Science 360:eaap8757.
59. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179:1469–1482.e11.
60. Agerbo E, Nordentoft M, Mortensen PB (2002): Familial, psychiatric, and socioeconomic risk factors for suicide in young people: Nested case-control study. BMJ 325:74.
61. Qin P, Agerbo E, Mortensen PB (2003): Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: A national register–based study of all suicides in Denmark, 1981–1997. Am J Psychiatry 160:765–772.
62. Lorant V, de Gelder R, Kapadia D, Borrell C, Kalediene R, Kovács K, *et al.* (2018): Socioeconomic inequalities in suicide in Europe: The widening gap. Br J Psychiatry 212:356–361.
63. Kaplan ML, Asnis GM, Sanderson WC, Keswani L, De Lecuona JM, Joseph S (1994): Suicide assessment: Clinical interview vs. self-report. J Clin Psychol 50:294–298.

ARTICLE OPEN

Check for updates

Gut microbiota signature in treatment-naïve attention-deficit/hyperactivity disorder

Vanesa Richarte[1,2,3], Cristina Sánchez-Mora[1,2,4], Montserrat Corrales[1,2,3], Christian Fadeuilhe[1,2,3], Laura Vilar-Ribó[1,2,4], Lorena Arribas[1,4], Estela Garcia[4], Silvia Karina Rosales-Ortiz[4,5], Alejandro Arias-Vasquez[6,7], María Soler-Artigas[1,2,4,8], Marta Ribasés[1,2,4,8]✉ and Josep Antoni Ramos-Quiroga[1,2,3,4]✉

Compelling evidence supports alterations in gut microbial diversity, bacterial composition, and/or relative abundance of several bacterial taxa in attention-deficit/hyperactivity disorder (ADHD). However, findings for ADHD are inconsistent among studies, and specific gut microbiome signatures for the disorder remain unknown. Given that previous studies have mainly focused on the pediatric form of the disorder and involved small sample sizes, we conducted the largest study to date to compare the gastrointestinal microbiome composition in 100 medication-naïve adults with ADHD and 100 sex-matched healthy controls. We found evidence that ADHD subjects have differences in the relative abundance of several microbial taxa. At the family level, our data support a lower relative abundance of Gracilibacteraceae and higher levels of Selenomonadaceae and Veillonellaceae in adults with ADHD. In addition, the ADHD group showed higher levels of *Dialister* and *Megamonas* and lower abundance of *Anaerotaenia and Gracilibacter* at the genus level. All four selected genera explained 15% of the variance of ADHD, and this microbial signature achieved an overall sensitivity of 74% and a specificity of 71% for distinguishing between ADHD patients and healthy controls. We also tested whether the selected genera correlate with age, body mass index (BMI), or scores of the ADHD rating scale but found no evidence of correlation between genera relative abundance and any of the selected traits. These results are in line with recent studies supporting gut microbiome alterations in neurodevelopment disorders, but further studies are needed to elucidate the role of the gut microbiota on the ADHD across the lifespan and its contribution to the persistence of the disorder from childhood to adulthood.

Translational Psychiatry (2021)11:382 ; https://doi.org/10.1038/s41398-021-01504-6

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by a persistent pattern of symptoms of inattention, hyperactivity, and impulsivity, resulting in dysfunction in two or more areas of an individual's life [1,2,]. ADHD is associated with deterioration in the social, family, academic, and/or occupational functioning of the affected subjects and has a high impact at the socioeconomic level [3].

The prevalence of ADHD in children is approximately 5.3%, and of these, 50−70% will still show symptoms in adulthood [4]. The etiology is complex and multifactorial, with an average heritability of 74% [5]. Through the largest meta-analyses of genome-wide association studies performed so far, the first genome-wide significant loci for ADHD were identified [6,7,]. Evidence for a strong genetic component of common variants in the polygenic architecture of ADHD was found, with an SNP-based heritability of 22% [6]. Given that the large proportion of heritability still needs to be explained, these data also suggest that other mechanisms may provide a means for integrating the effects of genetic and environmental risk factors and explaining additional phenotypic variance in ADHD. Among such factors, compelling evidence supports a possible role for the gut microbiome in ADHD.

The gut microbiome is essential for health and plays a role in the bidirectional regulation of the brain-gut axis. Microorganisms influence the brain through their ability to produce and modify many metabolic, immunological, and neurochemical factors in the gut that ultimately impact the central nervous system [8–10]; in turn, brain activity also impacts the gut microbiota composition [11,12,]. The gut microbiota influences gut barrier integrity and produce neuroactive compounds such as neurotransmitters, amino acids, and microbial metabolites, including short-chain fatty acids [10,13,]. These metabolites can interact with the host immune system, act on the central nervous system by regulating gene expression, epigenetics and neuroplasticity and affect local

[1]Department of Psychiatry, Mental Health and Addictions, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain. [2]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [3]Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain. [4]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain. [5]Department of Biomedicine, University of Barcelona, Barcelona, Spain. [6]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands. [7]Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands. [8]Department of Genetics, Microbiology & Statistics, University of Barcelona, Barcelona, Spain. ✉email: marta.ribases@vhir.org ; jaramos@vhebron.net

Received: 4 February 2021 Accepted: 24 June 2021
Published online: 08 July 2021

neuronal cells and afferent pathways that signal directly to the brain [13]. This dynamic bidirectional communication between the gut microbiota and the central nervous system influences brain function, cognition, and behavior and highlights the fact that gut microbiota imbalance may contribute to the pathophysiology of neurodevelopmental disorders and mental health outcomes.

Consistently, an increasing number of studies have shown gut microbiome alterations in neurodevelopmental disorders [14–19]. For ADHD, an increasing number of studies have reported that the gut microbial diversity, bacterial composition, and/or relative abundance of several bacterial taxa differ between patients and healthy controls [20–28]. Although not confirmed by others [20,22,27,], some studies found differences in microbiota alpha [23,28,] or beta diversity [23,24,] in ADHD. For example, Prehn-Kristensen et al. observed decreased alpha diversity in ADHD patients and differences in beta diversity between patients and controls [23]. Wang et al. [28] also reported differences in alpha diversity in ADHD, and Szopinska-Tokov found a significant reduction in beta diversity in patients with ADHD [24,28,]. When focusing on specific taxonomic groups, Aarts et al. reported a nominal increase in *Bifidobacterium* in individuals with ADHD, changes that were associated with a significantly enhanced predicted synthesis of the dopamine precursor phenylalanine [22]. Similarly, Jiang et al reported decreased amounts of the genera *Dialister, Lachnoclostridium, Sutterella*, and *Faecalibacterium* in treatment-naïve children with ADHD compared with healthy controls and a negative association between the abundance of the last taxonomic group and parental reports of ADHD symptoms [27]. These results were consistent with those from a recent study by Wan et al., which also detected a reduced relative abundance of *Faecalibacterium*, as well as higher amounts of *Odoribacter* and *Enterococcus*, in ADHD patients [20]. Moreover, Prehn-Kristensen observed distinct abundance in different microbial taxa, including increased *Neisseria* and decreased *Prevotella* and *Parabacteroides* in ADHD subjects [23]. Wang et al. compared the fecal microbiota composition between medication-naïve children with ADHD and healthy controls and found *Fusobacterium* genus as a marker for ADHD as well as enrichment of *Lactobacillus* in the control group [28]. Finally, a recent study conducted by Szopinska-Tokov et al. revealed an association between the relative abundance of the *Ruminococcaceae _UCG_004* genus and ADHD inattention symptoms [24]. All these previous studies, however, considered small sample sizes (from 14 to 51 ADHD patients), mainly focused on the childhood/adolescent form of the disorder, and showed no overlap or lack of concordance between findings.

Additionally, clinical evidence shows that probiotic intervention in early life may improve later outcomes and reduce the risk of neuropsychiatric disorders [29], and mice colonized by microbiota from subjects with ADHD displayed altered microbial composition and behavioral and brain abnormalities compared with mice transplanted with the microbiota from individuals without ADHD [30]. These data further support that the gut microbiome composition may influence brain function and behavior and play a role in the disorder [30–33].

Considering this background, we performed the largest characterization of the gastrointestinal microbiome composition in 100 medication-naïve adults with ADHD and 100 sex-matched healthy controls and assessed differences in the microbiota composition between both groups and whether such differences were associated with ADHD clinical symptoms.

MATERIALS AND METHODS

Participants and clinical assessment

The clinical sample consisted of 100 adult medication-naïve ADHD subjects (DSM-5 criteria) who were referred to an ADHD program from primary care centers and adult community mental health services. All subjects were evaluated and recruited prospectively from a restricted geographic area of Catalonia (Spain) in a specialized outpatient program for adult ADHD and by a single clinical group at Hospital Universitari Vall d'Hebron of Barcelona (Spain). A description of the sample is provided in Supplementary Table 1.

The clinical assessment consisted of structured interviews and self-report questionnaires in two different steps: (i) ADHD diagnosis was based on the results of the Structured Diagnostic Interview for Adult ADHD (DIVA 2.0) [34] by a psychiatrist; (ii) the severity of ADHD symptoms and levels of impairment and comorbid disorders were assessed by a psychologist. In this part of the evaluation, the following scales and questionnaires were used: the ADHD Rating Scale (ADHD-RS), the Clinical Global Impression (CGI), the Wender Utah Rating Scale (WURS), the Sheehan Disability Inventory (SDS), and the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II). Afterward, the psychiatrist and psychologist integrated the clinical information and self-reports for valid assessment of symptoms and impairments. In case of discordance between the different raters for ADHD symptoms or inconsistencies between the reporters in responses to items measuring similar symptoms, clinician-identified symptoms on the DIVA 2.0 prevailed. Clinical information was reordered at the moment of inclusion, at which time the stool specimen was also collected. Exclusion criteria were as follows: an intelligence quotient less than 70; lifelong or current history of mood, psychotic, anxiety, substance abuse, and personality disorders; pervasive developmental disorders; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight ≤ 1.5 kg; and other neurological or systemic disorders that might explain ADHD symptoms.

The control sample consisted of 100 unrelated healthy donors matched by sex and ethnicity with the clinical group. The exclusion criteria were ADHD symptomatology according to the Adult Self-Report Scale A.S.R. S v1.1. and any prior or current psychiatric comorbidity.

All subjects reported European ancestry, which was confirmed through principal component analysis (PCA) using genetic data. Exclusion criteria for all participants included treatment with antibiotics or probiotics up to before stool collection.

The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron. All methods were performed in accordance with the relevant guidelines and regulations, and written informed consent was obtained from all subjects before inclusion. None of the participants received any financial compensation.

Sample collection and DNA isolation

Human fecal samples were collected at home, stabilized with the OMNIgene·GUT (OM-200) (DNA Genotek Inc.) kit, and then transported to the laboratory. The samples were aliquoted into 1.5-ml tubes and stored at −80 °C. Microbial DNA was purified from 200 mg of each homogenized fecal sample using the QIAamp® PowerFecal® DNA extraction kit (QIAgen, Hilden, Germany). The isolated DNA was quantified using PicoGreen™ dsDNA Assay Kit [35].

Library preparation and Illumina sequencing

The V3−V4 hypervariable region of the bacterial 16S rRNA gene was amplified for microbiome composition profiling. DNA library construction was performed following the manufacturer's instructions (Illumina). We used the same workflow as described elsewhere [36] to perform cluster generation, template hybridization, isothermal amplification, linearization, blocking and denaturation, and hybridization of the sequencing primers. Briefly, the V3−V4 region was amplified using key-tagged eubacterial primers 5′CCTACGGGNGGCWGCAG3′ and 5′GACTACHVGGGTATC-TAATCC3′, and 300-nt paired-end amplicons were subsequently sequenced in two different rounds using the Illumina MiSeq platform. The raw Illumina paired-end reads were merged considering an overlap length > 70 bp with the PEAR software v. 0.9.1, providing a single FASTQ file for each of the samples [37]. High-quality reads were extracted by applying a minimum Phred score of 20 (Q20, 99% based call accuracy). After primer sequences trimming, reads without both primer sequences or with less than 200 bp were discarded with Cutadapt v.1.8.1 [38]. Chimeric sequences were removed using the UCHIME software [39]. After quality control filtering, we obtained 14.7 million high-quality sequences with 45 063−216 059 reads per sample from a total of 200 fecal samples. The raw and clean number of sequences, mean length, total mega bases sequenced, and mean quality per sample can be found in Supplementary Table 2. The remaining reads were clustered into operational taxonomic units (OTUs), in which unique sequences with a relative abundance above

0.1% were clustered into OTUs based on 97% sequence similarity [40] using the CD-HIT package [41] and the BLAST search against the NCBI 16S rRNA reference database (September 2019) with bastn v.2.10.0+. Taxonomic groups (phylum, family, and genus) were assigned with a Python script developed by ADM-BIOPOLIS (Paterna, Valencia, Spain). To remove genera with absent or low prevalence, the OTU table was filtered at the genus level. OTUs with nonzero values in less than 10% of the samples were removed. OUT counts were normalized by rarefaction with the phyloseq R package according to Weiss et al. [42].

Statistical analysis

Alpha diversity (within-sample diversity) was calculated on rarefied data with the Richness, Simpson and Shannon diversity indices and compared between individuals with ADHD and controls using the vegan R package (https://github.com/vegandevs/vegan). Beta diversity (between-sample diversity) was calculated by weighted and unweighted UniFrac and Bray Curtis distances, as represented by two-dimensional principal coordinates analysis (PCoA) plots, and compared between groups by permutation multivariate analysis of variance (PERMANOVA) using the phyloseq R package [43]. The local contribution to beta diversity (LCBD) test was applied to evaluate the contribution of each sample to the diversity between the groups using the adespatial R package (https://github.com/sdray/adespatial). Canonical correspondence analysis (CCA), a multivariate constrained ordination method, on rarefied OTUs was performed and significance regarding the microbial community composition between groups was assessed by permutational multivariate analysis of variance (ADONIS) using the vegan R package (https://github.com/vegandevs/vegan).

Differential abundance comparisons between groups were assessed in taxonomic groups showing an average of normalized counts (baseMean) > 10 using the DESeq2 and randomForest R packages for the classification, rfUtilities to estimate the significance of the classification and rfPermute to evaluate the significance of specific taxa, with 1000 permutations. All comparisons were performed at the phylum, family, and genus levels. Any unknown taxonomic level was assigned to the next highest known taxonomic rank.

Genera showing significant differences in relative abundance between ADHD cases and controls after multiple comparison corrections in DeSeq2 and the random forest comparisons were considered for downstream analyses. Multiple logistic regression models were applied to test the association between ADHD and all selected genera while adjusting for age, sex, and body mass index (BMI). Adjusted Pseudo-R2 was calculated with the McFaddenAdj method and the DescTools R package (https://github.com/AndriSignorell/DescTools); sensitivity and specificity were calculated with the caret R package (https://github.com/topepo/caret/). A likelihood ratio test with the lmtest R package (https://cran.r-project.org/web/packages/lmtest/) was employed to assess whether the inclusion of selected genera in the multiple logistic regression model fits the data significantly better than the model including only age, sex, and BMI. In the first model, we considered affectation status as dependent variable and age, sex and BMI as independent variables (ADHD ~ age + sex + BMI); in the second model, we included selected taxa as independent variables (ADHD ~ age + sex + BMI + *Megamonas* + *Anaerotaenia* + *Gracilibacter* + *Dialister)*. Spearman correlation tests were used to assess correlations between selected genera, age, BMI, and inattention and hyperactive/impulsivity subscale scores or total scores of the ADHD rating scale.

RESULTS

Bacterial composition based on 16S rRNA sequencing was available for 100 adult ADHD cases and 100 controls. No differences in intestinal microbial alpha diversity (microbial community richness and evenness) were found between ADHD cases and controls when measured by three different indices (Richness, Simpson, or Shannon indices; Supplementary Fig. 1). Beta diversity (between-sample community dissimilarity) according to weighted and unweighted UniFrac distances as well as the Bray-Curtis dissimilarity index showed no differences in the microbial composition between the groups (PERMANOVA P-value>0.05), with no evidence of separate clustering in PCoA representations (Supplementary Fig. 2). No significant differences in the gut microbiota composition between the ADHD and control groups were observed in the CCA either (ADONIS *P*-value = 0.31; Supplementary Fig. 3).

Compositional analysis of samples revealed that *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Actinobacteria*, and *Verrucomicrobia* were the most abundant phyla in our sample of 200 subjects (Supplementary Table 3), with no significant differences in relative abundance detected for any of them. When we explored the relative abundance of specific microbial taxa, however, we found evidence that several taxa differed significantly between ADHD cases and controls by two different methods, DeSeq2 and/or random forests: 1 phylum, 7 families, and 17 genera showed differential abundance ($P_{FDR} < 0{,}05$; DESeq2: 1 phylum, 5 families, and 15 genera; random forests: 5 families and 6 genera; Table 1, Fig. 1, Supplementary Table 4 and Supplementary Fig. 4). When combining the results of both methods, we found overlap for three families (Gracilibacteraceae, Selenomonadaceae, and Veillonellaceae) and four genera (*Anaerotaenia, Dialister, Gracilibacter, and Megamonas)* (Table 1 and Fig. 1).

For downstream analysis, we focused on genera that differed in relative abundance between ADHD and controls with both of the methods described above (*Anaerotaenia, Dialister, Gracilibacter, and Megamonas)*. When we assessed whether they correlated with each other, we found a moderate correlation between *Anaerotaenia* and *Gracilibacter* ($r = 0.35$; *P*-value = 3.6e−04), a weak correlation between *Anaerotaenia* and *Megamonas* ($r = -0.24$; *P*-value = 0.018), and no correlation between the others (Fig. 2). A model including the four genera and the covariates age, sex, and BMI explained 15% of the variance in ADHD, with significant improvement of the model which included only the covariates (*P*-value = 8.2e−07), which explained 5.9% of the variance (Supplementary Table 5). The microbial signature achieved an overall sensitivity of 74% and a specificity of 71% for the detection of individuals with ADHD versus healthy controls. We also assessed whether the selected genera correlated with age, BMI, or ADHD rating scale scores but found no evidence of correlation between relative abundance and any of the selected traits (Fig. 2).

DISCUSSION

To clarify the relationship between ADHD and the gut microbiome, we performed the largest study to date and compared the microbial composition between 100 medication-naïve adults with ADHD and 100 sex-matched unrelated healthy subjects. We found evidence that ADHD subjects exhibit differences in the relative abundance of several microbial taxa. At the family level, our data support a lower relative abundance of Gracilibacteraceae and higher levels of Selenomonadaceae and Veillonellaceae in adults with ADHD. In addition, the ADHD group showed higher levels of *Dialister* and *Megamonas* and lower abundances of *Anaerotaenia and Gracilibacter* at the genus level.

These results are in line with recent studies supporting gut microbiome differences in neurodevelopmental disorders. Although the mechanistic explanation for these associations remains unknown, a positive correlation between *Dialister* abundance and activity level has been described in toddlers [44]. Additionally, decreased levels of *Dialister* were found in autism spectrum disorder (ASD) patients [45,46,] or in treatment-naïve children with ADHD [27] compared with healthy controls and in ADHD individuals on medication compared with non-medicated individuals [24]. Furthermore, multiple taxonomic groups that differed in relative abundance between ADHD cases and controls in the present study, including Selenomonadaceae, Veillonellaceae, and *Megamonas*, have previously been associated with other psychiatric conditions that often coexist with ADHD, such as ASD or depression [27, 47–51]. Given that the ADHD subjects in this study displayed no comorbid psychiatric disorders, we cannot discount a possible pleotropic effect of these

Table 1. Summary of differential abundance results between ADHD patients and controls considering Deseq2 and random forest results.

		Relative abundance (% mean (SD)		Adjusted *P*-value	
		ADHD	**Controls**	**DEseq2**	**Random forests**
Phylum	*Candidatus Melainabacteria*	0.072 (0.24)	0.22 (0.76)	3.1E−03	0.11
Family	Eubacteriaceae	2.105 (1.51)	2.269 (1.35)	0.81	0.02
	Gracilibacteraceae	0,503 (0.85)	0,949 (1.49)	**0.035**	**0.05**
	Lactobacillaceae	0.965 (1.52)	1.077 (1.24)	0.93	0.02
	Peptostreptococcaceae	0,327 (0.55)	0,199 (0.23)	0.016	0.27
	Selenomonadaceae	0,387 (1.14)	0,071 (0.26)	**3.5E−07**	**0.05**
	Veillonellaceae	1,658 (1.90)	0,837 (1.43)	**0.012**	**9.9E−03**
	Verrucomicrobiaceae	0,036 (0.11)	0,063 (0.17)	0.012	0.73
Genus	*Acetivibrio*	0.021 (0.05)	0.056 (0.17)	6.1E−03	0.099
	Alloprevotella	0.380 (1.63)	0.182 (0.97)	4.4E−04	0.21
	Anaerotaenia	**0.072 (0.13)**	**0.248 (0.49)**	**2.3E−09**	**9.9E−03**
	Dialister	**1.377 (1.76)**	**0.649 (1.26)**	**0.041**	**0.02**
	Flintibacter	1.967 (1.46)	1.588 (1.37)	0.26	0.045
	Fucophilus	0.036 (0.11)	0.064 (0.17)	0.012	0.42
	Gracilibacter	**0.509 (0.86)**	**0.958 (1.50)**	**0.040**	**9.9E−03**
	Herbinix	0.024 (0.05)	0.042 (0.08)	0.024	0.24
	Leclercia	0.084 (0.42)	0.025 (0.12)	9.8E−03	0.30
	Megamonas	**0.323 (1.04)**	**0.029 (0.20)**	**3.2E−29**	**9.9E−03**
	Megasphaera	0.209 (0.72)	0.091 (0.42)	7.5E−20	0.80
	Odoribacter	0.547 (0.34)	0.751 (0.83)	0.039	0.14
	Parasutterella	0.751 (1.30)	1.588 (1.37)	0.70	9.9E−03
	Porphyromonas	0.129 (0.52)	0.110 (0.55)	6.1E−03	0.36
	Prevotellamassilia	0.356 (1.82)	0.340 (1.69)	6.4E−15	0.31
	Romboutsia	0.228 (0.52)	0.126 (0.16)	9.8E−03	0.93
	Vampirovibrio	0.073 (0.24)	0.225 (0.77)	2.6E−03	0.38

Differentially abundant taxa identified by both methods, DEseq2 and random forests, are shown in bold.

taxonomic groups and that their relative abundance may explain, in part, ADHD phenotypic variability.

Although previous gut microbiome analyses on ADHD have mainly focused on pediatric samples [20,23, 26–28] and there is limited research on adults [22,24,], we focused our study on adulthood ADHD. Nevertheless, given that the gut microbiome evolves throughout the lifespan [16,52,53,], whether early-life exposure to environmental risk factors contributes to the gut microbiota and impacts neurodevelopment and mental health outcomes later in life remain to be investigated. Further longitudinal studies are warranted to provide additional information on the role of the microbiome in ADHD symptom trajectories from childhood to adulthood as well as mental health outcomes and comorbid profiles across the lifespan.

We did not detect substantial changes in alpha or beta diversity between ADHD cases and controls. The high heterogeneity in terms of age, sample size, sex, clinical characteristics, and type of controls may explain nonreplicable results and discrepancies between studies. We sex-matched ADHD cases and controls and restricted the clinical sample to ADHD medication-naïve adult subjects, which is a major strength of our study design that may allow us to identify an imbalance in the gut microbiome composition that might be neglected by broader study designs. In addition, the sample sizes of previous studies on ADHD, were relatively small; although our study may also have limited statistical power to estimate the magnitude of the differences identified in microbial relative abundance, we assessed the largest sample size considered thus far. The results, however, need to be interpreted with caution given that we selected genera of interest and estimated the variance in ADHD explained by these taxa as well as the sensitivity and specificity of the regression model using the same dataset, which may have led to overfitting and further support the use of independent datasets to obtain more accurate estimates.

Microbiome composition is strongly influenced by environmental factors such as diet, overall health status, and medication use [52, 54–56]. The participants in this study were not on medication and had not used antibiotics or probiotics in the three months before sample collection, which may not explain the differences detected between ADHD cases and controls. Nonetheless, no other environmental exposures, including smoking, stress, dietary habits, or other lifestyle information, that may have an effect on microbiota composition were considered. For instance, animal models and population-based cross-sectional studies support an effect of nicotine or smoking status on the gut microbiome composition and the fecal metabolome [57–59]. In addition to environmental factors, consistent evidence suggests that the host genetic background impacts the composition of gut microbial communities and that genetic factors influence microbiome composition and explain a significant proportion of the variation in the gut microbiome [60–63]. Hence, further integrative studies considering multiple data sources (i.e., larger sample sizes), including environmental factors, human genetic variation, and gut microbial composition, are warranted to provide deeper insight into the mechanisms underlying the relationship between the microbiota, host genetics, and individual habits, and behavior, as

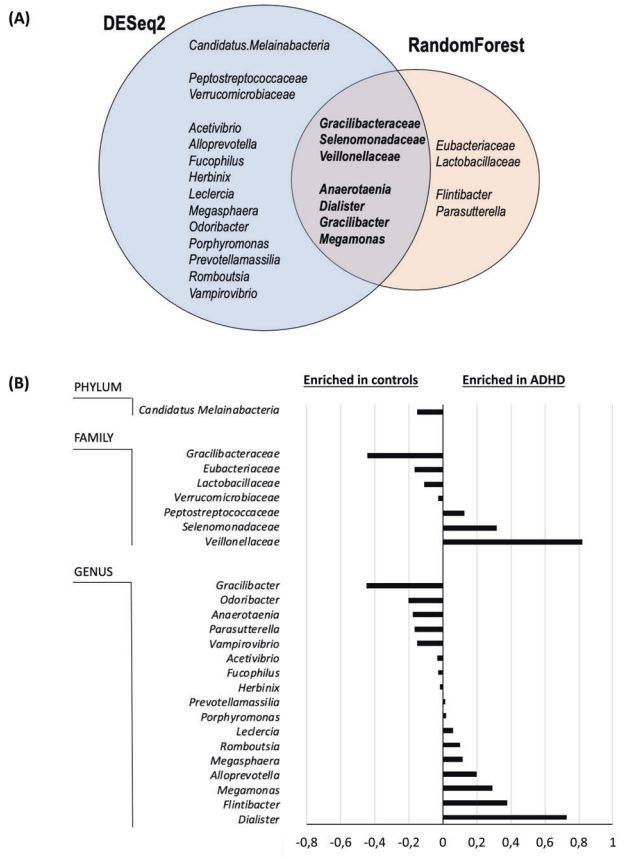


Fig. 1 Differentially abundant taxa between ADHD cases and controls. (**A**) Differential abundance results according to two different methods, DESeq2 andrandom forests. **B** Differences in relative abundance between ADHD cases and controls for taxonomic groups surpassing multiple comparison corrections in DeSeq2 and/or random forest analyses.

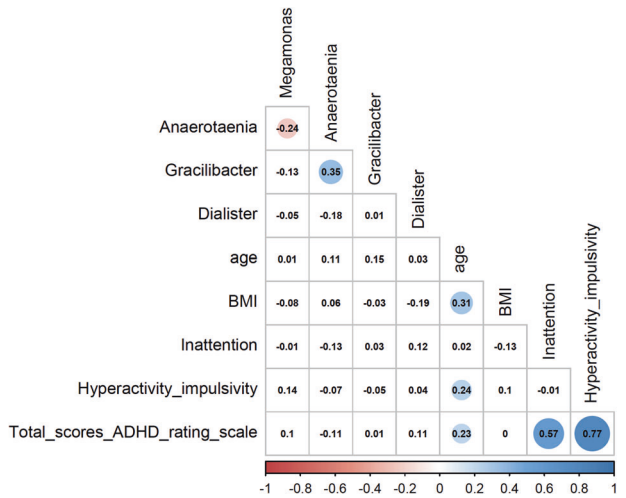


Fig. 2 Spearman correlation between the *relative abundance of* four bacterial genera (*Anaerotaenia, Dialister, Gracilibacter,* and *Megamonas)* and age, BMI, and ADHD rating scale scores. Colored correlations are statistically significant (*P*-value < 0.05), with positive and negative correlations in blue and red, respectively. Inattention: score of the inattention subscale of the ADHD rating scale; hyperactivity_impulsivity: score of the hyperactive/impulsivity subscale of the ADHD rating scale; total: total scores of the ADHD rating scale.

well as their roles in ADHD and other neurodevelopmental disorders across the lifespan.

DATA AVAILABILITY

The datasets supporting the conclusions of this article are included and available online. Raw fastq data will be available upon request to the corresponding author.

REFERENCES

1. Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Prim. 2015;1:15020. https://doi.org/10.1038/nrdp.2015.20.
2. Franke B, Michelini G, Asherson P, Banaschewski T, Bilbow A, Buitelaar JK. et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. Eur. Neuropsychopharmacol.2018;28:1059–88. https://doi.org/10.1016/j.euroneuro.2018.08.001.
3. Thapar A, Cooper M. Attention deficit hyperactivity disorder. Lancet. 2016;387:1240–50. https://doi.org/10.1016/S0140-6736(15)00238-X.
4. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am. J. Psychiatry. 2007;164:942–948. https://doi.org/10.1176/ajp.2007.164.6.942.
5. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. Mol Psychiatry. 2019;24:562−575. https://doi.org/10.1038/s41380-018-0070-0.
6. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet.2019;51:63–75. https://doi.org/10.1038/s41588-018-0269-7.
7. Rovira P, Demontis D, Sánchez-Mora C, Zayats T, Klein M, Mota NR. et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology. 2020;45:1617–26. https://doi.org/10.1038/s41386-020-0664-5.
8. Erny D, Hrabe de Angelis AL, Jaitin D, Wieghofer P, Staszewski O, David E. et al. Host microbiota constantly control maturation and function of microglia in the CNS HHS Public Access Author manuscript. Nat Neurosci.2015;18:965–77. https://doi.org/10.1038/nn.4030.
9. Clarke G, Grenham S, Scully P, Fitzgerald P, Moloney RD, Shanahan F. et al. The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. Mol Psychiatry. 2013;18:666–73. https://doi.org/10.1038/mp.2012.77.
10. Morais LH, Schreiber HL, Mazmanian SK. The gut microbiota–brain axis in behaviour and brain disorders. Nat Rev Microbiol.2020;19:241–55. https://doi.org/10.1038/s41579-020-00460-0.
11. Kuwahara, A. et al. Microbiota-gut-brain axis: enteroendocrine cells and the enteric nervous system form an interface between the microbiota and the central nervous system. Biomed Res. 2020;41:199−216.
12. Mohajeri MH, La Fata G, Steinert RE, Weber P. Relationship between the gut microbiome and brain function. Nutr Rev.2018;76:481–96. https://doi.org/10.1093/nutrit/nuy009.
13. Dalile B, Van Oudenhove L, Vervliet B, Verbeke K. The role of short-chain fatty acids in microbiota–gut–brain communication. Nat Rev Gastroenterol Hepatol.2019;16:461–78. https://doi.org/10.1038/s41575-019-0157-3.
14. Dam SA, Mostert JC, Szopinska-Tokov JW, Bloemendaal M, Amato M, Arias-Vasquez A. The role of the gut-brain axis in attention-deficit/hyperactivity disorder. Gastroenterol Clin North Am.2019;48:407–31. https://doi.org/10.1016/j.gtc.2019.05.001.
15. Ma Q, Xing C, Long W, Wang HY, Liu Q, Wang RF. Impact of microbiota on central nervous system and neurological diseases: the gut-brain axis. J. Neuroinflammation.2019;16:1–14. https://doi.org/10.1186/s12974-019-1434-3.
16. Borre YE, O'Keeffe GW, Clarke G, Stanton C, Dinan TG, Cryan JF. Microbiota and neurodevelopmental windows: implications for brain disorders. Trends Mol. Med.2014;20:509–18. https://doi.org/10.1016/j.molmed.2014.05.002.
17. Jurek, L. et al. Is there a dysbiosis in individuals with a neurodevelopmental disorder compared to controls over the course of development? A systematic review. Eur Child Adolesc Psychiatry. 2020. https://doi.org/10.1007/s00787-020-01544-1.
18. Rogers GB, Keating DJ, Young RL, Wong ML, Licinio J, Wesselingh S. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. Mol. Psychiatry. 2016;21:738–48. https://doi.org/10.1038/mp.2016.50.
19. Lacorte, E. et al. A systematic review of the microbiome in children with neurodevelopmental disorders. Front Neurol. 2019;10:727 https://doi.org/10.3389/fneur.2019.00727.
20. Wan, L. et al. Case-control study of the effects of gut microbiota composition on neurotransmitter metabolic pathways in children with attention deficit hyperactivity disorder. Front Neurosci. 2020;14:127. https://doi.org/10.3389/fnins.2020.00127.

21. Boonchooduang N, Louthrenoo O, Chattipakorn N, Chattipakorn SC. Possible links between gut–microbiota and attention-deficit/hyperactivity disorders in children and adolescents. Eur J Nutr.2020;59:3391–403. https://doi.org/10.1007/s00394-020-02383-1.
22. Aarts E, Ederveen T, Naaijen J, Zwiers MP, Boekhorst J, Timmerman HM. et al. Gut microbiome in ADHD and its relation to neural reward anticipation. PLoS One. 2017;12:e0183509 https://doi.org/10.1371/journal.pone.0183509.
23. Prehn-Kristensen A, Zimmermann A, Tittmann L, Lieb W, Schreiber S, Baving L. et al. Reduced microbiome alpha diversity in young patients with ADHD. PLoS One. 2018;13:e0200728 https://doi.org/10.1371/journal.pone.0200728.
24. Szopinska-Tokov J, Dam S, Naaijen J, Konstanti P, Rommelse N, Belzer C. et al. Investigating the gut microbiota composition of individuals with attention-deficit/hyperactivity disorder and association with symptoms. Microorganisms. 2020;8:406 https://doi.org/10.3390/microorganisms8030406.
25. Checa-Ros A, Jer A, Molina-Carballo A, Campoy C, Muñoz-Hoyos A. Current evidence on the role of the gut microbiome in ADHD: pathophysiology and therapeutic implications. Nutrients. 2021;13:1–32. https://doi.org/10.3390/nu13010249.
26. Sukmajaya A. C., Lusida, M. I., Soetjipto & Setiawati, Y. Systematic review of gut microbiota and attention-deficit hyperactivity disorder (ADHD). Ann Gen Psychiatry. 2021;20:12 https://doi.org/10.1186/s12991-021-00330-w.
27. Jiang HY, Zhou YY, Zhou GL, Li YC, Yuan J, Li XH. et al. Gut microbiota profiles in treatment-naïve children with attention deficit hyperactivity disorder. Behav Brain Res.2018;347:408–13. https://doi.org/10.1016/j.bbr.2018.03.036.
28. Wang LJ, Yang CY, Chou WJ, Lee MJ, Chou MC, Kuo HC. et al. Gut microbiota and dietary patterns in children with attention-deficit/hyperactivity disorder. Eur Child Adolesc Psychiatry. 2020;29:287–97. https://doi.org/10.1007/s00787-019-01352-2.
29. Pärtty A, Kalliomäki M, Wacklin P, Salminen S, Isolauri E. A possible link between early probiotic intervention and the risk of neuropsychiatric disorders later in childhood: a randomized trial. Pediatr Res.2015;77:823–8. https://doi.org/10.1038/pr.2015.51.
30. Tengeler AC, Dam SA, Wiesmann M, Naaijen J, van Bodegom M, Belzer C. et al. Gut microbiota from persons with attention-deficit/hyperactivity disorder affects the brain in mice. Microbiome. 2020;8:44 https://doi.org/10.1186/s40168-020-00816-x.
31. De Theije CG, Wopereis H, Ramadan M, van Eijndthoven T, Lambert J, Knol J. et al. Altered gut microbiota and activity in a murine model of autism spectrum disorders. Brain Behav Immun.2014;37:197–206. https://doi.org/10.1016/j.bbi.2013.12.005.
32. Sharon G, Cruz NJ, Kang DW, Gandal MJ, Wang B, Kim YM. et al. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. Cell . 2019;177:1600–e17. https://doi.org/10.1016/j.cell.2019.05.004.
33. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T. et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell. 2013;155:1451–63. https://doi.org/10.1016/j.cell.2013.11.024.
34. Ramos-Quiroga, J. A. et al. Criteria and concurrent validity of DIVA 2.0: a semi-structured diagnostic interview for adult ADHD. J Atten Disord. 2016;23:1−10. https://doi.org/10.1177/1087054716646451.
35. Mardis E, McCombie WR. Library quantification using picogreen fluorometry. Cold Spring Harb Protoc.2017;2017:432–5. https://doi.org/10.1101/pdb.prot094722.
36. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res.2013;41:1 https://doi.org/10.1093/nar/gks808.
37. Zhang J, Kobert K, Flouri T, Stamatakis A. Genome analysis PEAR: a fast and accurate illumina paired-end reAd mergeR. Bioinformatics.2014;30:614–20. https://doi.org/10.1093/bioinformatics/btt593.
38. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–12.
39. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194–2200. https://doi.org/10.1093/bioinformatics/btr381.
40. Goebel BM, Stackebrandt E. Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. Appl Environ Microbiol.1994;60:1614–21. https://doi.org/10.1128/aem.60.5.1614-1621.1994.
41. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9. https://doi.org/10.1093/bioinformatics/btl158.
42. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5:27. https://doi.org/10.1186/s40168-017-0237-y.
43. Mcmurdie, P. J. & Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome ceus data. PLoS ONE. 8(4):e61217. https://doi.org/10.1371/journal.pone.0061217.
44. Christian LM, Galley JD, Hade EM, Schoppe-Sullivan S, Kamp Dush C, Bailey MT. Gut microbiome composition is associated with temperament during early childhood. Brain Behav Immun.2015;45:118–27. https://doi.org/10.1016/j.bbi.2014.10.018.
45. Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J. et al. New evidences on the altered gut microbiota in autism spectrum disorders. Microbiome. 2017;5:1–11. https://doi.org/10.1186/s40168-017-0242-1.
46. Finegold SM, Dowd SE, Gontcharova V, Liu C, Henley KE, Wolcott RD. et al. Pyrosequencing study of fecal microflora of autistic and control children. Anaerobe. 2010;16:444–53. https://doi.org/10.1016/j.anaerobe.2010.06.008.
47. Kelly JR, Minuto C, Cryan JF, Clarke G, Dinan TG. Cross talk: the microbiota and neurodevelopmental disorders. Front Neurosci.2017;11:490 https://doi.org/10.3389/fnins.2017.00490.
48. Liu, S. et al. Altered gut microbiota and short chain fatty acids in Chinese children with autism spectrum disorder. Sci Rep. 2019;9:287. https://doi.org/10.1038/s41598-018-36430-z.
49. Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y. et al. Altered fecal microbiota composition in patients with major depressive disorder. Brain Behav Immun.2015;48:186–94. https://doi.org/10.1016/j.bbi.2015.03.016.
50. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X. et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatry. 2016;21:786–96. https://doi.org/10.1038/mp.2016.44.
51. Zou R, Xu F, Wang Y, Duan M, Guo M, Zhang Q. et al. Changes in the gut microbiota of children with autism spectrum disorder. Autism Res.2020;13:1614–25. https://doi.org/10.1002/aur.2358.
52. Hasan N, Yang H. Factors affecting the composition of the gut microbiota, and its modulation. PeerJ. 2019;2019:1–31. https://doi.org/10.7717/peerj.7502.
53. O'Toole PW. Changes in the intestinal microbiota from adulthood through to old age. Clin Microbiol Infect.2012;18:44–46. https://doi.org/10.1111/j.1469-0691.2012.03867.x.
54. Flandroy L, Poutahidis T, Berg G, Clarke G, Dao MC, Decaestecker E. et al. The impact of human activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems. Sci Total Environ.2018;627:1018–38. https://doi.org/10.1016/j.scitotenv.2018.01.288.
55. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K. et al. Population-level analysis of gut microbiome variation. Science. 2016;352:560–4. https://doi.org/10.1126/science.aad3503.
56. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352:565–9. https://doi.org/10.1126/science.aad3369.
57. Chi L, Mahbub R, Gao B, Bian X, Tu P, Ru H. et al. Nicotine alters the gut microbiome and metabolites of gut-brain interactions in a sex-specific manner. Chem Res Toxicol.2017;30:2110–9. https://doi.org/10.1021/acs.chemrestox.7b00162.
58. Lee SH, Yun Y, Kim SJ, Lee EJ, Chang Y, Ryu S. et al. Clinical medicine association between cigarette smoking status and composition of gut microbiota: population-based cross-sectional study. J Clin Med.2018;7:282 https://doi.org/10.3390/jcm7090282.
59. Biedermann L, Zeitz J, Mwinyi J, Sutter-Minder E, Rehman A, Ott SJ. et al. Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. PLoS One. 2013;8:e59260 https://doi.org/10.1371/journal.pone.0059260.
60. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med.2018;24:392–400. https://doi.org/10.1038/nm.4517.
61. Sommer F, Bäckhed F. The gut microbiota-masters of host development and physiology. Nat Rev Microbiol.2013;11:227–38. https://doi.org/10.1038/nrmicro2974.
62. Awany, D. et al. Host and microbiome genome-wide association studies: current state and challenges. Front Genet. 2019;10:637. https://doi.org/10.3389/fgene.2018.00637.
63. Hughes DA, Bacigalupe R, Wang J, Rühlemann MC, Tito RY, Falony G. et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat Microbiol.2020;5:1079–87. https://doi.org/10.1038/s41564-020-0743-8.

ACKNOWLEDGEMENTS

Over the course of this investigation C.S.M. was a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199), M.R. was a recipient of a Miguel de Servet

contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023), and M.S.A. was a recipient of a contract from the Biomedical Network Research Center on Mental Health (CIBERSAM), Madrid, Spain, and a recipient of a Juan de la Cierva Incorporación contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I). The research leading to these results received funding from the European Union H2020 Program (H2020/2014-20) under grant agreements no. 667302 (CoCA) and no. 728018 (Eat2beNICE), from the Instituto de Salud Carlos III (PI15/01789, PI16/01505, PI17/0289, PI18/01788, PI19/00721,P19/01224 and PI20/00041), from the Pla estratègic de recerca i innovació en salut (PERIS); Generalitat de Catalunya (METAL-Cat; SLT006/17/287) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461) and cofinanced by the European Regional Development Fund (ERDF) and by "la Marató de TV3" (092330/31). The work was also supported by the ECNP Network 'ADHD across the Lifespan' (https://www.ecnp.eu/researchinnovation/ECNP-networks/List-ECNP-Networks/).

COMPETING INTERESTS

Dra. Sánchez-Mora, Dra. Soler Artigas, Estela García, and Dra. Ribasés report no biomedical financial interests or potential conflicts of interest. Vanesa Richarte has served as a speaker for Rubió and Shire/Takeda in the last 5 years. She has received travel awards from Shire/Takeda for participating in psychiatric meetings. The ADHD Program received unrestricted educational and research support from Eli Lilly and Co., Janssen-Cilag, Shire/Takeda, Rovi, Psious, and Laboratorios Rubió in the past two years. Dra. Corrales received travel awards from Shire for participating in psychiatric meetings. Christian Fadeuilhe received travel awards from Rubió, Shire/Takeda, and Lundbeck for participating in psychiatric meetings. Prof. Ramos-Quiroga was on the speakers' bureau and/or acted as a consultant for Eli-Lilly, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogui, Lundbeck, Almirall, Braingaze, Sincrolab, Medice, and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) from Janssen-Cilag, Rubió, Shire, Takeda, Shionogui, Bial, Medice, and Eli-Lilly for participating in psychiatric meetings. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: EliLilly, Lundbeck, Janssen- Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió.

ADDITIONAL INFORMATION

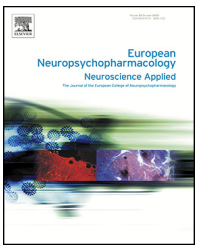
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-021-01504-6.

Correspondence and requests for materials should be addressed to M.R. or J.A.R-Q.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

European Neuropsychopharmacology (2020) 41, 160–166



www.elsevier.com/locate/euroneuro

SHORT COMMUNICATION

Transcriptome profiling in adult attention-deficit hyperactivity disorder

Niall Mortimer[a,b,c], **Cristina Sánchez-Mora**[a,b,d,e], **Paula Rovira**[a,b], **Laura Vilar-Ribó**[a,b], **Vanesa Richarte**[b,e,f], **Montse Corrales**[b,e,f], **Christian Fadeuilhe**[b,e,f], **Olga Rivero**[c], **Klaus-Peter Lesch**[c,g,h], **Miguel Casas**[a,b,e,f], **Josep Antoni Ramos-Quiroga**[a,b,e,f], **María Soler Artigas**[a,b,d,e,*], **Marta Ribasés**[a,b,d,e,*]

[a] *Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Passeig Vall d'Hebron, 119-129, Barcelona 08035, Spain*
[b] *Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain*
[c] *Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Germany*
[d] *Department of Genetics, Microbiology & Statistics, University of Barcelona, Barcelona, Spain*
[e] *Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain*
[f] *Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain*
[g] *Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University, Russia*
[h] *Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Netherlands*

Received 1 April 2020; received in revised form 4 October 2020; accepted 2 November 2020

KEYWORDS
Attention-deficit/hyperactivity disorder (ADHD);
Transcriptomic assays;
Gene-expression signatures

Abstract
Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder with an estimated heritability of around 70%. Although the largest genome-wide association study (GWAS) meta-analysis on ADHD identified independent loci conferring risk to the disorder, the molecular mechanisms underlying the genetic basis of the disorder remain to be elucidated. To explore ADHD biology, we ran a two-step transcriptome profiling in peripheral blood mononuclear cells (PBMCs) of 143 ADHD subjects and 169 healthy controls. Through this exploratory study we

* Corresponding authors.
E-mail addresses: maria.soler@vhir.org (M.S. Artigas), marta.ribases@vhir.org (M. Ribasés).

https://doi.org/10.1016/j.euroneuro.2020.11.005

found eight differentially expressed genes in ADHD. These results highlight promising candidate genes and gene pathways for ADHD and support the use of peripheral tissues to assess gene expression signatures for ADHD.

1. Introduction

The Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM5) defines attention-deficit/hyperactivity disorder (ADHD) as a clinically heterogeneous neurodevelopmental disorder characterized by inattention, hyperactivity and abnormal levels of impulsivity/emotionality (American Psychiatric Association, 2013). It has a childhood prevalence of around 3.4% (Polanczyk et al., 2015) and 2.5-3.4% in adults (Franke et al., 2018). Genetic factors play a critical role in the etiology of ADHD, with meta-analysis of multiple large scale twin studies estimating heritability of 70–80% for childhood ADHD (Nikolas and Burt, 2010) and around 70% for clinically diagnosed ADHD in adults (Faraone and Larsson, 2019).

A recent GWAS meta-analysis in 20,183 ADHD cases and 35,191 controls showed 12 genome-wide significant loci for ADHD that include genes involved in neurodevelopmental processes and evolutionarily conserved genomic regions (Demontis et al., 2019). In addition, gene expression studies in ADHD revealed novel regulatory networks underlying the disorder. However, most of these studies focused on specific genes, considered relatively small sample sizes and/or were not replicated in independent cohorts (Lorenzo et al., 2018; Nuzziello et al., 2019; Sanchez-Mora et al. 2019).

Recent transcriptome-wide association studies (TWAS) integrating GWAS summary data and expression data identified relevant genes and pathways associated with ADHD, including dopamine and norepinephrine-related pathways, and several genome-wide significant hits from the largest GWAS meta-analysis were found to impact downstream gene expression (Liao et al., 2019; Qi et al., 2019).

The aim of the present study is to perform gene expression profiling in peripheral mononuclear blood cell (PMBCs) from 94 ADHD subjects and 124 controls and to replicate the results in an independent dataset of 49 ADHD subjects and 45 controls.

2. Experimental procedures

2.1. Participants

Gene expression profiles were assessed in a discovery sample of 94 adult ADHD subjects (60.6% male, mean age=34.8 years, s.d = 11.3) and 124 healthy controls (55.6% male, mean age=36.7 years, s.d = 10.0) and a follow-up sample of 49 ADHD subjects (57.1% male, mean age=30.3 years, s.d = 11.2) and 45 healthy controls (55.6% male, mean age=51.7 years, s.d = 21.4). Subjects were of European ancestry and were evaluated and recruited prospectively from a restricted geographic area in a specialized out-patient program for adult ADHD at the Hospital Universitari Vall d'Hebron of Barcelona (Spain). Patients were medication-naïve without other psychiatric comorbidities and were distributed in two groups (discovery and follow-up) depending on the period in which they were recruited and clinically assessed.

2.2. Clinical assessment

Clinical assessment was conducted by structured interviews and self-reported questionnaires in two different steps: (i) assessment of ADHD diagnosis based on symptomatology using the Conner's Adult ADHD Diagnostic Interview for DSM-IV (CAADID) and (ii) assessment of the severity of ADHD symptoms, the levels of impairment and the presence of comorbid disorders to increase the diagnostic accuracy with the Conners' ADHD Rating Scale (CAARS), the ADHD Rating Scale (ADHD-RS), the Clinical Global Impression (CGI), the Wender Utah Rating Scale (WURS), the Sheehan Disability Inventory (SDS), and the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II). Exclusion criteria were IQ $<$ 70; lifelong and current history of mood, psychotic, anxiety, substance abuse, and DSM-IV axis II disorders; pervasive developmental disorders; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight $\leq$ 1.5 kg; and other neurological or systemic disorders that might explain ADHD symptoms.

The control sample consisted of unrelated healthy blood donors matched by sex with the clinical group. Individuals with ADHD symptomatology were excluded retrospectively under the following criteria: (1) diagnosed with ADHD previously and (2) answering positively to the life-time presence of the following ADHD symptoms: (a) often has trouble in keeping attention on tasks, (b) usually loses things needed for tasks, (c) often fidgets with hands or feet or squirms in seat, and (d) often gets up from seat when remaining in seat is expected.

The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, methods were performed in accordance with the relevant guidelines and regulations and written informed consent was obtained from all subjects before inclusion in the study.

2.3. RNA isolation

PBMCs were separated by the Ficoll density gradient method immediately after blood extraction, and total mRNA was isolated from PBMCs using Qiazol Lysis reagent and the RNAeasy Midi Kit (Qiagen). RNA concentration and integrity (RNA Integrity Number, RIN) was assayed using the 2100 Bioanalyzer (Agilent Technologies).

2.4. Microarray assays

RNA was reverse transcribed and amplified using the Ambion WT Expression Kit (Life technologies). cDNA produced was subsequently fragmented and labelled using the GeneChip WT Terminal Labelling and Hybridization Kit (Life technologies). Samples were hybridized in three batches to the Genechip Human Gene 1.1 ST 96-Array (Life technologies). The array processing and data generation were assessed using the Gene Titan Affymetrix microarray platform.

3. Statistical analyses

3.1. Discovery stage

The Robust Multichip Average (RMA) function in the oligo R package was used to background correct, normalize and summarize probe values for each batch independently. Probes that did not match to genes, matched with more than one gene or matched to genes in the X or Y chromosomes in the GRCh37/hg19 human genome build (release 32) were discarded. The study was restricted to 19,184 probes corresponding to 18,227 unique genes. Gene expression was adjusted for batch effect using the CombatR algorithm and linear regression models were used to compare gene expression patterns between ADHD cases and controls with the Limma R-package including gender and RNA Integrity Number (RIN) as covariates. Benjamini-Hochberg correction was applied for multiple comparisons ($P_{BH}<0.05$).

The set of genes showing suggestive evidence of differential expression between ADHD subjects and controls (*P*-value<0.05) were selected for further enrichment analyses of canonical pathways and downstream analyses, including diseases and functional annotations and gene networks using the Ingenuity Pathway Analysis software (IPA; Ingenuity Systems, Redwood City, CA, USA). Pathway analysis was restricted to the 25 most significant networks following IPA default settings. Gene networks were considered of relevance when the Network Score (P-score=-log10(*P*-value)) was over 8 (*P*-value$<$1e-08). IPA was also used to test for overrepresentation of candidate genes previously studied in ADHD selected from the gene list provided by the ADHDgene database (http://adhd.psych.ac.cn/index.do) and a comprehensive search for published reviews of ADHD genetic and pharmacogenetic studies. The full list of genes is available in Pagerols et al. (2018). Benjamini-Hochberg correction was applied for multiple comparisons ($P_{BH}<0.05$). The gene-set enrichment analysis of gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed with the Enrichr software (https://maayanlab.cloud/Enrichr/), with genes weighted according to *P*-values (1-*P*-value).

3.2. Follow-up stage

Gene expression profiling in the follow-up sample was restricted to the subset of genes differentially expressed in the discovery sample and followed the procedure described above. The correlation between differences in expression levels (logFC) of the subset of selected genes between the discovery and follow-up sample was assessed using a Pearson correlation test. Benjamini-Hochberg correction was applied for multiple comparisons ($P_{BH}<0.05$).

4. Results

After standard quality controls, expression data was available for 18,227 autosomal genes in a discovery sample of 94 medication-naïve ADHD patients and 124 healthy controls. Suggestive evidence of association (*P*-value<0.05 prior to correction for multiple testing) was detected for 1793 genes (Supplemental Table 2), which were enriched for genes previously linked with ADHD (48 of 1793 genes; *P*-value=2.1e-03). 25 gene networks were formed with the most significant one being the "*Cell-To-Cell Signaling and Interaction, Nervous System Development and Function, Developmental Disorder*" (Network Score=44; *P*-Value=1e-44; Figure 1 and Supplemental Table 3). Enrichment analyses also revealed significant over-representation for genes in 26 canonical pathways including pathways related to cell growth and differentiation, axon guidance and synaptic plasticity as well as immune and inflammatory response, among others, with the "*Retinoic acid receptor activation*" pathway being the most significant (P_{BH}=2.4e-03, ratio=0.174, 33 overlapping genes; Supplementary Table 4). In addition, three gene ontology biological processes associated with gene expression regulation were significantly enriched in this gene set, with the top process being "*Regulation of transcription, DNA-templated*" (P_{BH}=1.9e-05; Supplementary Table 5).

After multiple-comparison correction ($P_{BH}<0.05$), a total of 21 genes were differentially expressed between cases and controls, one up-regulated and 20 down-regulated in the ADHD group (Table 1). Expression profiles of the 21 genes differentially expressed in the discovery sample were subsequently assessed in an independent sample of 49 ADHD subjects and 45 controls. We found a strong correlation in expression level differences for this set of genes between the discovery and follow-up cohorts (r = 0.88; *P*-value=1.18e-07; Figure 2). Expression differences were replicated ($P_{BH}<0.0024$) for 8 of them (*KMT5A, IL7R, RAB11FIP1, LRRFIP1, KLF4, SLA, EGR2* and *SNORA38)* with the same direction of effect on ADHD across the two independent studies (Table 1).

5. Discussion

This transcriptome profiling study provides evidence of differential gene expression in peripheral blood mononuclear cells of subjects with ADHD. Through a two-step exploratory study we found eight genes showing differential expression levels in ADHD, with consistent direction of effect in two independent datasets. They include genes linked to schizophrenia and educational attainment (*KMT5A*) (Ripke et al., 2014; Lee et al., 2018), neural differentiation and migration (*KLF4* and *IL7R*) (Qin et al., 2011; Moors et al., 2010) or membrane trafficking and axon growth (*RAB11FIP1*) (Eva et al., 2010), and genes previously associated with alcohol dependence (*SLA*) (Wang et al., 2013), insomnia (*SNORA38*) (Jansen et al., 2019) and cognitive function (*EGR2*) (Watanabe et al., 2019). These results support previous studies revealing shared genetic background and common biological pathways underlying different psychiatric disorders and ADHD-comorbid conditions (Lee et al., 2019; Anttila et al., 2018).

Despite identification of multiple independent loci conferring risk to ADHD, the molecular mechanisms underlying the genetic basis of the disorder remain to be elucidated (Demontis et al., 2019). Although integrative approaches combining GWAS and expression data in previous studies provided biological insight into ADHD associated genes (Liao et al., 2019; Qi et al., 2019; Fahira et al. 2019),

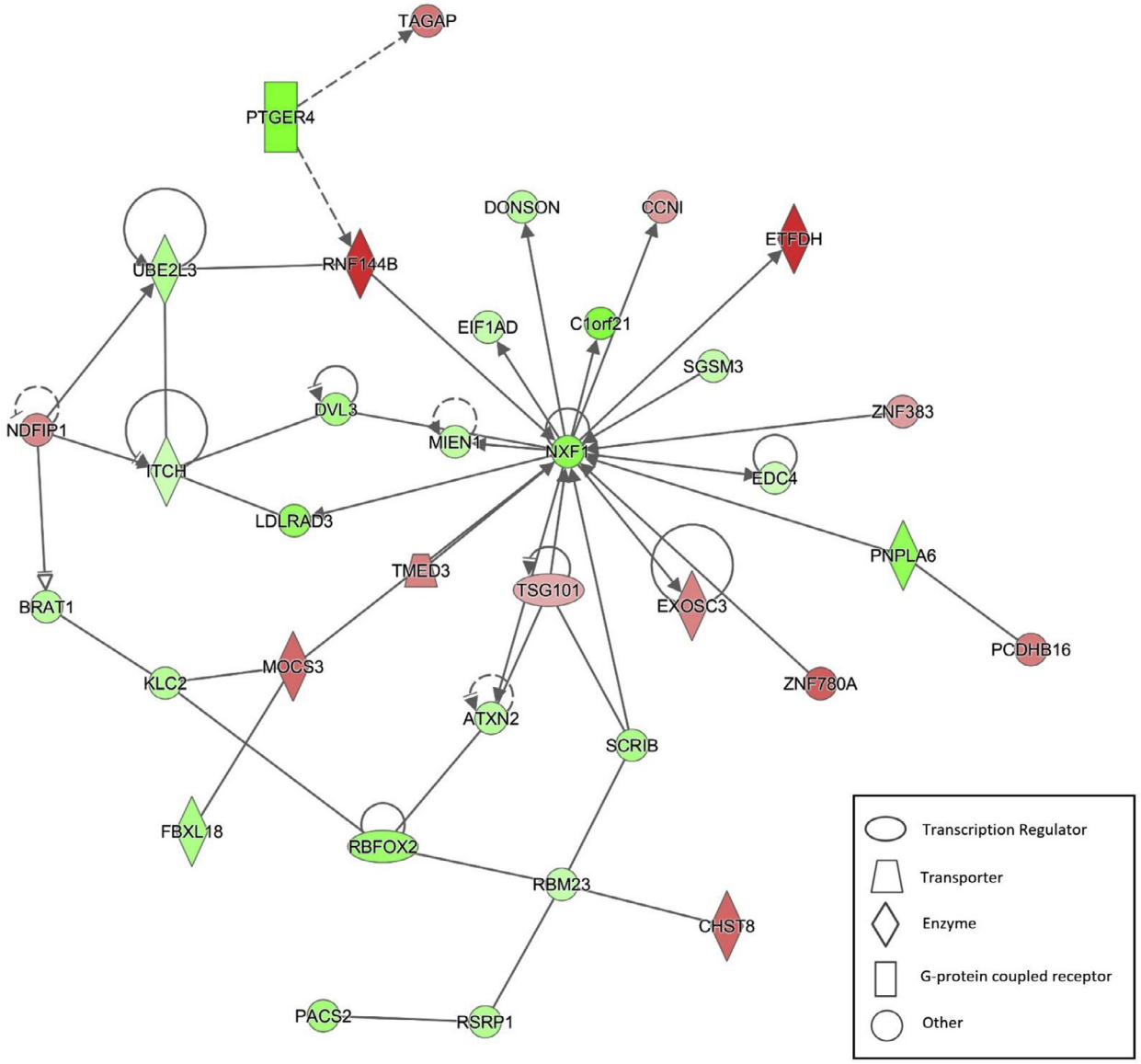


Fig. 1 "Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Developmental Disorder", top Ingenuity Pathway Analysis constructed gene network considering 1793 genes showing suggestive evidence of differential expression in ADHD. Gene interaction is shown with solid lines representing direct interaction between molecules and dashed lines representing indirect interaction. color intensity is proportional to degree of expression with red indicating upregulation and green representing downregulation. Node shape indicates a protein's primary function according to the legend insert.

we found no overlap between genes showing differential expression in the present study and GWAS hits for ADHD (Demontis et al., 2019; Rovira et al., 2020). These results are not surprising given that the majority of specific factors underlying ADHD remain unknown. Non-*cis* eQTL genetic components such as trans- eQTL, low-frequency/rare variants, and environmental factors may also account for a substantial proportion of gene expression (Grundberg et al., 2012). Although no GWAS hits were found among differentially expressed genes, we did found enrichment of ADHD candidate genes from the ADHDgene database and a comprehensive search from published reviews on genetics and pharmacogenetics on ADHD (Pagerols et al., 2018). These results support the potential relevance to ADHD pathogenesis of biological processes related with growth and differentiation, axon guidance, synaptic plasticity or nervous system development and are in agreement with previous results of RNA-seq blood transcriptome profiling described in familial ADHD (Lorenzo et al., 2018).

The results of the study should be interpreted in the context of several limitations. First, given the reduced sample size, the study is underpowered, which may have prevented us from detecting genes with small-effect sizes and emphasizes the need of further replication in larger samples to estimate the magnitude of the effect of the observed associations. Second, cases and controls were not paired-matched for gender and age and we cannot discard that gender or age-related differences in gene-expression

Table 1 List of differentially expressed genes in the discovery sample of 94 ADHD subjects and 124 healthy controls and in the follow-up sample of 49 ADHD subjects and 45 healthy controls after applying multiple testing corrections. Differentially expressed genes replicated in the follow-up sample are shown in bold.

Gene symbol	Gene name	Discovery sample			Follow-up sample		
		logFC	*P*-Value	P_{BH}	logFC	*P*-Value	P_{BH}
KMT5A	***lysine methyltransferase 5A***	**−0.12**	**7.60E-08**	**0.0015**	**−0.20**	**0.0003**	**0.0051**
NXF1	*nuclear RNA export factor 1*	−0.11	4.20E-07	0.0041	−0.06	0.3125	0.3860
KLF4	***Kruppel like factor 4***	**−0.38**	**7.30E-07**	**0.0047**	**−0.50**	**0.0040**	**0.0170**
LRRFIP1	***LRR binding FLII interacting protein 1***	**−0.2**	**3.30E-06**	**0.0160**	**−0.37**	**0.0028**	**0.0150**
PPP1R9B	*protein phosphatase 1 regulatory subunit 9B*	−0.11	5.00E-06	0.0160	−0.03	0.6498	0.6823
RAB11FIP1	***RAB11 family interacting protein 1***	**−0.2**	**6.00E-06**	**0.0170**	**−0.29**	**0.0024**	**0.0150**
ABCG1	*ATP binding cassette subfamily G member 1*	−0.19	1.20E-05	0.0270	−0.05	0.6113	0.6756
ETFDH	*electron transfer flavoprotein dehydrogenase*	0.13	1.60E-05	0.0310	0.13	0.0688	0.1606
C1QA	*complement C1q A chain*	−0.18	2.20E-05	0.0390	−0.13	0.1242	0.2078
ZC3H3	*zinc finger CCCH-type containing 3*	−0.11	2.90E-05	0.0470	0.12	0.1385	0.2078
SLA	***Src like adaptor***	**0.13**	**3.90E-05**	**0.0480**	**0.19**	**0.0086**	**0.0300**
TEPSIN	*TEPSIN. adaptor related protein complex 4 accessory protein*	−0.11	4.50E-05	0.0480	0.07	0.3621	0.4224
TNFSF8	*TNF superfamily member 8*	0.14	4.60E-05	0.0480	0.09	0.2575	0.3380
SMAP2	*small ArfGAP2*	0.11	4.70E-05	0.0480	0.10	0.1382	0.2078
ZBTB7A	*zinc finger and BTB domain containing 7A*	−0.12	5.10E-05	0.0480	−0.10	0.1205	0.2078
TXNIP	*thioredoxin interacting protein*	0.08	5.10E-05	0.0480	0.01	0.8618	0.8618
SNORA38	***small nucleolar RNA. H/ACA box 38***	**−0.2**	**5.30E-05**	**0.0480**	**−0.30**	**0.0150**	**0.0400**
EGR2	***early growth response 2***	**−0.46**	**5.50E-05**	**0.0480**	**−0.53**	**0.0130**	**0.0400**
IL7R	***interleukin 7 receptor***	**0.13**	**5.70E-05**	**0.0480**	**0.25**	**0.0013**	**0.0130**
TAGLN	*transgelin*	−0.27	6.00E-05	0.0480	−0.22	0.1115	0.2078
SIDT2	*SID1 transmembrane family member 2*	−0.21	6.00E-05	0.0480	−0.17	0.1656	0.2319

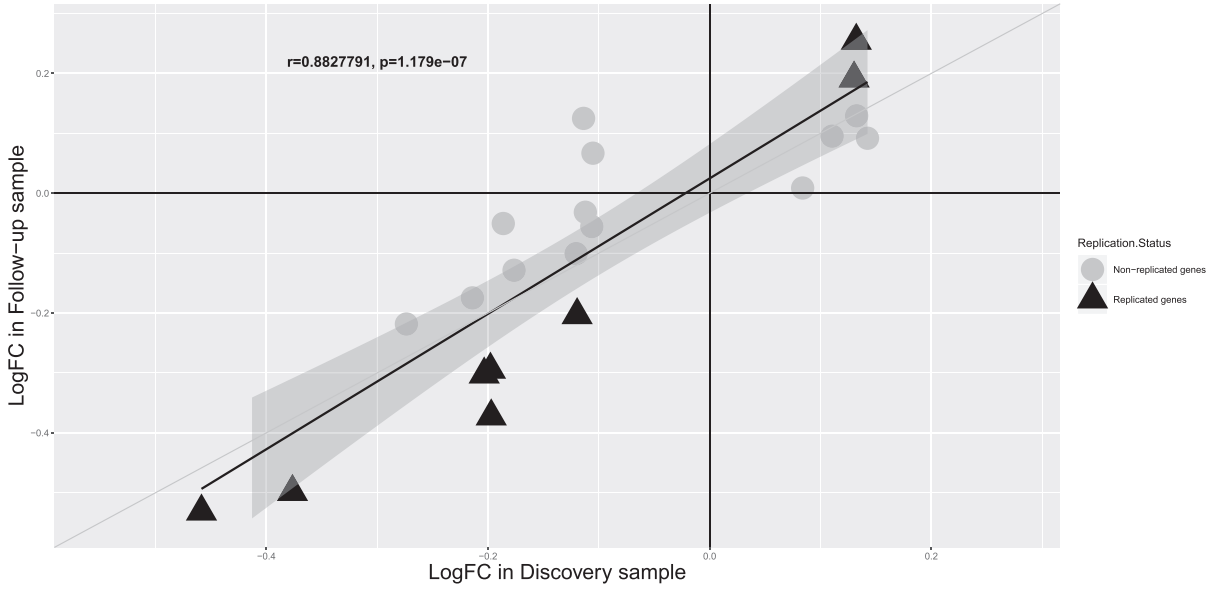


Fig. 2 Correlation in expression level differences (logFC) for the set of 21 genes differentially expressed in the discovery sample between the discovery and follow-up cohorts.

have caused some bias. Despite these methodological limitations, we ran a two-step approach, applied strict selection criteria and restricted the analysis to deeply phenotyped medication-naïve ADHD cases. With this design, we found a high degree of correlation between expression level differences across stages and validated differential expression for eight genes in two independent datasets, which support the robustness of our findings and the use of peripheral tissues to assess gene expression signatures for ADHD. Third, given that our clinical sample included medication-naïve subjects with no comorbid disorders, results are difficult to extend to a more realistic clinical context and further studies in independent clinical samples following less restrictive inclusion criteria are required. Lastly, microarray technology has some limitations when compared to RNA sequencing, including overestimation of target gene expression due to off-target hybridization between probes and homologs of the target gene, lower sensibility and specificity or the inability to detect novel or low abundance transcripts and different isoforms, which highlight RNAseq as an alternative technology to reveal more insight into molecular mechanisms underlying ADHD.

Our results highlight promising candidate genes and gene pathways for ADHD but should be considered as a proof of concept requiring further replication in larger datasets to identify biological subtracts underlying the disorder.

Role of the Funding Source

The funding entities had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions statement

N.M., C.S.M., P.R., and L.V. participated in the RNA isolation and preparation of samples. N.M., C.S.M. and M.S. undertook the statistical analyses. V.R, M.Corrales, C.F, M.Casas and J.A.R.Q. participated in the clinical assessment and in the recruitment of patients. N.M., M.S. and M.R. participated in the study design, clinical assessment and coordination of the clinical research. M.S and M.R. conceived of the project concept, wrote the protocol, and coordinated the study design and the statistical analyses. M.S. and M.R supervised the project and the manuscript preparation. All authors contributed to and have approved the final manuscript.

Conflict of Interest

Josep Antoni Ramos-Quiroga was on the speakers' bureau and/or acted as consultant for Eli-Lilly, Janssen-Cilag, Novartis, Shire, Takeda, Bial, Shionogui, Lundbeck, Almirall, Braingaze, Sincrolab, Medice and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Takeda, Shionogui, Bial, Medice and Eli- Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli-Lilly, Lundbeck, Janssen- Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió. Christian Fadeuilhe has received fees to give talks for Shire, Ferrer, Italfarmaco and Otsuka in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Lundbeck, Otsuka and Ferrer. Vanesa Richarte was on the speakers' bureau and/or acted as consultant for Takeda and Rubió in the last 5 years. She also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Rubió, Shire, Takeda and Lundbeck. Miguel Casas was on the speakers' bureau and/or acted as consultant for Janssen-Cilag in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag. All other authors declare that they have no conflicts of interest.

Acknowledgments

Over the course of this investigation, NM was supported by MiND Marie Sklodowska-Curie which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 643051. P.R. is a recipient of a predoctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI_B 00899). C.S.M. was a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199). L.V. is a recipient of a pre-doctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285). MSA was a recipient of a contract from the Biomedical Network Research Center on Mental Health (CIBERSAM), Madrid, Spain and was also a recipient of a Juan de la Cierva Incorporacion contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I). M.R. is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023). The research leading to these results has received funding from the European Union H2020 Programme (H2020/2014-2020) under grant agreements No. 667302 (CoCA) and No. 728018 (Eat2beNICE), from the Instituto de Salud Carlos III (PI16/01505, PI17/00289, PI18/01788, PI19/00721 and PI19/01224), and cofinanced by the European Regional Development Fund (ERDF), from the Pla estratègic de recerca i innovació en salut (PERIS); Generalitat de Catalunya (MENTAL-Cat; SLT006/17/287) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461). The work was also supported by the ECNP Network 'ADHD across the Lifespan' (https://www.ecnp.eu/research-innovation/ECNP-networks/List-ECNP-Networks/). We thank all the individuals who kindly participate in this study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroneuro.2020.11.005.

References

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th ed.).

Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Patsopoulos, N.A., 2018. Analysis of shared heritability in common disorders of the brain. Science 360 (6395), eaap8757.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Cerrato, F., 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51 (1), 63.

Eva, R., Dassie, E., Caswell, P.T., Dick, G., Norman, J.C., Fawcett, J.W., 2010. Rab11 and its effector Rab coupling protein contribute to the trafficking of β1 integrins during axon growth in adult dorsal root ganglion neurons and PC12 cells. J. Neurosci. 30 (35), 11654–11669.

Faraone, S.V., Larsson, H., 2019. Genetics of attention deficit hyperactivity disorder. Mol. Psychiatry 24 (4), 562–575.

Franke, B., Michelini, G., Asherson, P., Banaschewski, T., Bilbow, A., Buitelaar, J.K., Kuntsi, J., 2018. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. Eur. Neuropsychopharmacol. 28 (10), 1059–1088.

Grundberg, E., Small, K.S., Hedman, Å.K., Nica, A.C., Buil, A., Keildson, S., Nisbett, J., 2012. Mapping cis-and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44 (10), 1084–1089.

Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., Savage, J.E., 2019. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat. Genet. 51 (3), 394–403.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Fontana, M.A., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50 (8), 1112–1121.

Lee, P.H., Anttila, V., Won, H., Feng, Y.C.A., Rosenthal, J., Zhu, Z., Wang, M.M.J., 2019. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179 (7), 1469–1482.

Liao, C., Laporte, A.D., Spiegelman, D., Akçimen, F., Joober, R., Dion, P.A., Rouleau, G.A., 2019. Transcriptome-wide association study of attention déficit associated genes and phenotypes. Nat. Commun. 10 (1), 1–7.

Lorenzo, G., Braun, J., Muñoz, G., Casarejos, M.J., Bazán, E., Jimenez-Escrig, A., 2018. RNA-Seq blood transcriptome profiling in familial attention deficit and hyperactivity disorder (ADHD). Psychiatry Res. 270, 544–546.

Moors, M., Vudattu, N.K., Abel, J., Krämer, U., Rane, L., Ulfig, N., Maeurer, M.J., 2010. Interleukin-7 (IL-7) and IL-7 splice variants affect differentiation of human neural progenitor cells. Genes Immunity 11 (1), 11–20.

Nikolas, M.A., Burt, S.A., 2010. Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: a meta-analysis. J. Abnorm. Psychol. 119 (1), 1.

Nuzziello, N., Craig, F., Simone, M., Consiglio, A., Licciulli, F., Margari, L., Grillo, G., Liuni, S., Liguori, M., 2019. Integrated analysis of microRNA and mRNA expression profiles: an attempt to disentangle the complex interaction network in attention deficit hyperactivity disorder. Brain Sci. 9 (10), 288. doi:10.3390/brainsci9100288.

Pagerols, M., Richarte, V., Sánchez-Mora, C., Rovira, P., Artigas, M.S., Garcia- Martínez, I., Victor, M.M., 2018. Integrative genomic analysis of methylphenidate response in attention-deficit/hyperactivity disorder. Sci. Rep. 8 (1), 1–11.

Polanczyk, G.V., Salum, G.A., Sugaya, L.S., Caye, A., Rohde, L.A., 2015. Annual research review: a meta- analysis of the worldwide prevalence of mental disorders in children and adolescents. J. Child Psychol. Psychiatry 56 (3), 345–365.

Qi, X., Wang, S., Zhang, L., Liu, L., Wen, Y., Ma, M., Liang, X., 2019. An integrative analysis of transcriptome-wide association study and mRNA expression profile identified candidate genes for attention-deficit/hyperactivity disorder. Psychiatry Res. 282, 112639.

Qin, S., Liu, M., Niu, W., Zhang, C.L., 2011. Dysregulation of Kruppel-like factor 4 during brain development leads to hydrocephalus in mice. Proc. Natl. Acad. Sci. 108 (52), 21117–21121.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T., Farh, K.H., Holmans, P.A., Pers, T.H., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511 (7510), 421–427.

Rovira, P., Demontis, D., Sánchez-Mora, C., et al., 2020. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology 45 (10), 1617–1626.

Sánchez-Mora, C., Artigas, M.S., Garcia-Martínez, I., Pagerols, M., Rovira, P., Richarte, V., Franke, B., 2019. Epigenetic signature for attention- deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191- 5p as potential biomarkers in peripheral blood mononuclear cells. Neuropsychopharmacology 44 (5), 890–897.

Wang, J.C., Foroud, T., Hinrichs, A.L., Le, N.X., Bertelsen, S., Budde, J.P., Almasy, L., 2013. A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. Mol. Psychiatry 18 (11), 1218–1224.

Watanabe, K., Stringer, S., Frei, O., Mirkov, M.U., de Leeuw, C., Polderman, T.J., Posthuma, D., 2019. A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51 (9), 1339–1348.

Rovira et al. *Translational Psychiatry* (2020)10:199
https://doi.org/10.1038/s41398-020-0860-4

Translational Psychiatry

ARTICLE Open Access

Epigenome-wide association study of attention-deficit/hyperactivity disorder in adults

Paula Rovira[1,2], Cristina Sánchez-Mora[1,2,3,4], Mireia Pagerols[1,2], Vanesa Richarte[2,3,5], Montserrat Corrales[2,3,5], Christian Fadeuilhe[2,3,5], Laura Vilar-Ribó[1,2], Lorena Arribas[1,2], Gemma Shireby[6], Eilis Hannon[6], Jonathan Mill[6], Miquel Casas[2,3,5], Josep Antoni Ramos-Quiroga[1,2,3,5], María Soler Artigas[1,2,3,4] and Marta Ribasés[1,2,3,4]

Abstract

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable neurodevelopmental disorder that often persists into adulthood. There is growing evidence that epigenetic dysregulation participates in ADHD. Given that only a limited number of epigenome-wide association studies (EWASs) of ADHD have been conducted so far and they have mainly focused on pediatric and population-based samples, we performed an EWAS in a clinical sample of adults with ADHD. We report one CpG site and four regions differentially methylated between patients and controls, which are located in or near genes previously involved in autoimmune diseases, cancer or neuroticism. Our sensitivity analyses indicate that smoking status is not responsible for these results and that polygenic risk burden for ADHD does not greatly impact the signatures identified. Additionally, we show an overlap of our EWAS findings with genetic signatures previously described for ADHD and with epigenetic signatures for smoking behavior and maternal smoking. These findings support a role of DNA methylation in ADHD and emphasize the need for additional efforts in larger samples to clarify the role of epigenetic mechanisms on ADHD across the lifespan.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder characterized by age-inappropriate levels of inattention, impulsivity and hyperactivity[1]. ADHD is a disabling condition in childhood and adolescence which often persists into adulthood, interfering with the quality of social, academic, or occupational functioning[2,3].

ADHD is a multifactorial disorder with an estimated heritability of 76%. Twenty-two percent of its phenotypic variance is explained by common genetic variants[1,4] and the proportion of variance still to be explained might be, to some extent, accounted for by gene by environment interactions. In this context, epigenetic processes have emerged as a plausible mechanism by which environmental exposures can lead to long-lasting alterations, such as variation in brain structure or neuronal circuits, found in psychiatric disorders[5–7]. There is growing evidence that epigenetic dysregulation is a feature of ADHD[6,8–11], depression[12], autism[13–16], schizophrenia[17,18] and bipolar disorder[19].

Studies of DNA methylation profiles in ADHD have been conducted using peripheral blood, cord blood, buccal samples or saliva[6,9–11,20–28]. Candidate gene studies have revealed differential methylation patterns in genes involved in the dopaminergic, serotoninergic and neurotrophic systems, including *SLC6A4*, *DRD4*, *COMT*, *ANKK1*, *BDNF*, or *NGFR*, associated with ADHD symptomatology and severity[23–28]. Seven epigenome-wide association studies (EWASs) on ADHD have been run to date, with sample sizes ranging from 54 subjects for clinical samples[21] to 4,689 individuals in a meta-analysis considering ADHD symptomatology in general population[9], yielding non-overlapping findings across them[6,9–11,20–22]. There is

Correspondence: María Soler Artigas (maria.soler@vhir.org) or Marta Ribasés (marta.ribases@vhir.org)
[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain
[2]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain
Full list of author information is available at the end of the article

limited research on adults using this approach, given that most of the EWASs have focused on pediatric samples[6,10,11,20,22]. To the best of our knowledge, only two studies evaluated methylome-wide patterns on adults[9,21]. One identified methylation changes associated with ADHD symptomatology that did not remain significant when results were meta-analyzed across cohorts[9]. The second one found hypermethylated regions in genes involved in fatty acid metabolism and fatty acid oxidation pathways associated with ADHD persistence when compared to remittance[21]. In the childhood period, Wilmot et al. analyzed a population cohort of school-age boys and found lower methylation levels at the *VIPR2* gene in ADHD subjects compared to their age- and sex- matched controls[10], results that were recently replicated in the largest EWAS on ADHD in children conducted so far[22]. In a similar aged population cohort, Walton et al. investigated ADHD symptom trajectories from birth to adolescence and pointed to epigenetic marks in genes related to neural tube development and peroxisomal mechanisms as candidates to be involved in the different ADHD symptom trajectories across time[6]. In the most recent EWAS evaluating ADHD symptoms in population-based cohorts, aberrant methylation patterns at birth in different regions, lying in the *ERC2* and *CREB5* genes among others, were associated with later ADHD symptoms in childhood or adolescence[11]. And finally, the latest and largest EWAS conducted in a clinical sample of children with ADHD supported the association between ADHD polygenic risk and DNA methylation patterns at the *GART* and *SON* genes[22].

Recent evidence supports a large genetic overlap between ADHD in children and adults[29], but little is known about the co-occurrence between the epigenetic signatures characterizing both groups of age. In addition, although various studies report shared genetics between ADHD and several psychiatric and behavioral traits[4,29], this overlap has not been assessed yet using epigenome-wide data.

Whereas most previous studies considered pediatric clinical samples or adult population-based cohorts with measures of ADHD symptoms, we report an EWAS on a clinical sample of adults with ADHD. With these data we (i) assessed DNA methylation signatures for ADHD in adults through an EWAS in peripheral blood mononuclear cells, (ii) tested whether either polygenic risk burden for ADHD or smoking status had an impact on those DNA methylation signatures, (iii) examined whether exposure to stressful life events had an effect on these methylation patterns in ADHD subjects and (iv) explored the overlap between these findings and results from previous meta-analyses of genome-wide association studies (GWAS-MA) on clinical ADHD or ADHD symptoms in population-based samples, and EWAS on ADHD symptoms or exposure to stressful life events.

Materials and methods

Participants and clinical assessment

The clinical sample consisted of 103 ADHD subjects that were referred to an ADHD program from primary care centers and adult community mental health services. All subjects were evaluated and recruited prospectively from a restricted geographic area of Catalonia (Spain) in a specialized out-patient program for Adult ADHD and by a single clinical group at Hospital Universitari Vall d'Hebron of Barcelona (Spain).

The clinical assessment consisted of structured interviews and self-reported questionnaires in two different steps: (i) assessment of ADHD diagnosis based on symptomatology using the Conner's Adult ADHD Diagnostic Interview for DSM-IV (CAADID) by a psychiatrist and, (ii) assessment of the severity of ADHD symptoms, the levels of impairment and the presence of comorbid disorders by a psychologist to increase the diagnostic accuracy and reduce the likelihood of misdiagnosis with the Conners ADHD Rating Scale (CAARS), the ADHD Rating Scale (ADHD-RS), the Clinical Global Impression (CGI), the Wender Utah Rating Scale (WURS), the Sheehan Disability Inventory (SDS), and the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II). Afterwards, the psychiatrist and psychologist integrate the clinical information and self-reports for the valid assessment of symptoms and impairments. In case of discordance between different raters of ADHD symptoms or inconsistencies between reporters in responses to items measuring similar symptoms, the clinician-identified symptoms on the CAADID prevailed. Exclusion criteria were IQ < 70; lifelong and current history of mood, psychotic, anxiety, substance abuse, and DSM-IV axis II disorders; pervasive developmental disorders; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight≤1.5 kg; and other neurological or systemic disorders that might explain ADHD symptoms. For more detailed information on clinical assessment see Sánchez-Mora et al.[30].

Data pertaining to exposure to 17 stressful life events (six gestational and 11 postnatal) were collected retrospectively with the CAADID Part I[31] and were available from 98 subjects with ADHD. No information was available from controls. Specifically, this questionnaire includes: premature birth, illegal drug abuse during pregnancy, maternal smoking, prenatal exposure to drugs, maternal health problems during pregnancy, other problems during maternal pregnancy, exposure to heavy metals, malnutrition, financial stress and/or poverty, extreme familial stress, neglect, familiar violence, emotional and physical maltreatment, sexual abuse, death or

separation from a loved one, and other trauma in childhood or adolescence.

The control sample consisted of 100 unrelated healthy blood donors matched by sex and ethnicity with the clinical group. Individuals with ADHD symptomatology were excluded retrospectively under the following criteria: (1) having been diagnosed with ADHD previously or (2) answering positively to the lifetime presence of the following ADHD symptoms: (a) often has trouble in keeping attention on tasks, (b) usually loses things needed for tasks, (c) often fidgets with hands or feet or squirms in seat, and (d) often gets up from seat when remaining in seat is expected.

All subjects reported European ancestry, which was confirmed through principal component analysis (PCA) using genetic data. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, all methods were performed in accordance to the relevant guidelines and regulations and written informed consent was obtained from all subjects before inclusion into the study.

DNA isolation, quantification, and genome-wide DNA methylation assays

Peripheral blood mononuclear cells (PBMCs) of patients with ADHD and controls were isolated using the Ficoll density gradient method, and DNA was extracted using the QIAamp DNA Mini Kit DNA Purification following manufacturer's instructions (Qiagen, Hilden, Germany). The quality of the samples was checked by NanoDrop® ND-1000 (Thermo Fisher Scientific, MA) and by PicoGreen® (Thermo Fisher Scientific, MA). Genome-wide DNA methylation was assessed with the Illumina Infinium MethylationEPIC BeadChip Kit (EPIC array) (Illumina, San Diego, CA, USA) following sodium bisulfite treatment of genomic DNA.

DNA methylation analysis based on ADHD diagnosis

Data preprocessing and normalization

The 203 samples included in this study were assayed in three batches, which were preprocessed and normalized separately. Raw signal intensities of each probe were extracted using the Illumina Genome Studio software (https://support.illumina.com) and were imported into the R software (3.6.0 version; https://www.R-project.org) using the minfiData 0.2 package[32]. The bisulfite conversion control probes and the 59 single nucleotide polymorphism (SNP) probes of the EPIC array were used to calculate the bisulfite conversion reaction efficiency and to confirm the absence of sample contamination, respectively. Sex was confirmed for all samples using the *getSex* function of the minfi R package[33]. The Horvath Epigenetic Clock algorithm[34] implemented by the *agep* function of the wateRmelon R package was used to calculate the estimated age of participants according to their DNA methylation data, which correlated with their reported age ($\rho = 0.82$, $SE = 0.04$, $P < 2.00E{-}16$). Poorly performing probes or samples were removed using the wateRmelon R package (version 0.9.9;[35]). The exclusion criteria for the probes included detection *P*-values >0.05 for >1% of the samples and a beadcount <3 for >5% of the samples. Probes that were cross-reactive, present in sexual chromosomes or that contained polymorphisms were also excluded from the study[36,37]. Samples with >1% of probes with a detection *P*-value >0.01 were also removed. Probes that passed the quality control filters were quantile normalized with the *dasen* function of the wateRmelon R package.

Bioinformatic and statistical analyses

PCA of methylation values was conducted using the *prcomp* function of the stats R package, first separately for each batch and then across all batches. Within batch, non-biological experimental variation (Sentrix Position and chip ID) of normalized methylation values was tested for association with the Principal Component loadings (PCs). Chip ID was associated with the first PC (PC1) in all three batches, which accounted for the 99% of the variation of samples. We therefore adjusted the beta values with the *ComBat* function of the SVA R package[38] for this variable. The effect of batch and sex on adjusted methylation values of probes present in the three batches after quality control ($n = 744{,}227$) was tested for association with the PCs estimated in the overall sample. Evidence of clustering according to batch was visually detected and statistically confirmed with a significant association of PC1 with batch (*P*-value < 2.20E−16).

Given that detailed smoking information was not available for each individual, an individual smoking score (continuous measure) was generated based on DNA methylation sites known to be associated with current smoking using a method developed by Elliot and colleagues[39]. To account for methylation differences between cell types, we estimated the cell-type composition using the *estimateCellCounts* function of the FlowSorted.Blood.450k R package[40].

Probe-wise differential methylation analysis was performed using the *lmFit* function of the limma R package[41]. Each CpG site was tested individually in a linear regression model with normalized, corrected beta values as the dependent variable and ADHD status as independent predictor, including covariates for sex, age, batch, smoking score and cell-type composition. Age was included as covariate in all the analysis, since it was significantly different between cases and controls. Multiple testing corrections were applied using false discovery rate (FDR) with a cut-off of 5%[42]. The qqman R package was used to generate the Manhattan plot.

The post-hoc power analysis in our sample calculated with the EPIC array online tool (https://epigenetics.essex.ac.uk/shiny/EPICDNAmPowerCalcs/)[43] using the default significance threshold (P-value < 9.42E−08) showed that 6.12% of sites had > 90% power to detect a mean methylation difference of 1%.

At the differentially methylated CpG site, we tested the association between DNA methylation and the exposure to at least one stressful life event, and to each stressful life event separately using the *lmFit* function of the limma R package. As 17 stressful life events were tested, Bonferroni correction was set at $P < 2.94E{-}03$. We also tested the correlation between the number of stressful life events (sum of overall stressful life events and also separated in pre- and post-natal periods) and DNA methylation levels using Spearman's correlation.

To identify differentially methylated regions (DMRs), we used the Python module *comb-p*[44] to group spatially correlated CpG sites with a seed of P-value < 0.01 and 500 base pairs (bp) as the maximum distance. DMR P-values were corrected for multiple testing using the Šidák correction[45] and significant regions were defined as those with at least two probes and an adjusted P-value < 0.05. DMRs were mapped to genes using the interface provided by the minfi R package or the UCSC Genome Browser to identify the closest gene when no genes were mapped to a region (https://genome.ucsc.edu/cgi-bin/hgGateway).

Sensitivity analyses were conducted with the same parameters described above for the probe-wise and regional analyses excluding smoking score as covariate in the model.

DNA methylation analysis based on ADHD diagnosis controlling for ADHD polygenic burden

Bioinformatic and statistical analyses

ADHD polygenic burden was inferred using a Polygenic Risk Score (PRS) built in a subset of 195 individuals with genotype data available, from three different genotyping waves (Illumina HumanOmni1-Quad BeadChip ($n = 3$), Illumina HumanOmni2.5-8 BeadChip ($n = 29$) and Infinium™ Global Screening Array-24 v2.0 ($n = 163$) (Illumina, San Diego, CA, USA), using summary statistics of the largest GWAS-MA performed to date on ADHD[4], with different P-value thresholds ((P_T) < 1e−04, 5e−04, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1). None of the samples used in this study were included in this GWAS-MA[4], and thus did not contribute to defining the variants included in the PRS.

In this subset of 195 individuals, sensitivity analyses for the differentially methylated sites and regions were conducted with the same parameters used in the original EWAS but including the PRS explaining the most variance (Nagelkerke's R^2) as an additional covariate to control for ADHD polygenic risk burden.

ADHD PRSs for each individual were generated with PRSice2 (https://choishingwan.github.io/PRSice/) including sex and the first five PCs as covariates in the model. To set an empirical threshold for the best-fit PRS, 1,000 permutations were run. Information about the pre-imputation quality control at individual and SNP level for the 195 individuals in the target sample and about the phasing and imputation software used is described elsewhere[29]. The European ancestry panel of the 1000 Genomes Project was considered as reference for the imputation (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/) and best guess genotypes were filtered by excluding variants with MAF < 0.05, missing rate>0.01, Hardy-Weinberg Equilibrium ($P < 1.00E{-}06$). Ambiguous strand and multiallelic variants were removed and independent SNPs (obtained using the clumping parameters $p1 = 1$, $p2 = 1$, $r2 = 0.2$, kb=250 in PLINK1.9[46]) present in all individuals were included ($n = 37{,}527$).

Enrichment analyses

We assessed whether probes in different categories: (i) showing a statistically significant proportion of methylation variance explained by additive genetic effects as reported by Zeng et al.[47]; (ii) probes identified in previous EWASs on exposure to adverse live events[48–50]; (iii) probes identified in previous EWASs on ADHD[21,22] or ADHD symptoms[6,9] or (iv) probes located in ADHD-associated loci identified through GWAS[4,29,51] showed, on average, a stronger association with adult ADHD than other methylation sites by regressing our EWAS test statistics (Z_{score}) on each CpG category as described by van Dongen et al. [9]:

$$|Z_{\text{score}}| = \text{Intercept} + \beta_{\text{category x}} * \text{category x},$$

where $|Z_{\text{score}}|$ represents the absolute value of the Z_{score} from our EWAS on adult ADHD, category x represents whether a CpG belongs or not to a specific category and $\beta_{\text{category x}}$ represents the effect estimate for that category. A CpG was assigned to a category if it was associated to the phenotype of interest according to the P-value thresholds shown in Supplementary Table 1 [excel file]. For GWAS, we considered CpG sites within windows of 10 kb, 100 kb, and 1 Mb around significant variants (Supplementary Table 1 [excel file]). For each enrichment test, bootstrap standard errors were computed with 2,000 bootstraps using the "simpleboot" R package. Bonferroni correction was applied for multiple comparison correction ($P_{\text{Bootstrap}} < 3.85E{-}03$; accounting for the 13 analyses conducted).

We also tested for enrichment of regulatory domains, ontological categories and pathways, using CpG sites with P-value<1.00E−05 in our results. For the enrichment

analysis of regulatory domains, ontological categories and pathways, probes were annotated with the Illumina Human EPIC array annotation R package ("IlluminaHumanMethylationEPICanno.ilm10b2.hg19"). The enrichment analyses for transcription factor binding sites (TFBS) and DNase I hypersensitive sites (DHS) from the ENCODE project[52] were performed using a two-sided Fisher's 2×2 exact test. The enrichment analyses for GO terms and KEGG, Reactome or Biocarta pathways were assessed using the *gsameth* function of the missMethyl R package[53]. Gene sets denoting canonical pathways were downloaded from MSigDB (http://www.broadinstitute.org/gsea/msigdb), which integrates Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/), BioCarta (http://www.biocarta.com/), Reactome (https://reactome.org/) and Gene Ontology (GO) (http://www.geneontology.org/) resources.

The datasets for this article are not publicly available because of limitations in ethical approvals and the summary data will be available upon request.

Results

Our sample consisted of 103 cases and 100 controls after quality control. The distribution of sexes was not significantly different between groups ($\chi^2 = 2.60$, $P = 0.11$), with 56% and 45% of cases and controls being male, respectively. Age of participants was significantly different between cases and controls ($P = 3.61E{-}04$), with a mean age of 31.90 (SD = 11.45) years in cases and of 37.25 (SD = 9.47) years in controls. In the case group, 35% of participants experienced no stressful life events, 35% were exposed to at least one prenatal stressful life event and 54% were exposed to at least one of them after birth (Supplementary Table 2; Supplementary Fig. 1).

We identified one differentially methylated CpG site, cg07143296, in the EWAS (P.adj = 0.033; Fig. 1a, b; Table 1; EWAS inflation factor $\lambda = 0.67$). This CpG lies 77 bp upstream the *PCNXL3* gene and was hypermethylated in patients, with a mean difference of 0.2% between groups (Table 1, Fig. 2). When evaluating the effect of prenatal and postnatal stressful life events on the methylation patterns of ADHD subjects at this CpG site, we found no significant differences in the methylation levels between individuals with ADHD exposed to stressful life events compared to those not exposed. The combined analysis of multiple correlated CpG sites showed evidence of association between ADHD and methylation levels in four genomic regions (P.adj < 0.02), with the most significant one spanning six CpG sites and located in the *DENND2D* gene (P.adj = 2.52E−07; Table 2). The smoking score was not significantly different between cases and controls (mean score in cases = −2.42, mean in controls = −3.34, $P = 0.05$). When we excluded it from the fitted model as a sensitivity analysis, cg07143296 (logFC = 0.0059, $P = 1.19E{-}07$, P.adj=0.07) and the region in chromosome 11 were no longer significant and the other regions remained significant (Table 2).

We subsequently tested whether the polygenic risk burden for ADHD had an effect on the DNA methylation signatures. After constructing PRSs at different P-value thresholds from the largest GWAS-MA on ADHD in children and adults[4], the PRS explaining the most variance in our sample was found for $P_T = 0.001$ ($N_{SNPs} = 490$, $R^2 = 0.052$, $P_{perm} = 0.029$), and was significantly higher in ADHD patients than controls ($P = 3.10E{-}03$; Supplementary Fig. 2). After adding it as a covariate to the model fitted for the EWAS, we found that the cg07143296 CpG site (logFC = 0.066, $P = 1.60E{-}08$, P.adj=0.012) and three of the four genomic regions identified remained significant (Table 2).

We then tested whether CpG sites whose methylomic variation is mainly explained by additive genetic effects showed, on average, a stronger association with adult ADHD than other methylation sites included in the array, and found a significant enrichment of signal for adult ADHD among them ($P_{Bootstrap} = 2.39E{-}04$). In addition, when we assessed the overlap between genetic and epigenetic signatures of ADHD, we found suggestive evidence of overlap between our EWAS results and probes annotated to ADHD-associated loci in the largest GWAS meta-analyses on ADHD across the lifespan or GWAS-MA on ADHD symptoms in children ($P_{Bootstrap} = 6.75E{-}03$ and $P_{Bootstrap} = 1.36E{-}02$, respectively), but not with results of previous GWAS-MA on ADHD conducted separately in adults or children (Supplementary Table 1 [excel file]). We also considered CpG sites differentially methylated in previous EWAS on individuals exposed to adverse life events, on clinical ADHD or on ADHD symptoms and found that CpG sites previously associated with current vs never smoking and with maternal smoking showed a highly significant enrichment of signal for adult ADHD ($P_{Bootstrap} = 9.03E{-}18$ and $P_{Bootstrap} = 4.62E{-}14$, respectively) (Supplementary Table 1 [excel file]). However, no overlap was detected with findings of previous EWASs on ADHD, on ADHD symptoms and on physical/emotional neglect or abuse (Supplementary Table 1 [excel file]).

When we focused on the top 15 differentially methylated CpG sites ($P < 1.00E{-}05$) in our EWAS, we found no enrichment of regulatory domains (TFBS and DHS) from the ENCODE project[52] nor ontological categories or pathways from GO terms, KEGG, Reactome or Biocarta (Supplementary Table 3 [excel file]).

Discussion

To the best of our knowledge, this is the first study evaluating DNA methylation signatures in a clinical sample of adults with ADHD and testing whether

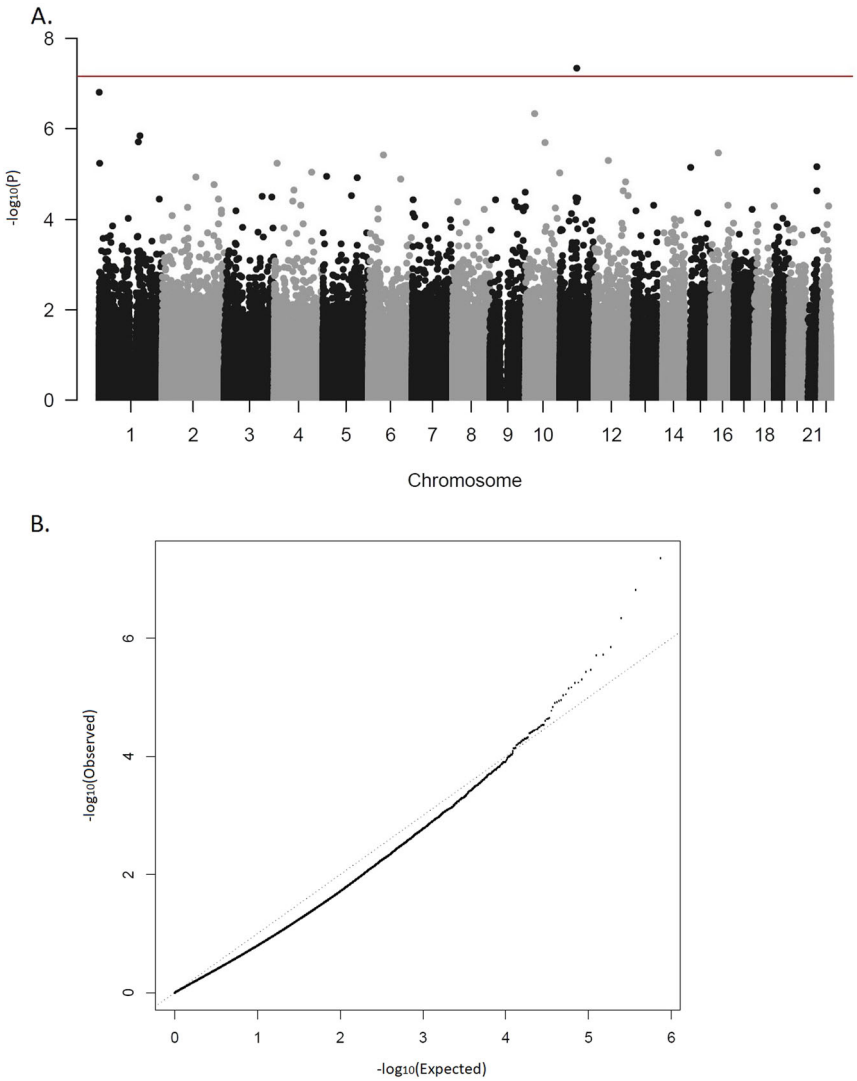


Fig. 1 Results of the epigenome-wide association study. a Manhattan plot. Horizontal line indicates 5% FDR significance threshold (*P*-value=6.72E−08). **b** Quantile-quantile plot.

smoking status, polygenic risk burden for ADHD or exposure to stressful life events had an impact on the methylation signatures identified.

Methylation differences were found in regions that include genes related to cancer and pulmonary function (*DENND2D*)[54,55], neuroticism and regulation of histone acetylation dynamics (*PWWP2B*)[56,57] or regulation of immune signaling (*UBASH3A*)[58]. We also identified a CpG site (cg07143296) significantly hypermethylated in ADHD, located close to *PCNXL3*, a gene related to autoimmune diseases[59]. Although not achieving significance after multiple comparison correction, CpG sites in ADHD-related genes were found among the top ten signals of the EWAS, including *CREM*, which has been previously associated with impulsivity, hyperactivity, anxiety-like behavior, circadian rhythmicity and drug addiction[60–62], *ADK*, whose deficiency may result in altered dopaminergic function, attentional impairment, and learning impairments[63,64], or *LAT*, whose genetic variation has been associated with educational attainment[65].

The lack of overlap between our EWAS results and those from previous EWASs on ADHD in childhood[6,10,11,20,22] is in line with the fact that genome-wide DNA methylation is highly age dependent[34]. Contrary to some risk factors stably involved in ADHD throughout the lifespan, DNA methylation is developmental-stage specific and hence the patterns contributing to ADHD susceptibility may differ over time. The absence of overlap between our results and findings from previous EWASs

Table 1 Top 10 ADHD-associated differentially methylated CpG sites.

Probe ID	Location	%	logFC	P	P.adj	Gene	DHS	DHS location	Overlap transcription factor binding site
cg07143296	11: 65383707	0.24	0.0062	4.41E−08	0.033	*PCNXL3* (+77 bp)	Yes	TSS200	POLR2A | SIN3A | RBBP5 | EGR1 | YY1 | MAX | ZBTB7A | MYC
cg05041517	1: 1021029	0.24	0.0105	1.53E−07	0.057	*C1orf159*	No	–	–
cg01174734	10: 35485053	0.83	0.0152	4.54E−07	0.113	*CREM*	Yes	1st Exon; Body; 5′UTR	POLR2A | TCF7L2 | TCF12 | STAT5A | TBP | PML | NFATC1 | FOXM1 | BCL3 | EP300 | MXI1 | STAT3 | KAP1 | ATF3 | REST | GTF2F1 | JUND | MTA3 | NFIC | CEBPB | GABPA | MAX | RFX5 | TRIM28 | ATF2 | RELA | E2F1 | TAF1 | CBX3 | USF1 | RUNX3 | BCLAF1 | CREB1 | MAZ | CTCF | E2F4 | SMC3 | RAD21 | ZNF143
cg25889770	1: 161068063	0.48	0.0073	1.40E−06	0.242	*KLHDC9* (+88 bp)	Yes	TSS200	NFIC | PML | FOXM1 | SETDB1 | TFAP2C | POLR2A | RCOR1 | TBP | RFX5 | SIN3A | MXI1 | MAX | SREBP1 | HSF1 | TAF1 | SIN3AK20 | RELA | RUNX3 | YY1 | CTCF | PAX5 | POU2F2 | ZNF143 | ELF1 | BCL3 | ATF3 | SMC3
cg15705054	1: 154927709	0.36	0.0063	1.89E−06	0.242	*PBXIP1*	Yes	5′UTR	POLR2A | GATA1 | CBX3 | MAZ | MAX | SIN3AK20 | GATA2 | EP300 | SIN3A | RCOR1 | EGR1 | GABPA | TAL1 | MXI1 | RUNX3 | TCF3 | TBP | SPI1 | RFX5 | TCF12 | ELF1 | TBL1XR1 | FOXA1 | PAX5
cg05770546	10: 76004016	2.17	0.0295	1.96E−06	0.242	*ADK*	Yes	Body	CTCF | ZNF143 | RAD21 | SMC3
cg05529890	16: 28997459	0.93	0.0124	3.40E−06	0.349	*LAT*	Yes	Exon boundary; Body	–
cg23874234	6: 57087441	0.81	0.0129	3.75E−06	0.349	*RAB23* (−0.34 kb)	Yes	TSS1500	POLR2A | PHF8 | TAF1 | TCF7L2 | E2F1 | MBD4 | TBP
cg16449840	12: 54718606	0.28	0.0058	5.00E−06	0.385	*COPZ1 (+0.27 kb)*	Yes	TSS1500	HDAC8 | MYC | UBTF | E2F1 | ELK1 | MBD4 | RCOR1 | FOXM1 | RXRA | CBX3 | GTF2F1 | FOSL2 | TBP | NR2F2 | JUND | BHLHE40 | IRF1 | CTCF | STAT3 | EP300 | MAFF | ZNF143 | MAX | MAZ | STAT1 | REST | GATA3 | ARID3A | SMC3 | RAD21 | RUNX3 | MAFK | MXI1 | GATA2 | USF2 | ATF3 | SIN3AK20 | USF1 | CTCFL | NFYB | GATA1 | RFX5 | E2F4 | ELK4 | ATF1
cg24479820	1: 2388073	0.91	0.0147	5.62E−06	0.385	*PLCH2* (10.83 kb)	No	–	–

Location (chromosome: base pair), DNA methylation change between groups (%), log fold change estimates (logFC), *P*-values (*P*) and adjusted *P*-values (P.adj) are shown for each CpG site.

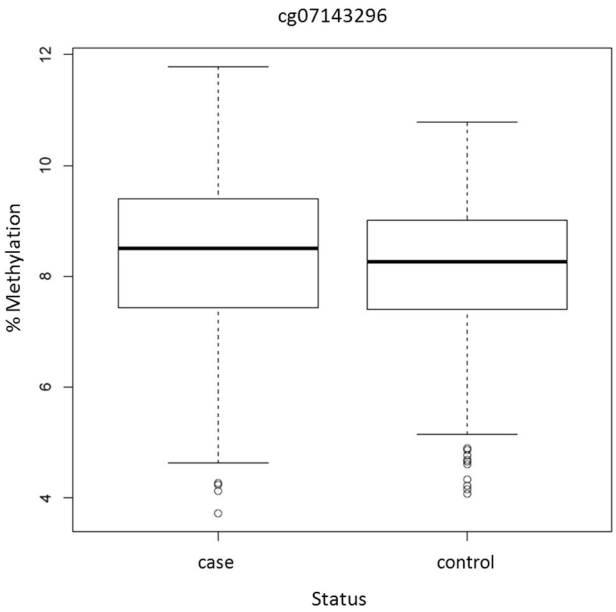


Fig. 2 CpG-specific DNA methylation levels. Boxplot showing the levels of DNA methylation in cases and controls at cg07143296.

on ADHD in the adulthood period[9,21] could be ascribed to differences in the characteristics of the samples and on the array used (clinical vs population-based samples and EPIC vs Infinium Human Methylation 450K array[9]), to random variation and limited statistical power or, as previously suggested by Meijer et al.[21], to the fact that the epigenetic effects identified may not be those with the strongest effect sizes on the phenotype[21].

Results on the relationship between genetic and epigenetic signatures in ADHD were not conclusive. We found enrichment of signal for adult ADHD in CpGs whose methylation variance is mainly explained by additive genetic effects[47] and suggestive evidence of enrichment in loci described in the largest GWAS-MA on ADHD[4] and on ADHD symptoms[51]. However, no evidence was found for overlap between our EWAS results and loci from smaller GWAS-MAs on ADHD[28] or for a substantial effect of the polygenic burden for ADHD on the methylation patterns identified. These inconsistent results should be interpreted in the context of the limited statistical power of the EWAS and warrant further investigation.

Our EWAS findings do not seem to be driven by an effect of current smoking since they were significant when we adjusted the model for it. When excluding smoking status from the model, we did not detect an effect of methylation on ADHD through smoking for cg07143296 or for the region in chromosome 11 but we cannot rule out a mediating effect for the remaining regions as their signal becomes more significant. Although bearing in mind that we used an estimated smoking score that might be a less accurate tool than clinical data, it has been postulated as a valid marker for current tobacco exposure[13,39].

We also report preliminary data supporting overlap between epigenetic signatures of ADHD and smoking-related traits or behaviors. Enrichment of top-ranking CpGs from previous EWASs on smoking behavior[49] or maternal smoking[50] was obtained. In addition, methylation differences were identified in regions lying in or near genes (such as *DENND2D* or *PWWP2B*) related to phenotypes where tobacco exposure is a key risk factor[66–68], and maternal smoking, which increases risk of ADHD in the offspring[69–71], was the most frequently prenatal stressful life event reported by participants with ADHD.

To note, sixty-five percent of individuals with ADHD reported having been exposed to stressful life events, a circumstance that has been associated with the persistence of the disorder into adulthood[72]. Extreme familial stress was found among the most frequently reported postnatal exposures in individuals with ADHD, which is not surprising given that the presence of ADHD has been associated to varying degrees of disturbances in family and marital functioning[73–75]. However, no effect of stressful live events on DNA methylation patterns was found in ADHD subjects. Given that our study lacked data on exposure to stressful live events in controls, larger studies including cases and controls are needed to understand the impact of environmental factors on DNA methylation patterns associated with ADHD.

The results of the present study should be interpreted in the context of several limitations. First, the limited sample size of the present EWAS, which should be viewed as a pilot study whose findings await further replication. Second, our study design allowed the assessment of methylation patterns in a restricted clinical sample of medication-naïve subjects with no comorbid disorders. This design may have facilitated the identification of novel epigenetic signatures, which may not have been possible using a broader recruitment strategy. However, given that patients under medication and/or with lifetime comorbidities were excluded and this group accounts for a not negligible proportion of the overall ADHD group, further studies in larger samples including cases and controls meeting common inclusion criteria, more relaxed in terms of medication or comorbid disorders, will be required to clarify whether the results obtained could be generalized to a more realistic clinical situation. Third, the low inflation factor obtained indicates that the distribution of effect sizes in the present EWAS were not driven by systematic biases but also suggests that our study had limited statistical power and that the data may have been overcorrected, which may have prevented us from detecting methylation signatures with small effect sizes.

Table 2 ADHD-associated differentially methylated regions identified using *comb-p*.

Chr	Coordinates	Gene annotation	CpGs	Covariates in the model	Min P	P	P.adj
1	111743038- 111743411	*DENND2D*	cg18924738, cg19269039, cg20317872, cg23184711, cg00619207, cg19268695	Sex, age, batch, smoking score and cell-type composition	8.48E−05	1.26E−10	2.52E−07
				Sex, age, batch, smoking score, cell-type composition and PRS	5.19E−06	7.25E−12	1.45E−08
	111743200- 111743411		cg18924738, cg19269039, cg20317872, cg23184711, cg00619207, cg19268695	Sex, age, batch and cell-type composition	7.73E−05	3.49E−10	1.23E−06
10	134231487- 134231549	*PWWP2B (-0.12 kb)*	cg09287328, cg03860038, cg23249922	Sex, age, batch, smoking score and cell-type composition	7.61E−03	1.50E−07	1.80E−03
				Sex, age, batch, smoking score, cell-type composition and PRS	4.61E−03	8.49E−08	1.02E−03
				Sex, age, batch and cell-type composition	9.41E−04	1.52E−08	1.82E−04
11	69241075- 69241093	*LOC102724265 / AK094674*	cg22203628, cg13647725	Sex, age, batch, smoking score and cell-type composition	7.61E−03	1.18E−07	4.86E−03
				Sex, age, batch, smoking score, cell-type composition and PRS	2.28E−03	1.84E−08	7.60E−04
				Sex, age, batch and cell-type composition	1.92E−03	6.18E−08	2.55E−03
21	43823797- 43823863	*UBASH3A (+ 0.11 kb)*	cg20272209, cg27280688, cg10690747, cg13578652	Sex, age, batch, smoking score and cell-type composition	1.53E−03	1.84E−06	0.02
				Sex, age, batch, smoking score, cell-type composition and PRS	4.61E−03	7.84E−06	0.09
	43823749- 43823863		cg20272209, cg27280688, cg10690747, cg13578652	Sex, age, batch and cell-type composition	2.52E−04	2.03E−09	1.33E−05

For each region, its coordinates (chromosome and start-end bp), CpGs included, minimum *P*-value in the region (Min *P*), non-corrected DMR P-value (*P*) and corrected *P*-value (P.adj) are shown along with the gene annotation obtained from the Illumina Human EPIC array annotation file and from UCSC when a gene was not annotated in a region. Results including polygenic risk score (PRS) for ADHD or excluding smoking score as covariate are also shown.

And fourth, peripheral tissues used generally as proxies have limited utility for inferring variation in the brain[76], although these novel signatures identified in blood might be used as biomarkers for the disorder.

In summary, we conducted the largest study assessing DNA methylation signatures in a clinical sample of adult patients with ADHD. Our results suggest that ADHD polygenic risk burden or current smoking status do not change substantially the methylomic variation between cases and controls, suggest an overlap between epigenetic signatures of ADHD and smoking-related traits, and point to an overlap between genetic and epigenetic signatures in ADHD. These results emphasize the need of additional efforts in larger samples and the inclusion of stressful life events in future studies to clarify the role of epigenetic mechanisms and environmental risk factors on ADHD across the lifespan.

Acknowledgements
Over the course of this investigation, P.R. is a recipient of a predoctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI_B 00899). C.S.M. is a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199). M.P has been a recipient of a predoctoral fellowship from the Vall d'Hebron Research Institute (PRED-VHIR-2013). L.V. is a recipient of a predoctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285). M.S.A. was a recipient of a contract from the Biomedical Network Research Center on Mental Health (CIBERSAM), Madrid, Spain, and now is a recipient of a Juan de la Cierva Incorporación contract from the Ministry of Science, Innovation and Universities, Spain (IJC2018-035346-I). M.R. is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023). The research leading to these results has received funding from the European Union H2020 Programme (H2020/2014-2020) under grant agreements no. 667302 (CoCA) and no. 728018 (Eat2beNICE), from the Instituto de Salud Carlos III (PI15/01789, PI16/01505, PI17/00289, PI18/01788, PI19/00721, and P19/01224), from the Pla estratègic de recerca i innovació en salut (PERIS); Generalitat de Catalunya (MENTAL-Cat; SLT006/17/287) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461) and cofinanced by the European Regional Development Fund (ERDF) and by "la Marató de TV3" (092330/31),. The work was also supported by the ECNP Network 'ADHD across the Lifespan' (https://www.ecnp.eu/research-innovation/ECNP-networks/List-ECNP-Networks/). This paper reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained therein. The authors are grateful to patients and controls who kindly participated in this research. The Microarray-based DNA methylation profiling was performed by the Spanish "Centro Nacional de Genotipado" (CEGEN-ISCIII) (www.cegen.org). We thank Dr. Walton for providing the list of 783 probes with $P < 1.00E{-}03$ in their EWAS on childhood ADHD symptoms. We thank Dr. Meijer for providing the lists of 723 and 658 probes with $P < 1.00E{-}03$ in their EWAS on persistent ADHD versus healthy controls and in their EWAS on persistent versus remittent ADHD, respectively. We thank Dr. Mooney for providing the list of 679 probes with $P < 1.00E{-}03$ in their EWAS on childhood ADHD. We thank Dr. Zeng and Dr. Haley for providing the list of probes of the Illumina EPIC array with a significant proportion of methylation variance explained by additive genetic effects.

Author details
[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [2]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [3]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. [4]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, University of Barcelona, Catalonia, Spain. [5]Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [6]University of Exeter Medical School, University of Exeter, Exeter, UK

Conflict of interest
Paula Rovira, Dr. Sánchez-Mora, Dr. Pagerols, Laura Vilar-Ribó, Lorena Arribas, Gemma Shireby, Dr. Hannon, Prof. Mill, Dr. Soler Artigas and Dr. Ribasés reported no biomedical financial interests or potential conflicts of interest. Vanesa Richarte has served on the speakers for Eli Lilly, Rubio and Shire in the last 5 years. She has received travel awards from Eli Lilly and Co. and Shire for participating in psychiatric meetings. The ADHD Program has received unrestricted educational and research support from Eli Lilly and Co., Janssen-Cilag, Shire, Rovi, Psious and Laboratorios Rubió in the past 2 years. Montserrat Corrales received travel awards for taking part in psychiatric meetings from Shire. Christian Fadeuilhe received fees to give talks for Shire, Ferrer, Italfarmaco and Otsuka in the last 5 years. He also received travel awards (air tickets +hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Lundbeck, Otsuka and Ferrer. Prof. Casas has received travel grants and research support from Eli Lilly and Co., Janssen-Cilag, Shire and Lundbeck and served as consultant for Eli Lilly and Co., Janssen-Cilag, Shire and Lundbeck. Prof. Ramos-Quiroga was on the speakers' bureau and/or acted as consultant for Eli-Lilly, Janssen-Cilag, Novartis, Shire, Lundbeck, Almirall, Braingaze, Sincrolab, Medice, and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Medice and Eli- Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli-Lilly, Lundbeck, Janssen- Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-020-0860-4).

Received: 13 November 2019 Revised: 5 May 2020 Accepted: 15 May 2020
Published online: 19 June 2020

References

1. Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Prim.* **1**, 15020 (2015).
2. Franke, B. et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur. Neuropsychopharmacol.: J. Eur. Coll. Neuropsychopharmacol.* **28**, 1059–1088 (2018).
3. Kupper, T. et al. The negative impact of attention-deficit/hyperactivity disorder on occupational health in adults and adolescents. *Int. Arch. Occup. Environ. Health* **85**, 837–847 (2012).
4. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
5. Rutten, B. P. & Mill, J. Epigenetic mediation of environmental influences in major psychotic disorders. *Schizophrenia Bull.* **35**, 1045–1056 (2009).
6. Walton, E. et al. Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Mol. Psychiatry* **22**, 250–256 (2017).
7. Mill, J. & Petronis, A. Pre- and peri-natal environmental risks for attention-deficit hyperactivity disorder (ADHD): the potential role of epigenetic processes in mediating susceptibility. *J. Child Psychol. Psychiatry Allied Discip.* **49**, 1020–1030 (2008).
8. Hamza, M. et al. Epigenetics and ADHD: toward an integrative approach of the disorder pathogenesis. *J. Atten. Disord.* **23**, 655–664 (2019).
9. van Dongen, J. et al. Epigenome-wide association study of ADHD symptoms in adults. *Biol. Psychiatry* https://doi.org/10.1016/j.biopsych.2019.02.016 (2019).
10. Wilmot, B. et al. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. *J. Child Psychol. Psychiatry Allied Discip.* **57**, 152–160 (2016).
11. Neumann, A. et al. Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. Preprint at https://www.biorxiv.org/content/10.1101/ (2019).

12. Li, M. et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* **47**(D1), D983–D988 (2019).
13. Hannon, E. et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med.* **10**, 19 (2018).
14. Loke, Y. J., Hannan, A. J. & Craig, J. M. The role of epigenetic change in autism spectrum disorders. *Front. Neurol.* **6**, 107 (2015).
15. Sun, W. et al. Histone acetylome-wide association study of autism spectrum disorder. *Cell* **167**, 1385–1397 e1311 (2016).
16. Tremblay, M. W. & Jiang, Y. H. DNA methylation and susceptibility to autism spectrum disorder. *Annu. Rev. Med.* **70**, 151–166 (2019).
17. Pries, L. K., Guloksuz, S. & Kenis, G. DNA methylation in schizophrenia. *Adv. Exp. Med. Biol.* **978**, 211–236 (2017).
18. Viana, J. et al. Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. *Hum. Mol. Genet.* **26**, 210–225 (2017).
19. Ludwig, B. & Dwivedi, Y. Dissecting bipolar disorder complexity through epigenomic approach. *Mol. psychiatry* **21**, 1490–1498 (2016).
20. Gervin, K., Nordeng, H., Ystrom, E., Reichborn-Kjennerud, T. & Lyle, R. Long-term prenatal exposure to paracetamol is associated with DNA methylation differences in children diagnosed with ADHD. *Clin. Epigenetics* **9**, 77 (2017).
21. Meijer et al. Genome-wide DNA methylation patterns in persistent attention-deficit/hyperactivity disorder and in association with impulsive and callous traits. *Front. Genet* **11**, 16 (2020).
22. Mooney, M. A. et al. Large epigenome-wide association study of childhood ADHD identifies peripheral DNA methylation associated with disease and polygenic risk burden. *Transl. Psychiatry* **10**, 8 (2020).
23. Heinrich, H. et al. Attention, cognitive control and motivation in ADHD: linking event-related brain potentials and DNA methylation patterns in boys at early school age. *Sci. Rep.* **7**, 3823 (2017).
24. Sengupta, S. M., Smith, A. K., Grizenko, N. & Joober, R. Locus-specific DNA methylation changes and phenotypic variability in children with attention-deficit hyperactivity disorder. *Psychiatry Res.* **256**, 298–304 (2017).
25. van Mil, N. H. et al. DNA methylation profiles at birth and child ADHD symptoms. *J. Psychiatr. Res.* **49**, 51–59 (2014).
26. Xu, Y. et al. Multiple epigenetic factors predict the attention deficit/hyperactivity disorder among the Chinese Han children. *J. Psychiatr. Res.* **64**, 40–50 (2015).
27. Dadds, M. R., Schollar-Root, O., Lenroot, R., Moul, C. & Hawes, D. J. Epigenetic regulation of the DRD4 gene and dimensions of attention-deficit/hyperactivity disorder in children. *Eur. Child Adolesc. Psychiatry* **25**, 1081–1089 (2016).
28. Park, S. et al. Associations between serotonin transporter gene (SLC6A4) methylation and clinical characteristics and cortical thickness in children with ADHD. *Psychological Med.* **45**, 3009–3017 (2015).
29. Rovira, P. et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology 1–10 (2020).
30. Sanchez-Mora, C. et al. Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. *Neuropsychopharmacology* **44**, 890–897 (2019).
31. Epstein, J. N., Johnson, D. E. & Conners, C. K. *Conners Adult ADHD Diagnostic Interview for DSM-IV*. (Multi-Health Systems, North Tonawanda, NY, 1999).
32. Hansen, K. D., Aryee, M. & Timp, W. minfiData: example data for the Illumina Methylation 450k array. R package version 0.24.0. Bioconductor. https://doi.org/10.18129/B9.bioc.minfiData (2017).
33. Aryee, M. J. et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
34. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
35. Pidsley, R. et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC genomics* **14**, 293 (2013).
36. Chen, Y. A. et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).
37. McCartney. et al. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22–24 (2016).
38. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
39. Elliott, H. R. et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin. Epigenetics* **6**, 4 (2014).
40. Jaffe, A. E. FlowSorted.Blood.450k: Illumina HumanMethylation data on sorted blood cell populations. R package version 1.22.0. Bioconductor. https://doi.org/10.18129/B9.bioc.FlowSorted.Blood.450k (2019).
41. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
42. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).
43. Mansell, G. et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC genomics* **20**, 366 (2019).
44. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986–2988 (2012).
45. Šidák, Z. K. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* **62**, 626–633 (1967).
46. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
47. Zeng, Y. et al. Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nat. Commun.* **10**, 1383 (2019).
48. Cecil, C. A. et al. Epigenetic signatures of childhood abuse and neglect: implications for psychiatric vulnerability. *J. Psychiatr. Res.* **83**, 184–194 (2016).
49. Joehanes, R. et al. Epigenetic signatures of cigarette smoking. *Circulation Cardiovascular Genet.* **9**, 436–447 (2016).
50. Joubert, B. R. et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am. J. Hum. Genet.* **98**, 680–696 (2016).
51. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 896–905 e896 (2016).
52. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
53. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).
54. Kanda, T., Ishii, K., Kawaguchi, H., Kitajima, K. & Takenaka, D. High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: relationship with increasing cumulative dose of a gadolinium-based contrast material. *Radiology* **270**, 834–841 (2014).
55. Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).
56. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
57. Zhang, T. et al. A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes. *Nat. Commun.* **9**, 3798 (2018).
58. Ge, Y., Paisie, T. K., Newman, J. R. B., McIntyre, L. M. & Concannon, P. UBASH3A mediates risk for type 1 diabetes through inhibition of T-cell receptor-induced NF-kappaB signaling. *Diabetes* **66**, 2033–2043 (2017).
59. Sun, C. et al. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat. Genet.* **48**, 323–330 (2016).
60. Maldonado, R., Smadja, C., Mazucchelli, C. & Sassone-Corsi, P. Altered emotional and locomotor responses in mice deficient in the transcription factor CREM. *Proc. Natl Acad. Sci.* **96**, 14094–14099 (1999).
61. Lahti, T. A. & Partonen, T. CREM mutations and ADHD symptoms. *Med. Hypotheses* **72**, 544–545 (2009).
62. Miller, M. L. et al. Ventral striatal regulation of CREM mediates impulsive action and drug addiction vulnerability. *Mol. Psychiatry* **23**, 1328–1335 (2018).
63. Boison, D. & Aronica, E. Comorbidities in neurology: is adenosine the common link? *Neuropharmacology* **97**, 18–34 (2015).
64. Sandau, U. S. et al. Adenosine kinase deficiency in the brain results in maladaptive synaptic plasticity. *J. Neurosci.* **36**, 12117–12128 (2016).
65. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
66. Ling, B. et al. Suppression of non-small cell lung cancer proliferation and tumorigenicity by DENND2D. *Lung Cancer* **79**, 104–110 (2013).

67. Zhang, T. et al. Downregulation of miR-522 suppresses proliferation and metastasis of non-small cell lung cancer cells by directly targeting DENN/MADD domain containing 2D. *Sci. Rep.* **6**, 19346 (2016).
68. Park, S. L. et al. Mercapturic acids derived from the toxicants acrolein and crotonaldehyde in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. *PLoS ONE* **10**, e0124841 (2015).
69. Joelsson, P. et al. Prenatal smoking exposure and neuropsychiatric comorbidity of ADHD: a finnish nationwide population-based cohort study. *BMC psychiatry* **16**, 306 (2016).
70. Huang, L. et al. Maternal smoking and attention-deficit/hyperactivity disorder in offspring: a meta-analysis. *Pediatrics* **141**, e20172465 (2018).
71. He, C. W., Liao, C. P. & Pan, C. L. Wnt signalling in the development of axon, dendrites and synapses. *Open Biol.* **8**, 180116 (2018).
72. Friedrichs, B., Igl, W., Larsson, H. & Larsson, J. O. Coexisting psychiatric problems and stressful life events in adults with symptoms of ADHD-a large Swedish population-based study of twins. *J. Atten. Disord.* **16**, 13–22 (2012).
73. Eakin, L. et al. The marital and family functioning of adults with ADHD and their spouses. *J. Atten. Disord.* **8**, 1–10 (2004).
74. Harpin, V. A. The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. *Arch. Dis. Child* **90**(Suppl 1), i2–i7 (2005).
75. Schermerhorn, A. C. et al. Offspring ADHD as a risk factor for parental marital problems: controls for genetic and environmental confounds. *Twin Res. Hum. Genet.* **15**, 700–713 (2012).
76. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10**, 1024–1032 (2015).

Neuropsychopharmacology www.nature.com/npp

ARTICLE OPEN

Shared genetic background between children and adults with attention deficit/hyperactivity disorder

Paula Rovira[1,2], Ditte Demontis[3,4,5], Cristina Sánchez-Mora[1,2,6,7], Tetyana Zayats[8,9,10], Marieke Klein[11,12], Nina Roth Mota[11,13,14], Heike Weber[15,16], Iris Garcia-Martínez[1,2,17,18], Mireia Pagerols[1,2], Laura Vilar-Ribó[1,2], Lorena Arribas[1,2], Vanesa Richarte[2,6,19], Montserrat Corrales[2,6,19], Christian Fadeuilhe[2,6,19], Rosa Bosch[2,6,19], Gemma Español Martin[2,19,20], Peter Almos[21], Alysa E. Doyle[22,23], Eugenio Horacio Grevet[13,24], Oliver Grimm[16], Anne Halmøy[8,25], Martine Hoogman[11], Mara Hutz[26], Christian P. Jacob[21], Sarah Kittel-Schneider[16], Per M. Knappskog[27,28], Astri J. Lundervold[29], Olga Rivero[21], Diego Luiz Rovaris[13,26,30], Angelica Salatino-Oliveira[26], Bruna Santos da Silva[13,26], Evgeniy Svirin[21,31], Emma Sprooten[32], Tatyana Strekalova[21,31,33], ADHD Working Group of the Psychiatric Genomics Consortium, 23andMe Research team, Alejandro Arias-Vasquez[11,14], Edmund J. S. Sonuga-Barke[34,35], Philip Asherson[36], Claiton Henrique Dotto Bau[13,26], Jan K. Buitelaar[32,37], Bru Cormand[7,38,39,40], Stephen V. Faraone[41], Jan Haavik[8,25], Stefan E. Johansson[27,28], Jonna Kuntsi[36], Henrik Larsson[42,43], Klaus-Peter Lesch[21,31,33], Andreas Reif[16], Luis Augusto Rohde[44], Miquel Casas[1,2,6,19], Anders D. Børglum[3,4,5], Barbara Franke[11,14], Josep Antoni Ramos-Quiroga[1,2,6,19], María Soler Artigas[1,2,6,7] and Marta Ribasés[1,2,6,7]

Attention deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder characterized by age-inappropriate symptoms of inattention, impulsivity, and hyperactivity that persist into adulthood in the majority of the diagnosed children. Despite several risk factors during childhood predicting the persistence of ADHD symptoms into adulthood, the genetic architecture underlying the trajectory of ADHD over time is still unclear. We set out to study the contribution of common genetic variants to the risk for ADHD across the lifespan by conducting meta-analyses of genome-wide association studies on persistent ADHD in adults and ADHD in childhood separately and jointly, and by comparing the genetic background between them in a total sample of 17,149 cases and 32,411 controls. Our results show nine new independent loci and support a shared contribution of common genetic variants to ADHD in children and adults. No subgroup heterogeneity was observed among children, while this group consists of future remitting and persistent individuals. We report similar patterns of genetic correlation of ADHD with other ADHD-related datasets and different traits and disorders among adults, children, and when combining both groups. These findings confirm that persistent ADHD in adults is a neurodevelopmental disorder and extend the existing hypothesis of a shared genetic architecture underlying ADHD and different traits to a lifespan perspective.

Neuropsychopharmacology (2020) 0:1–10; https://doi.org/10.1038/s41386-020-0664-5

[1]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health, and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain; [2]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain; [3]Department of Biomedicine (Human Genetics), and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark; [4]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark; [5]Center for Genomics and Personalized Medicine, Aarhus, Denmark; [6]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain; [7]Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain; [8]Department of Biomedicine, University of Bergen, Bergen, Norway; [9]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA; [10]Stanley Center for Psychiatric Research, Broad Institute of MIT, and Harvard, Cambridge, MA, USA; [11]Department of Human Genetics, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; [12]University Medical Center Utrecht, UMC Utrecht Brain Center, Department of Psychiatry, Utrecht, The Netherlands; [13]ADHD Outpatient Program, Adult Division, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [14]Department of Psychiatry, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; [15]Department of Psychiatry, Psychosomatics, and Psychotherapy, University of Würzburg, Würzburg, Germany; [16]Department of Psychiatry, Psychosomatic Medicine, and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany; [17]Banc de Sang i Teixits (BST), Barcelona, Spain; [18]Grup de Medicina Transfusional, Vall d'Hebron Institut de Recerca, Universitat Autònoma de Barcelona (VHIR-UAB), Barcelona, Spain; [19]Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain; [20]Group of Psychiatry, Mental Health, and Addiction, Vall d'Hebron Research Institute (VHIR), Barcelona, Catalonia, Spain; [21]Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg, Germany; [22]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; [23]Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [24]Department of Psychiatry, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [25]Division of Psychiatry, Haukeland University Hospital, Bergen, Norway; [26]Department of Genetics, Institute of Biosciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [27]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; [28]Department of Clinical Science, University of Bergen, Bergen, Norway; [29]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway; [30]Department of Physiology and Biophysics, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil; [31]Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, IM Sechenov First Moscow State Medical University, Moscow, Russia; [32]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands; [33]Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands; [34]Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK; [35]Department of Child and Adolescent Psychiatry, Aarhus University, Aarhus, Denmark; [36]Social Genetic and Developmental Psychiatry, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, UK; [37]Karakter Child and Adolescent Psychiatry, Nijmegen, The Netherlands; [38]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain; [39]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain; [40]Institut de Recerca Sant Joan de Déu (IRSJD), Esplugues de Llobregat, Catalonia, Spain; [41]Departments of Psychiatry, of Neuroscience, and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA; [42]School of medical Sciences, Örebro University, Örebro, Sweden; [43]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden and [44]Division of Child Psychiatry, Hospital de Clinicas de Porto Alegre, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
Correspondence: María. Soler Artigas (maria.soler@vhir.org) or Marta Ribasés (marta.ribases@vhir.org)
These authors contributed equally: Paula Rovira, Ditte Demontis, Anders D. Børglum, Barbara Franke, Josep Antoni Ramos-Quiroga, María Soler Artigas, Marta Ribasés.
A list of members of ADHD Working Group of the Psychiatric Genomics Consortium and 23andMe Research team and their affiliations are listed in Supplementary information.

Received: 23 December 2019 Revised: 25 February 2020 Accepted: 3 March 2020
Published online: 12 April 2020

INTRODUCTION

Attention deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder that severely impairs the daily functioning of patients due to age-inappropriate levels of impulsivity and hyperactivity, and/or difficulties in focusing attention [1]. ADHD has a prevalence of 5–6% in childhood, and impairing symptoms persist into adulthood in around two-thirds of children with ADHD diagnosis, with an estimated adult prevalence around 3.4% [1, 2].

ADHD is a multifactorial disorder with heritability averaging 76% throughout the lifespan [3–5]. There is consistent evidence that both common and rare variants make an important contribution to the risk for the disorder [6–11]. Several genome-wide association studies (GWAS) and meta-analyses across those have been conducted [7], but only the largest GWAS meta-analysis (GWAS-MA) performed to date reported genome-wide significant loci [6]. This study concluded that common genetic variants (minor allele frequency, MAF > 0.01) account for 22% of the heritability of the disorder [6] and supported substantial genetic overlap between ADHD and other brain disorders and behavioral/cognitive traits [6, 12].

The presentation of ADHD symptoms changes from childhood to adulthood, with lower levels of hyperactivity in adulthood but a high risk for ongoing attention problems, disorganization, and emotional dysregulation [13, 14]. As in the general population, the pattern of psychiatric and somatic comorbid conditions in ADHD also changes substantially over time, with learning disabilities, oppositional defiant disorder, and conduct disorder being more prevalent in children, and substance use disorders, social phobia, insomnia, obesity, and mood disorders becoming more pronounced in adulthood [1, 15–18]. In addition, persistent ADHD in adults is, compared with the general population (and to cases with remitting ADHD), associated with higher risk for a wide range of functional and social impairments, including unemployment, accidents, and criminal behavior [7, 19–23].

Several risk factors measured in childhood predict the persistence of ADHD symptoms into adulthood, such as the presence of comorbid disorders, the severity of ADHD symptoms, being exposed to psychosocial adversity, as well as having a high polygenic risk score (PRS) for childhood ADHD [24–28]. Twin studies suggest that both stable and dynamic genetic influences affect the persistence of ADHD symptoms [4, 5, 29, 30]. However, specific genetic factors differentiating childhood and persistent ADHD into adulthood are not well understood due to the lack of longitudinal studies. Molecular studies, including the most recent GWAS-MA of ADHD [6], have been performed in children and adults either separately or jointly [6, 31–40], but large-scale analyses comparing their genetic basis are yet to be conducted.

Given this background, we set out to study the contribution of common genetic variants to the risk for ADHD from a lifespan perspective by conducting the largest GWAS-MAs performed so far on persistent ADHD in adults (diagnosed according to DSM-IV/ICD-10 criteria) and on ADHD in childhood (that may include remittent and persistent forms of the disorder) separately and jointly. For the first time, we estimated the genetic correlation between childhood and persistent ADHD, compared their patterns of genetic correlation with other traits and disorders, assessed the effect of childhood ADHD PRSs on persistent ADHD, and explored whether individuals in which ADHD symptoms may persist into adulthood could be distinguished already in childhood using genetic data.

MATERIAL AND METHODS

Sample description

A total of 19 GWAS of ADHD comprising 49,560 individuals (17,149 cases and 32,411 controls), provided by the Psychiatric Genomics Consortium (PGC), the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), and the International Multi-centre persistent ADHD CollaboraTion (IMpACT), were analyzed. All participants were of European ancestry, had provided informed consent, and all sites had documented permission from local ethics committees.

The meta-analysis on persistent ADHD was conducted in 22,406 individuals (6,532 ADHD adult cases and 15,874 controls) using six datasets from the IMpACT consortium, two datasets from the PGC, and the adult subset from the iPSYCH cohort included in Demontis and Walters et al. [6]. The meta-analysis on ADHD in childhood included 27,154 individuals (10,617 cases and 16,537 controls), comprising two Brazilian and Spanish cohorts, seven datasets from the PGC, and the children subset from the iPSYCH cohort included in Demontis and Walters et al. [6]. All patients met DSM-IV/ICD-10 diagnostic criteria. In total, 7,086 new samples not included in Demontis and Walters et al. [6] were considered in the present study. Detailed information on each dataset is provided in Table S1 and in Supplementary Methods.

GWAS and meta-analyses

Genotyping platforms and quality control (QC) filters for each of the datasets are shown in Table S1. Pre-imputation QC at individual and SNP level were performed using the Rapid Imputation and COmputational PIpeLIne with the default settings (https://sites.google.com/a/broadinstitute.org/ricopili/). Non-European ancestry samples, related and duplicated individuals, and subjects with sex discrepancies were excluded. Phasing of genotype data was performed using the SHAPEIT2 algorithm, and imputation for unrelated samples and trios was performed with MaCH, IMPUTE2, or MINIMAC3 (http://genome.sph.umich.edu/wiki/Minimac3) depending on software availability at the time of imputation (Table S1) [41–43]. The European ancestry panel of the 1000 Genomes Project using genome build hg19 was considered as reference for genotype imputation (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/). After imputation, the association with ADHD of genotype dosages was tested using logistic regression in PLINK 1.9 [44], assuming an additive genetic model and including sex, the first ten principal components, and other relevant covariates for each case-control study (Table S1). GWAS summary statistics were filtered prior to meta-analysis, excluding variants with MAF < 0.01, and imputation quality scores (INFO) ≤ 0.8. Inverse-variance weighted fixed-effects meta-analyses were conducted using METAL [45] and results were filtered by effective sample size > 70% of the total, defined as Neff $= \frac{2}{\left(\frac{1}{Nca}\right)+\left(\frac{1}{Nco}\right)}$ [46]. The genome-wide significance threshold was set at $P < 5.00E{-}08$ to correct for multiple testing. Independent loci for variants exceeding this threshold were defined based on clumping using PLINK 1.9. Variants that were ±250 kb away from the index variant (variant with smallest P value in the region), with P value < 0.001, and with an estimated linkage disequilibrium (LD) of $r^2 > 0.2$ with the index variant were assigned to a clump ($p_1 = 5.00E{-}08$, $p_2 = 0.001$, $r^2 = 0.2$, kb = 250). Manhattan and Forest plots were generated using the "qqman" and "forestplot" R packages (3.4.4R version), respectively. The LocusZoom software [47] was used to generate regional association plots.

Details of downstream analyses for top signals identified are provided in the online supplement and include conditional analysis, Bayesian credible set analysis, and functional characterization of the significant variants.

SNP-based heritability (SNP-h^2)

The SNP-h^2 was estimated by single-trait LD score regression using summary statistics, HapMap 3 LD-scores, considering default SNP QC filters (INFO > 0.9 and MAF > 0.01) and assuming population prevalence of 3.4, 5.5, and 5% for persistent ADHD, ADHD on childhood, and ADHD across the lifespan, respectively, [48]. Data of 1,113,287, 1,072,558, and 1,092,418 SNPs from the GWAS-MA of

persistent ADHD, ADHD on childhood, and ADHD across the lifespan, respectively, were considered to estimate the liability-scale SNP-h^2. Partitioning and enrichment of the heritability by functional categories was analyzed using the 24 main annotations (no window around the functional categories) described by Finucane et al. [49]. Statistical significance was set using Bonferroni correction ($P < 2.08E{-}03$).

Gene-based and gene-set analyses

MAGMA software was undertaken for gene-based and gene-set association testing using summary data from our GWAS-MAs [50]. Variants were mapped to a gene if they were within 20 kb upstream or downstream from the gene according to dbSNP build 135 and NCBI 37.3 gene definitions. Genes in the MHC region (hg19:chr6:25-35M) were excluded from the analyses. LD patterns were estimated using the European ancestry reference panel of the 1000 Genomes Project. Gene sets denoting canonical pathways were downloaded from MSigDB (http://www.broadinstitute.org/gsea/msigdb), which integrates Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/), BioCarta (http://www.biocarta.com/), Reactome (https://reactome.org/), and Gene Ontology (GO) (http://www.geneontology.org/) resources. Bonferroni correction ($P < 2.77E{-}06$ for 18,038 genes in persistent ADHD; $P < 2.75E{-}06$ for 18,218 genes in childhood ADHD; $P < 2.79E{-}06$ for 17,948 genes in ADHD across the lifespan) and 10,000 permutations were used for multiple testing correction in the gene-based and gene-set analyses, respectively.

BUHMBOX analysis

The Breaking Up Heterogeneous Mixture Based On cross(X)-locus correlations (BUHMBOX) analysis [51] was used to test whether the genetic correlation between persistent ADHD and ADHD in childhood was driven by subgroup heterogeneity, found when there is a subset of children enriched for persistent ADHD-associated alleles. Subgroup heterogeneity was tested in each childhood dataset considering independent SNPs ($r^2 = 0.1$, kb = 10,000) with MAF > 0.05 from the GWAS-MA of persistent ADHD using two different P value thresholds of $P < 5.00E{-}05$ (62 SNPs) and $P < 1.00E{-}03$ (710 SNPs). Results were meta-analyzed using the standard weighted sum of z-score approach, where z-scores are weighted by the square root of the effective sample size. The statistical power was calculated using 1,000 simulations, considering the ADHD children meta-analysis sample size, the odds ratios and risk allele frequencies from the GWAS-MA of persistent ADHD, and assuming 65% of heterogeneity proportion (π).

Sign test

The direction of the effect of variants associated with ADHD in childhood was tested in persistent ADHD and vice versa, using strict clumping ($r^2 = 0.05$, kb = 500, $p_2 = 0.5$) and different P value thresholds (1.00E−07, 5.00E−07, 1.00E−06, 5.00E−06, 1.00E−05, 5.00E−05, 1.00E−04, and 5.00E−04). The concordant direction of effect was evaluated using a one sample test of the proportion with Yates' continuity correction against a null hypothesis of $P = 0.50$ with the "stats" R package.

Polygenic risk scoring

PRSs were constructed using different P value thresholds ($P < 0.001$, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) to select independent variants ($p_1 = 1$, $p_2 = 1$, $r^2 = 0.1$, kb = 250) from the childhood GWAS-MA of ADHD and were then tested for association with persistent ADHD in each of the nine datasets, adjusting for the covariates included in the GWAS and using PRSice-2 (https://choishingwan.github.io/PRSice/). Best guess genotypes for nonambiguous strand variants present in all the persistent ADHD studies (missing rate < = 0.02) were included ($N_{SNPs} = 32{,}584$ for $P = 1$). Results from the nine PRS analyses at each P value threshold were combined using inverse-variance weighted meta-analysis.

Genetic correlation

Cross-trait LD score regression with unconstrained intercept was used to calculate genetic correlations (rg) between pairs of traits, considering HapMap3 LD-scores, markers with INFO ≥ 0.90, and excluding the MHC region (hg19:chr6:25-35M) [48]. Other ADHD datasets [6, 52] and phenotypes from the LD-hub centralized database [53] with heritability z-scores (observed heritability/observed standard error) >4 and with an observed heritability > 0.1 were considered ($N = 139$ out of 689 available traits). Statistical significance was set using Bonferroni correction ($P < 3.60E{-}04$). Pearson's correlation coefficient (Pearson's r) was calculated between the genetic correlations of persistent ADHD with the phenotypes from the LD-hub and the genetic correlations of ADHD in childhood with the phenotypes from the LD-hub.

RESULTS

GWAS-MA of persistent ADHD in adults

The GWAS-MA of persistent ADHD in adults included 6,532 adult ADHD cases and 15,874 controls. Minimal population stratification or other systematic biases were detected (LD score regression intercept = 1.01, Fig. S1a). The proportion of heritability of persistent ADHD attributable to common single-nucleotide polymorphisms on the liability scale (SNP-h^2) was 0.19 (SE = 0.024), with a nominally significant enrichment in the heritability of variants located in conserved genomic regions ($P = 5.18E{-}03$) and in the cell-specific histone mark H3K4me1 ($P = 3.17E{-}02$) (Fig. S2a). The gene-based analysis revealed six genes in four loci (*ST3GAL3, FRAT1/FRAT2, CGB1*, and *RNF225/ZNF584*) significantly associated with persistent ADHD, with *ST3GAL3* being the most significant one ($P = 8.72E{-}07$) (Table S2a). The single-marker analysis showed no variants exceeding genome-wide significance, with the most significant signal being rs3923931 ($P = 1.69E{-}07$) (Fig. 1a and Table S3a). Similarly, no significant gene sets were identified in the pathway analysis after correction for multiple comparisons (Table S4a [excel file]).

GWAS-MA of ADHD in childhood

To compare the genetic background between persistent ADHD in adults and ADHD in childhood (that may include future remittent and persistent forms of the disorder), we conducted a GWAS-MA on children with ADHD in a total of 10,617 ADHD cases and 16,537 controls. We found no evidence of genomic inflation or population stratification (LD score regression intercept = 1.02, Fig. S1b). The liability-scale SNP-h^2 for ADHD in childhood was 0.19 (SE = 0.021), with a significant enrichment in the heritability of variants located in conserved genomic regions after Bonferroni correction ($P = 1.21E{-}06$) (Fig. S2b). The gene-based analysis highlighted a significant association between *FEZF1* and ADHD in childhood ($P = 5.42E{-}07$) (Table S2b). No single genetic variant exceeded genome-wide significance, with the top signal being in rs55686778 ($P = 1.67E{-}07$) (Fig. 1b and Table S3b), and no significant gene sets were identified in the pathway analysis after correction for multiple comparisons (Table S4b [excel file]).

Comparison of the genetic background of persistent ADHD in adults and ADHD in childhood

We found a strong genetic correlation between persistent ADHD in adults and ADHD in childhood (rg = 0.81, 95% CI: 0.64–0.97), significantly different from 0 ($P = 2.13E{-}21$) and from 1 ($P = 0.02$). Sign test results provided evidence of a consistent direction of effect of genetic variants associated with ADHD in childhood in persistent ADHD and vice versa ($P = 6.60E{-}04$ and $P = 4.47E{-}03$, respectively, for variants with $P < 5.00E{-}05$ in each dataset) (Table S5). In addition, PRS analyses showed that childhood ADHD PRSs were associated with persistent ADHD at different predefined P value thresholds, with the $P = 0.40$ threshold ($N_{SNPs} = 20{,}398$) explaining the most variance ($r^2 = 0.0041$ and $P = 1.20E$

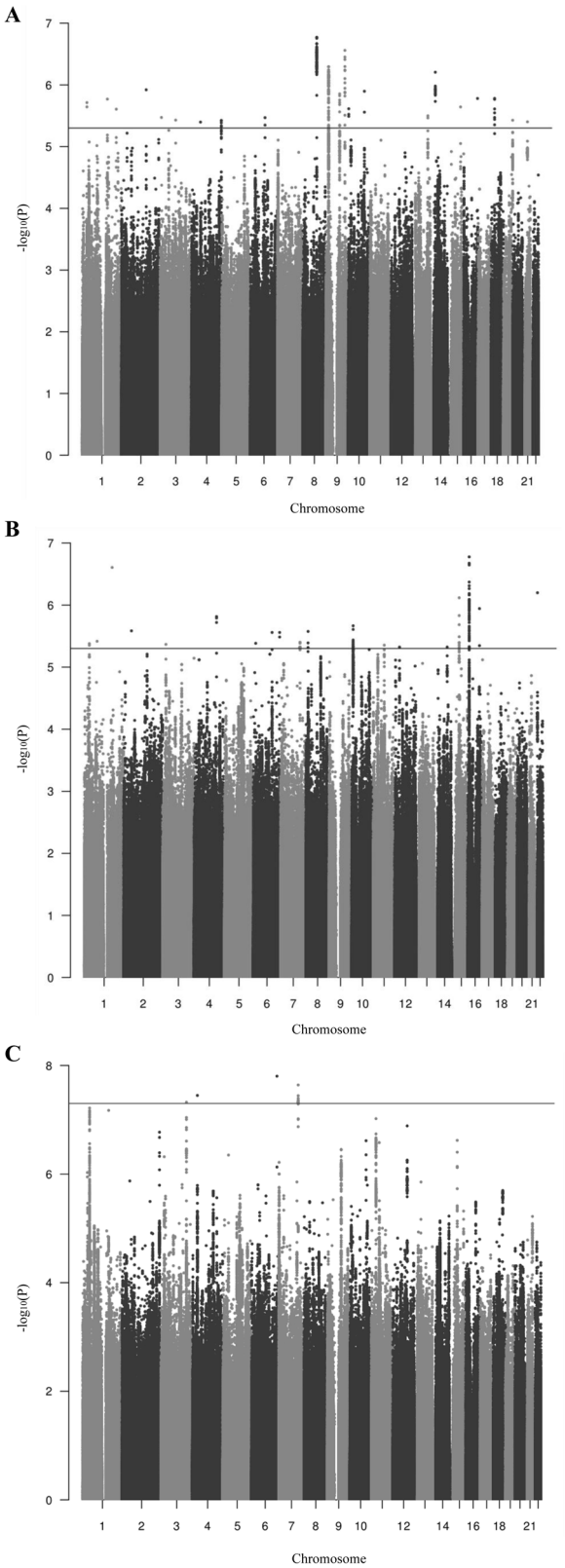


Fig. 1 Manhattan plots of the three GWAS meta-analyses conducted. (**a**) GWAS-MA of nine cohorts of persistent ADHD in adults, (**b**) GWAS-MA of ten cohorts of ADHD in childhood, and (**c**) GWAS-MA of all datasets of ADHD across the lifespan (ADHD in childhood + persistent ADHD). Horizontal lines indicate suggestive (*P* value = 5.00E−06) and genome-wide significant (*P* = 5.00E−08) thresholds in **a**-**b**, and **c**, respectively.

−27) (Fig. 2a). The quintiles of the PRS built using this threshold showed the expected trend of higher ADHD risk for individuals in higher quintiles (Fig. 2b, Table S6).

We then tested whether the genetic correlation between persistent ADHD and ADHD in childhood was driven by a subset of children enriched for persistent ADHD-associated alleles using the Breaking Up Heterogeneous Mixture Based On Cross-locus correlations (BUHMBOX) analysis. We found no evidence of subgroup genetic heterogeneity in children, supporting that the sharing of persistent ADHD-associated alleles between children and adults was driven by the whole group of children, with a statistical power of 98.4 and 100% for thresholds of $P < 5.00E{-}05$ and $P < 1.00E{-}03$, respectively (Table S7).

GWAS-MA of ADHD across the lifespan

Given the strong genetic correlation between persistent ADHD in adults and in childhood, we performed a GWAS-MA of ADHD across the lifespan considering all datasets included in the GWAS-MAs. In total, 17,149 ADHD cases and 32,411 controls were included, and no evidence of genomic inflation or population stratification was found (LD score regression intercept = 1.03, Fig. S1c). The liability-scale SNP-h^2 for ADHD across the lifespan was 0.17 (SE = 0.013), and a significant enrichment in the heritability of variants located in conserved genomic regions was observed after Bonferroni correction ($P = 1.53E{-}06$) (Fig. S2c). We identified four genome-wide significant variants (Figs. 1c and 3, Table 1a, and Fig. S3) and nine genes in seven loci (*FEZF1, DUSP6, ST3GAL3/KDM4A, SEMA6D, C2orf82/GIGYF2, AMN*, and *FBXL17*) significantly associated with ADHD across the lifespan (Table 1b). The most significantly associated locus was on chromosome 6 (index variant rs183882582-T, OR = 1.43 (95% CI: 1.26–1.60), $P = 1.57E{-}08$), followed by loci on chromosome 7 (index variant rs3958046), chromosome 4 (index variant rs200721207), and chromosome 3 (index variant rs1920644) (Table 1a, Fig. 3). The gene-set analysis showed a significant association of the "ribonucleoprotein complex" GO term with ADHD across the lifespan (*P*.adj = 0.021) (Table S4c [excel file]).

One of the four loci identified in the single-variant analysis also reached genome-wide significance in the previous GWAS-MA on ADHD [6], and all of them showed consistent direction of the effect in that study (Table S8a). Significant loci reported by Demontis et al. [6] showed nominal association with ADHD across the lifespan in our study (Table S8b, c), with single variant hits showing the same direction of the effect (Table S8b).

Analyses conditioning on the index variant for the four ADHD-associated loci did not reveal new independent markers. These four significant loci were functionally characterized by obtaining Bayesian credible sets and searching for expression quantitative trait loci (eQTL) using available data in blood or brain [54, 55]. We found that credible sets for three of the four loci contained at least one eQTL within 1 Mb of the index variant. The credible set on chromosome 6 included the index variant (rs183882582) and rs12197454. This variant, in LD with the index variant ($r^2 = 0.56$), was associated with the expression of *RSPH3* in blood and brain (*P*.adj < 1.65E−05 and *P*.adj = 2.36E−07, respectively), and with the expression of *VIL2* in blood (*P*.adj = 3.21E−03). The credible set for the second most associated locus on chromosome 7 included 24 variants. The index variant, rs3958046, and other variants in this set, were eQTLs for *CADPS2* in brain (maximum *P*.adj = 2.91E−03). The credible set for the locus on chromosome 4 contained 50 variants, most of them located in or near *PCDH7*, but no eQTLs were identified. In the credible set for the locus on chromosome 3, which included 98 variants, the index variant, rs1920644, was associated with the expression of *KPNA4*, *IFT80*, and *KRT8P12* in brain (*P*.adj = 1.16E−04, *P*.adj = 1.40E−03, and *P*.adj = 1.77E−03, respectively). Many other variants in this set were eQTLs for these genes and also for *TRIM59*, *OTOL1*, and/or *C3orf80* in brain (*P*.adj < 0.05) (Table S9 [excel file]).

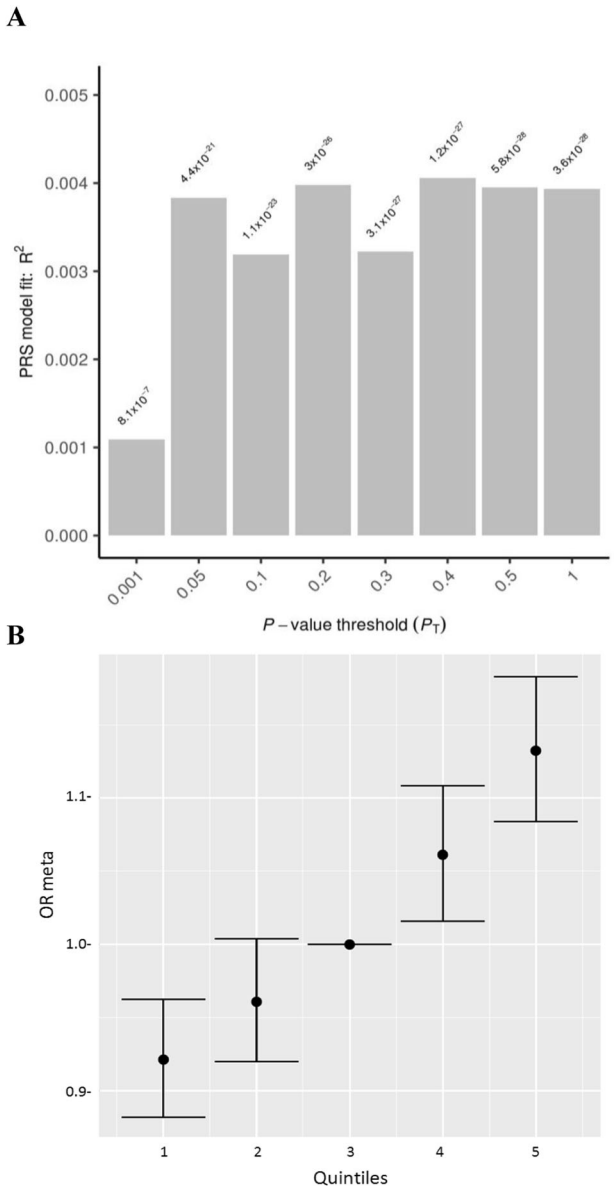


Fig. 2 Polygenic risk scores for ADHD in childhood tested on persistent ADHD as target sample. a Bar plot and **b** quintile plot of meta-analysis odds ratios (OR meta) with 95% confidence intervals for *P* value threshold = 0.4 using the third quintile as baseline.

In a summary-data-based Mendelian randomization (SMR) analysis, we used summary data from the GWAS-MA of ADHD across the lifespan and the eQTL data in blood and brain from Westra et al. [54] and Qi et al. [55] to identify gene expression levels associated with ADHD. We found a significant association between ADHD across the lifespan and *RMI1* expression in blood (P_{SMR} = 5.36E−06) (Table S10 [in excel]), finding not likely to be an artifact due to LD between eQTL and other ADHD-associated variants given that the P_{HEIDI} was 0.47.

Genetic correlation with other ADHD datasets and phenotypes

We found significant genetic correlations of ADHD in children and adults from the previous GWAS-MA [6] (*N* = 53,296) and persistent ADHD (rg = 0.85, SE = 0.04, *P* = 5.49E−99), ADHD in childhood (rg = 0.99, SE = 0.03, *P* = 5.02E−273), and ADHD across the lifespan (rg = 0.98, SE = 0.01, *P* < 2.23E−308) (Table S11). When removing sample overlap (LD score genetic covariance intercept = 0.75) and considering only the subset of new samples included in our GWAS-MA on ADHD across the lifespan (*N* = 7086), a significant genetic correlation was also obtained between their sample and ours (rg = 0.91, SE = 0.35, *P* = 8.70E−03).

We also observed significant genetic correlations between childhood ADHD symptom scores from a GWAS-MA in a population of children reported by the EAGLE consortium [52] (*N* = 17,666) and persistent ADHD (rg = 0.65, SE = 0.20, *P* = 1.10E−03), ADHD in childhood (rg = 0.98, SE = 0.21, *P* = 2.76E−06), and ADHD across the lifespan (rg = 0.87, SE = 0.19, *P* = 4.80E−06). Similarly, significant genetic correlations between GWAS of self-reported ADHD status from 23andMe (*N* = 952,652) and persistent ADHD (rg = 0.75, SE = 0.05, *P* = 2.49E−45), ADHD in childhood (rg = 0.63, SE = 0.05, *P* = 1.39E−42), and ADHD across the lifespan (rg = 0.72, SE = 0.04, *P* = 4.86E−88) were observed (Table S11).

We also estimated the genetic correlation of persistent ADHD in adults, ADHD in childhood, and ADHD across the lifespan with all available phenotypes in LD-hub. Results for 139 phenotypes passed the QC parameters and 41 genetic correlations were significant after Bonferroni correction in both children and adults with persistent ADHD (Table S12 [excel file]). Again, the genetic correlations with ADHD were consistent across the lifespan, with similar patterns found in adulthood and childhood (Pearson's *r* = 0.89) (Fig. 4a, Table S12 [excel file]). The strongest genetic correlations with ADHD were found for traits related to academic performance, intelligence, and risk-taking behaviors, including smoking and early pregnancy (Fig. 4b).

DISCUSSION

In the current study, we set out to explore the contribution of common genetic variants to the risk of ADHD across the lifespan by conducting GWAS-MAs separately for children and adults with persistent ADHD that meet DSM-IV/ICD-10 criteria. Using the largest GWAS datasets available from the PGC, the iPSYCH, and IMpACT consortia we found evidence for a common genetic basis for ADHD in childhood and persistent ADHD in adults and identified nine new loci associated with the disorder.

We found a highly similar proportion of the heritability of ADHD explained by common variants in children and in adults (SNP-h^2 = 0.19), which is consistent with the SNP-h^2 estimate reported in the recent GWAS-MA on ADHD [6] (SNP-h^2 = 0.22), that included children and adults, and is in line with multiple studies supporting the stability of ADHD's heritability from childhood to adulthood [3–5]. These results together with the 0.81 genetic correlation found between children and adults with persistent ADHD reinforce the hypothesis of the neurodevelopmental nature of persistent ADHD in adults. Consistently, the sign test and the PRS analysis confirmed the extensive overlap of common genetic risk variants for ADHD in childhood and adulthood.

In the view of the fact that children with ADHD may be an admixed group of individuals whose ADHD symptoms will persist or remit in adulthood, we ran a BUHMBOX analysis to elucidate if the potential "persistent" individuals could be distinguishable already in childhood. Our data supported genetic similarities in ADHD across the lifespan with no evidence of a subset of patients enriched for persistent ADHD-associated alleles within the group of children.

Despite not having identified specific genetic contributions for ADHD in children or persistent ADHD, our results are not inconsistent with evidence suggesting changes in the genetic contribution to ADHD symptoms from childhood into adulthood, as described in previous twin studies in the general population [4, 5, 29, 30]. Our study design and the still limited statistical power of the GWAS-MAs may have facilitated the identification of the shared genetic basis rather than specific genetic factors for persistence. Also, differences between the origin of the samples (population-based versus clinical) and/or discrepancies between

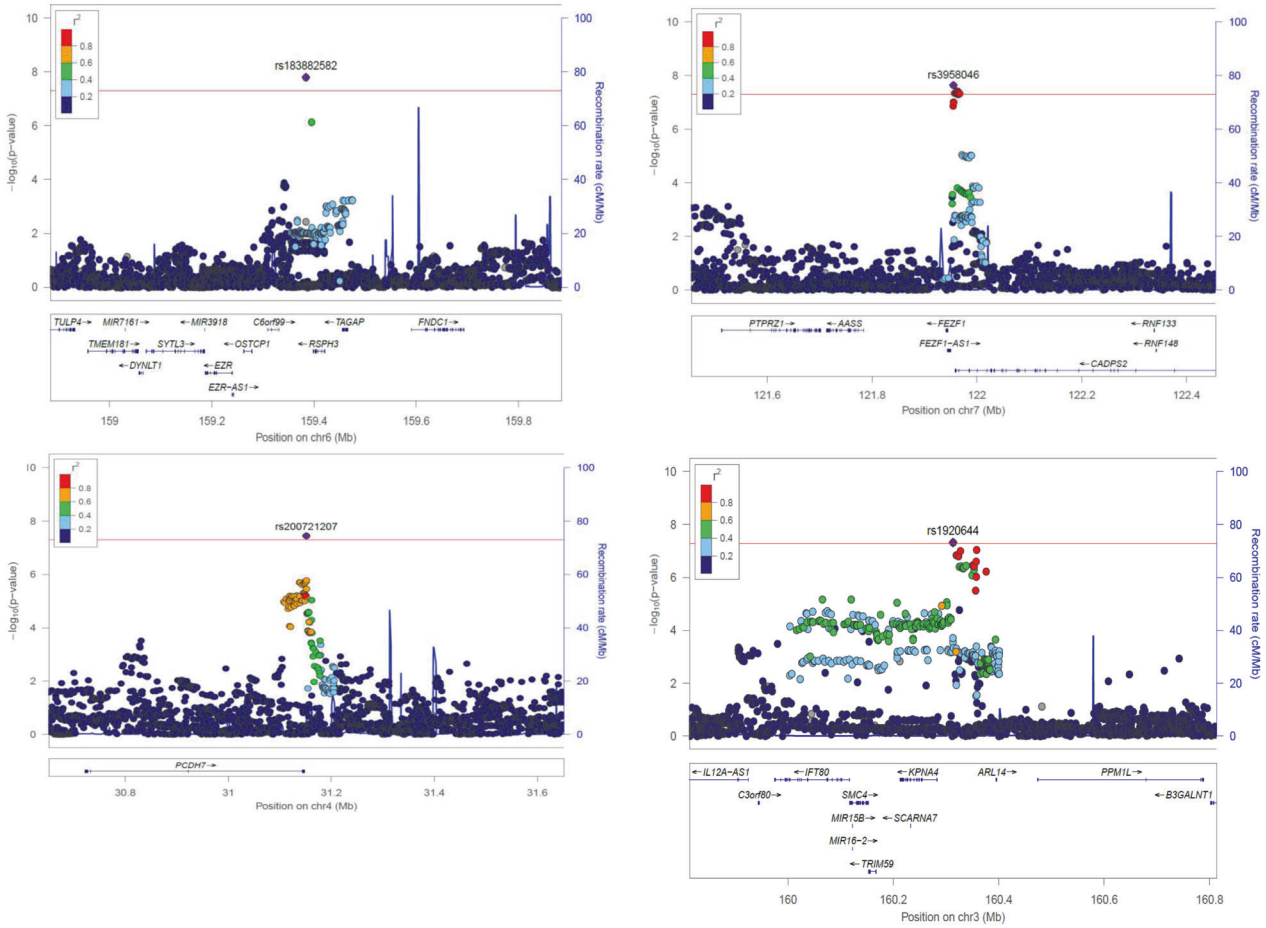


Fig. 3 Regional association plots for genome-wide significant loci identified in the GWAS meta-analysis of ADHD across the lifespan. Each plot includes information about the locus, the location and orientation of the genes in the region, the local estimates of recombination rate (in the right corner), and the LD estimates of surrounding SNPs with the index SNP (r^2 values are estimated based on 1000 Genomes European reference panel), which is indicated by color (in the upper left corner).

self- and medical reports could explain why we found no group-specific genetic variants. In addition, given that Chen et al. [56] and Biederman et al. [57] reported that persistence of ADHD into adulthood indexed stronger familial aggregation of ADHD, we cannot yet discard influences of non-additive genetic effects, or other types of genetic variation, such as rare mutations or copy number variation, playing a role in the different ADHD trajectories across the lifespan.

We also found strong and significant positive genetic correlations of ADHD ascertained in clinical populations of adults, children, or both with other ADHD-related measures from general population samples, including the largest GWAS of self-reported ADHD status from 23andMe participants ($N = 952{,}652$) and the GWAS-MA of childhood rating scales of ADHD symptoms in the general population [52]. In agreement with previous reports, these data suggest that a clinical diagnosis of ADHD in adults is an extreme expression of continuous heritable traits [6] and that a single question about ever having received an ADHD diagnosis, as in the 23andMe sample, may be informative for molecular genetics studies.

Similar patterns of genetic correlation of ADHD with different somatic disorders and anthropometric, cognitive, and educational traits were identified for children and adults. These findings were highly similar to those observed in the recent GWAS-MA [6] and further extend the existing hypothesis of a shared genetic architecture underlying ADHD and these traits to a lifespan perspective.

We report 13 loci in gene- and SNP-based analyses for childhood ADHD, adult ADHD, and/or ADHD across the lifespan. Four ADHD-associated loci were previously identified by Demontis et al. [6], which was expected due to the sample overlap between the two datasets. The new loci identified in the present study mainly included genes involved in brain formation and function, such as *FEZF1*, a candidate for autism spectrum disorder implicated in the formation of the diencephalon [58, 59], *RSPH3*, which participates in neuronal migration in embryonic brain [60], *CADPS2*, which has been associated with psychiatric conditions due to its role in monoamine and neurotrophin neurotransmission [61–64], *AMN*, which is involved in the uptake of vitamin B12 [65, 66], essential for brain development, neural myelination, and cognitive function [67], and *FBXL17*, which has previously been related to intelligence [68].

The main limitation of this study is the sample overlap (85.7%) between the present GWAS-MAs and the previous one by Demontis et al. [6], which highlighted loci previously associated with ADHD. Although sample overlap may have inflated the genetic correlation found between these studies, the estimate remained strong and significant when excluding nonoverlapping datasets.

In summary, the present cross-sectional analyses identify new genetic loci associated with ADHD and, more importantly, support the hypothesis that persistent ADHD in adults is a neurodevelopmental disorder that shows a high and significant genetic overlap with ADHD in children. Future longitudinal studies will be required

Table 1. Genome-wide significant loci in the GWAS meta-analysis of ADHD across the lifespan identified through (A) single-variant analysis and (B) gene-based analysis.

Chr	BP	SNP	Effect allele	Freq effect allele	OR	CI 95%	*P* value	Gene
A								
6	159384224	rs183882582	T	0.98	1.43	1.26–1.60	1.57E−08	*RSPH3* (+14 kb)
7	121955328	rs3958046	T	0.40	1.09	1.06–1.10	2.28E−08	*CADPS2* (+3.2 kb)/*FEZF1* (−13.9 kb)/*FEZF1-AS1* (+5.2 kb)
4	31151465	rs200721207	T	0.66	1.10	1.06–1.13	3.56E−08	*PCDH7* (−3.0 kb)
3	160313354	rs1920644	T	0.52	1.09	1.05–1.12	4.74E−08	*BC125159* (+27.9 kb)/*KPNA4* (−30 kb)/*ARL14* (−81.6 kb)

Gene	Chr	Start	Stop	*N* SNPs*	*N* PARAM**	Z-STAT	*P* value
B							
FEZF1	7	121921373	121971173	108	18	5.6	9.57E−09
DUSP6	12	89721837	89766296	103	12	5.4	3.51E−08
ST3GAL3	1	44153204	44416837	521	19	5.4	3.58E−08
SEMA6D	15	47456403	48086420	1565	55	5.3	7.24E−08
KDM4A	1	44095797	44191189	169	13	4.9	4.34E−07
C2orf82	2	233713724	233761111	138	17	4.8	7.74E−07
GIGYF2	2	233542015	233745287	511	19	4.8	8.36E−07
AMN	14	103368993	103417179	101	21	4.6	2.56E−06
FBXL17	5	107174734	107738080	1273	35	4.6	2.59E−06

The location (chromosome (Chr) and base position (BP)), effect allele and its frequency, odds ratio (OR) of the effect allele with 95% confidence interval (CI 95%) and association *P* values, along with genes in the locus are shown for each index variant ID (SNP). For the gene-based results, the number of single-nucleotide polymorphisms in the genes (*) and the number of relevant parameters used in the model by MAGMA software (**) are given.

Fig. 4 Genetic correlation of ADHD and several traits. a Dots represent genetic correlations (rg) for all traits considered (with $h^2 > 0.1$ and *z*-score > 4) and those traits that met Bonferroni correction in both children and adult ADHD groups are presented in grey. *r* indicates Pearson's correlation coefficient. **b** The ten strongest genetic correlations (with 95% confidence intervals) surpassing Bonferroni corrections in the children and persistent ADHD analysis are shown for each trait and ADHD.

to disentangle the role of common genetic variants on ADHD remittance and/or persistence.

FUNDING AND DISCLOSURE

VR has served on the speakers for Eli Lilly, Rubio, and Shire in the last 5 years. She has received travel awards from Eli Lilly and Co. and Shire for participating in psychiatric meetings. The ADHD Program has received unrestricted educational and research support from Eli Lilly and Co., Janssen-Cilag, Shire, Rovi, Psious, and Laboratorios Rubió in the past 2 years. MC received travel awards for taking part in psychiatric meetings from Shire. CF received travel awards for taking part in psychiatric meetings from Shire and Lundbeck. GEM received travel awards for taking part in psychiatric meetings from Shire. EJSS-B received speaker fees, consultancy, research funding, and conference support from Shire Pharma. Consultancy from Neurotech Solutions, Copenhagen University and Berhanderling, Skolerne & KU Leuven. Book royalties from OUP and Jessica Kingsley. Financial support—Arhus University and Ghent University for visiting Professorship. Editor-in-Chief JCPP—supported by a buy-out of time to University of Southampton and personal Honorarium. King's College London

received payments for work conducted by PA: consultancy for Shire, Eli Lilly, Novartis, and Lundbeck; educational and/or research awards from Shire, Eli Lilly, Novartis, Vifor Pharma, GW Pharma, and QbTech; speaker at events sponsored by Shire, Eli Lilly, Janssen-Cilag, and Novartis. JKB has been in the past 3 years a consultant to/member of advisory board of and/or speaker for Shire, Roche, Medice, and Servier. He is not an employee of any of these companies, and not a stock shareholder of any of these companies. He has no other financial or material support, including expert testimony, patents, royalties. In the past year, SVF received income, potential income, travel expenses continuing education support and/or research support from Tris, Otsuka, Arbor, Ironshore, Shire, Akili Interactive Labs, Enzymotec, Sunovion, Supernus, and Genomind. With his institution, he has US patent US20130217707 A1 for the use of sodium–hydrogen exchange inhibitors in the treatment of ADHD. He also receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health, Oxford University Press: Schizophrenia: The Facts and Elsevier: ADHD: Non-Pharmacologic Interventions. He is principal investigator of www.adhdinadults.com. JK has given talks at educational events sponsored by Medice; all funds are received by King's College London and used for studies of ADHD. HL has served as a speaker for Evolan Pharma and Shire and has received research grants from Shire; all outside the submitted work. K-PL served as a speaker for Eli Lilly and received research support from Medice, and travel support from Shire, all outside the submitted work. LAR reported receiving honoraria, serving on the speakers' bureau/advisory board, and/or acting as a consultant for Eli Lilly, Janssen-Cilag, Novartis, and Shire in the last 3 years; receiving authorship royalties from Oxford Press and ArtMed; and receiving travel awards from Shire for his participation in the 2015 WFADHD meetings and from Novartis to take part of the 2016 AACAP meeting. The ADHD and juvenile bipolar disorder outpatient programs chaired by him received unrestricted educational and research support from the following pharmaceutical companies in the last 3 years: Janssen-Cilag, Novartis, and Shire. MC has received travel grants and research support from Eli Lilly and Co., Janssen-Cilag, Shire, and Lundbeck and served as consultant for Eli Lilly and Co., Janssen-Cilag, Shire, and Lundbeck. BF has received educational speaking fees from Medice and Shire. JAR-Q was on the speakers' bureau and/or acted as consultant for Eli Lilly, Janssen-Cilag, Novartis, Shire, Lundbeck, Almirall, Braingaze, Sincrolab, Medice, and Rubió in the last 5 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Janssen-Cilag, Rubió, Shire, Medice, and Eli- Lilly. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 5 years: Eli Lilly, Lundbeck, Janssen- Cilag, Actelion, Shire, Ferrer, Oryzon, Roche, Psious, and Rubió. The other authors have nothing to disclose. All members of the 23andMe Research Team are current or former employees of 23andMe, Inc. and hold stock or stock options in 23andMe. Authors of the ADHD Working group of the PGC that participated in this study: Catharina A. Hartman, Ziarih Hawi, Jennifer Crosbie, Sandra Loo, Josephine Elia, Russell Schachar, Christie Burton, Ted Reichborn-Kjennerud, Aribert Rothenberger, Søren Dalsgaard, Irwin Waldman, Mark Bellgrove, Hakon Hakonarson, Johannes Hebebrand, Anke Hinney, and Robert Oades have nothing to disclose. Joseph Biederman 2019–2020: he received research support from Genentech, Headspace Inc., Lundbeck AS, Neurocentria Inc, Pfizer Pharmaceuticals, Roche TCRC Inc., Shire Pharmaceuticals Inc., Sunovion Pharmaceuticals Inc., and Tris. He was a consultant for Akili, Avekshan, Jazz Pharma, and Shire/Takeda. Through MGH CTNI, he participated in a scientific advisory board for Supernus. Dr Biederman's program has received departmental royalties from a copyrighted rating scale used for ADHD diagnoses, paid by Bracket Global, Ingenix, Prophase, Shire, Sunovion, and Theravance; these royalties were paid to the Department of Psychiatry at MGH. Tobias Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire, and Infectopharm. He received conference support or speaker's fee by Lilly, Medice, and Shire. He received royalties from Hogrefe, Kohlhammer, CIP Medien, and Oxford University Press. James McGough is an expert testimony for Eli Lilly; DSMB for Sunovion. Benjamin Neale is a member of the scientific advisory board at Deep Genomics and consultant for Camp4 Therapeutics, Takeda Pharmaceutical, and Biogen. Ole Andreas Andreassen received speaker's honorarium from Lundbeck, and is a consultant for HealthLytix.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 602805 (Aggressotype) as well as from the European Union H2020 Programme (H2020/2014–2020) under grant agreements n° 643051 (MiND), n° 667302 (CoCA), n° 728018 (Eat2beNICE), and n° 278948 (TACTICS). The work was also supported by the ECNP Network "ADHD across the Lifespan" (https://www.ecnp.eu/research-innovation/ECNP-networks/List-ECNP-Networks/ADHD.aspx). Over the course of this investigation, PR is a recipient of a pre-doctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI_B 00899). The iPSYCH project is funded by the Lundbeck Foundation (grant nos. R102-A9118 and R155-2014-1724) and the universities, and university hospitals of Aarhus and Copenhagen. ADB and NRM's work is also supported by the EU's Horizon 2020 programme (grant no. 667302, CoCA). Data handling and analysis was supported by NIMH (1U01MH109514-01 to Michael O'Donovan and ADB). CSM is a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199). K-PL and his team are supported by the Deutsche Forschungsgemeinschaft (DFG: CRU 125, CRC TRR 58 A1/A5, no. 44541416), ERA-Net NEURON/RESPOND, no. 01EW1602B, ERA-Net NEURON/DECODE, no. 01EW1902, and 5-100 Russian Academic Excellence Project. JH thanks Stiftelsen K.G. Jebsen, University of Bergen, the Western Norwegian Health Authorities (Helse Vest). HL thanks the Swedish research council. BC received financial support from the Spanish "Ministerio de Economía y Competitividad" (SAF2015-68341-R) and AGAUR (2017SGR738). MSA is a recipient of a contract from the Biomedical Network Research Center on Mental Health (CIBERSAM), Madrid, Spain. MR is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023). The research leading to these results has received funding from the Instituto de Salud Carlos III (PI16/01505, PI17/00289, PI18/01788, PI19/00721 and P19/01224), and cofinanced by the European Regional Development Fund (ERDF), from the Pla estratègic de recerca i innovació en salut (PERIS); Generalitat de Catalunya (METAL-Cat; SLT006/17/287), from "la Marató de TV3" (092330/31) and from the Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2017SGR1461). ES's work is supported by a personal Hypatia grant from the Radboud University Medical Center. MH received a Veni grant from of the Netherlands Organization for Scientific Research (NWO, grant number 91619115). The NeuroIMAGE study was supported by NIH Grant R01MH62873 (to SVF), NWO Large Investment Grant 1750102007010 (to JKB), ZonMW grant 60-60600-97-193, NWO grants 056-13-015 and 433-09-242, and matching grants from Radboud University Nijmegen Medical Center, University Medical Center Groningen and Accare, and Vrije Universiteit Amsterdam. Organization for Scientific Research (NWO; grant 016-130-669). BF and MK's work is supported by the Dutch National Science Agenda (NWA) for the NeurolabNL project (grant 400-17-602). This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and

King's College London. BF received additional funding from a personal Vici grant of the Dutch. The authors declare no competing interests

ACKNOWLEDGEMENTS

All the authors thank the patients and their families. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark, (grant to ADB) and Center for Genomics and Personalized Medicine, Aarhus, Denmark. This study is part of the International Multicentre persistent ADHD Collaboration (IMpACT; www.impactadhdgenomics.com). IMpACT unites major research centres working on the genetics of ADHD persistence across the lifespan and has participants in The Netherlands, Germany, Spain, Norway, the United Kingdom, the United States of America, Brazil, and Sweden. Principal investigators of IMpACT are: BF (chair), AR (co-chair), SVF, JH, BC, JAR-Q, PA, K-PL, JK, CHDB, JKB, SEJ, HL, AED, and EHG. This study uses data from the ADHD Working Group of the PGC, many of which contributed to this study and are named under the consortium author for this working group; additional members are at the time of submission of this manuscript: Anke R. Hammerschlag, Alexandra Philipsen, Alexandre Todorov, Alice Charach, Allison Ashley-Koch, Amaia Hervás, Ana Miranda, André Scherag, Anita Thapar, Anna Rommel, Anne Wheeler, Christine Cornforth, Danielle Posthuma, Elizabeth Corfield, Felecia Cerrato, Fernando Mulas, Franziska Degenhardt, Gláucia Chiyoko Akutagava Martins, Gun Peggy Strømstad Knudsen, Hans-Christoph Steinhausen, Herbert Roeyers, Hyo-Won Kim, Joanna Martin, Joel Gelernter, Joseph Sergeant, Juanita Gamble, Julia Pinsonneault, Jurgen Deckert, Kate Langley, Li Yang, Lindsey Kent, Manuel Mattheisen, Maria Jesús Arranz Calderón, Martin Steen Tesli, Meg Mariano, Michael Gill, Michael O'Donovan, Monica Bayes, Nick Martin, Niels Peter Ole Mors, Nigel Williams, Pak Sham, Patrick Sullivan, Patrick WL Leung, Paul Arnold, Paul Lichtenstein, Peter Holmans, Preben Bo Mortensen, Rachel Guerra, Raymond Walters, Richard Anney, Richard Ebstein, Ridha Joober, Sarah Anthony, Sarah Medland, Sarojini Sengupta, Søren Dinesen, Steve Nelson, Susan Smalley, Susann Scherag, Tammy Biondi, Tim Silk, Tinca Polderman, Tony Altar, Yanli Zhang-James, Yufeng Wang.

AUTHOR CONTRIBUTIONS

Conception or design of the work: PR, DD, CS-M, TZ, CPJ, OR, ES, AA-V, PA, BC, SVF, JH, SEJ, K-PL, MC, ADB, BF, JAR-Q, MSA, and MR. Acquisition, analysis or interpretation of the data: PR, DD, CS-M, TZ, MK, NRM, HW, IG-M, MP, LV, LA, VR, MC, CF, RB, GE, PA, AED, OG, MH, CPJ, SK-S, OR, ES, EJSS-B, PA, CHDB, JKB, BC, JH, SEJ, HL, K-PL, AR, LAR, MC, ADB, BF, JAR-Q, MSA, MR, Ole Andreas Andreassen, Tobias Banaschewski, Mark Bellgrove, Joseph Biederman, Christie Burton, Jennifer Crosbie, Søren Dalsgaard, Josephine Elia, Hakon Hakonarson, Catharina A. Hartman, Ziarih Hawi, Johannes Hebebrand, Anke Hinney, Sandra Loo, James McGough, Benjamin Neale, Robert Oades, Ted Reichborn-Kjennerud, Aribert Rothenberger, Russell Schachar, Irwin Waldman, Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Xin Wang, and Catherine H. Wilson. Drafted the work or substantively revised it: PR, DD, CS-M, PA, EHG, AH, MH, PMK, AJL, OR, DLR, AS-O, BSdS, Ted Reichborn-Kjennerud, ES, TS, EJSS-B, Mark Bellgrove, Ziarih Hawi, PA, JK, CHDB, BC, Johannes Hebebrand, SVF, JKB, HL, AR, LAR, MC, ADB, BF, JAR-Q, MSA, and MR. Final approval of the version to be published: all authors. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: PR, MSA, and MR.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41386-020-0664-5).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Franke B, Michelini G, Asherson P, Banaschewski T, Bilbow A, Buitelaar JK, et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. Eur Neuropsychopharmacol. 2018;28:1059–88.
2. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015;56:345–65.
3. Brikell I, Kuja-Halkola R, Larsson H. Heritability of attention-deficit hyperactivity disorder in adults. Am J Med Genet B Neuropsychiatr Genet. 2015;168:406–13.
4. Kan KJ, Dolan CV, Nivard MG, Middeldorp CM, van Beijsterveldt CE, Willemsen G, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry. 2013;52:12–25.
5. Kuntsi J, Rijsdijk F, Ronald A, Asherson P, Plomin R. Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. Biol Psychiatry. 2005;57:647–54.
6. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.
7. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. Mol Psychiatry. 2019;24:562–75.
8. Ganna A, Satterstrom FK, Zekavat SM, Das I, Kurki MI, Churchhouse C, et al. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. Am J Hum Genet. 2018;102:1204–11.
9. Martin J, O'Donovan MC, Thapar A, Langley K, Williams N. The relative contribution of common and rare genetic variants to ADHD. Transl Psychiatry. 2015;5:e506.
10. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. ASD and ADHD have a similar burden of rare protein-truncating variants. https://doi.org/10.1101/277707 (2018).
11. Hawi Z, Cummins TD, Tong J, Arcos-Burgos M, Zhao Q, Matthews N, et al. Rare DNA variants in the brain-derived neurotrophic factor gene increase risk for attention-deficit hyperactivity disorder: a next-generation sequencing study. Mol Psychiatry. 2017;22:580–4.
12. Brainstorm C, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360:eaap8757.
13. Shaw P, Stringaris A, Nigg J, Leibenluft E. Emotion dysregulation in attention deficit hyperactivity disorder. Am J Psychiatry. 2014;171:276–93.
14. Wilens TE, Biederman J, Faraone SV, Martelon M, Westerberg D, Spencer TJ. Presenting ADHD symptoms, subtypes, and comorbid disorders in clinically referred adults with ADHD. J Clin Psychiatry. 2009;70:1557–62.
15. Anker E, Bendiksen B, Heir T. Comorbid psychiatric disorders in a clinical sample of adults with ADHD, and associations with education, work and social characteristics: a cross-sectional study. BMJ Open. 2018;8:e019700.
16. Biederman J, Faraone SV, Spencer T, Wilens T, Norman D, Lapey KA, et al. Patterns of psychiatric comorbidity, cognition, and psychosocial functioning in adults with attention deficit hyperactivity disorder. Am J Psychiatry. 1993;150:1792–8.
17. Kessler RC, Adler LA, Barkley R, Biederman J, Conners CK, Faraone SV, et al. Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: results from the national comorbidity survey replication. Biol Psychiatry. 2005;57:1442–51.
18. Larson K, Russ SA, Kahn RS, Halfon N. Patterns of comorbidity, functioning, and service use for US children with ADHD, 2007. Pediatrics. 2011;127:462–70.
19. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet. 2015;385:2190–6.
20. Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Prim. 2015;1:15020.
21. Kupper T, Haavik J, Drexler H, Ramos-Quiroga JA, Wermelskirchen D, Prutz C, et al. The negative impact of attention-deficit/hyperactivity disorder on occupational health in adults and adolescents. Int Arch Occup Environ Health. 2012;85:837–47.
22. Voigt RG, Katusic SK, Colligan RC, Killian JM, Weaver AL, Barbaresi WJ. Academic achievement in adults with a history of childhood attention-deficit/hyperactivity disorder: a population-based prospective study. J Dev Behav Pediatr. 2017;38:1–11.
23. Strine TW, Lesesne CA, Okoro CA, McGuire LC, Chapman DP, Balluz LS, et al. Emotional and behavioral difficulties and impairments in everyday functioning among children with a history of attention-deficit/hyperactivity disorder. Prev Chronic Dis. 2006;3:A52.
24. Biederman J, Petty CR, Evans M, Small J, Faraone SV. How persistent is ADHD? A controlled 10-year follow-up study of boys with ADHD. Psychiatry Res. 2010;177:299–304.
25. Caye A, Spadini AV, Karam RG, Grevet EH, Rovaris DL, Bau CH, et al. Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis. Eur Child Adolesc Psychiatry. 2016;25:1151–9.
26. Cheung CH, Rijdijk F, McLoughlin G, Faraone SV, Asherson P, Kuntsi J. Childhood predictors of adolescent and young adult outcome in ADHD. J Psychiatr Res. 2015;62:92–100.

27. Riglin L, Collishaw S, Thapar AK, Dalsgaard S, Langley K, Smith GD, et al. Association of genetic risk variants with attention-deficit/hyperactivity disorder trajectories in the general population. JAMA Psychiatry. 2016;73:1285–92.
28. Roy A, Hechtman L, Arnold LE, Swanson JM, Molina BSG, Sibley MH, et al. Childhood predictors of adult functional outcomes in the multimodal treatment study of attention-deficit/hyperactivity disorder (MTA). J Am Acad Child Adolesc Psychiatry. 2017;56:687–95. e7.
29. Pingault JB, Viding E, Galera C, Greven CU, Zheng Y, Plomin R, et al. Genetic and environmental influences on the developmental course of attention-deficit/hyperactivity disorder symptoms from childhood to adolescence. JAMA Psychiatry. 2015;72:651–8.
30. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. J Child Psychol Psychiatry. 2004;45:577–88.
31. Akutagava-Martins GC, Salatino-Oliveira A, Kieling C, Genro JP, Polanczyk GV, Anselmi L, et al. COMT and DAT1 genes are associated with hyperactivity and inattention traits in the 1993 Pelotas Birth Cohort: evidence of sex-specific combined effect. J Psychiatry Neurosci. 2016;41:405–12.
32. Bonvicini C, Faraone SV, Scassellati C. Attention-deficit hyperactivity disorder in adults: a systematic review and meta-analysis of genetic, pharmacogenetic and biochemical studies. Mol Psychiatry. 2016;21:1643.
33. Bonvicini C, Faraone SV, Scassellati C. Common and specific genes and peripheral biomarkers in children and adults with attention-deficit/hyperactivity disorder. World J Biol Psychiatry. 2018;19:80–100.
34. Johansson S, Halleland H, Halmoy A, Jacobsen KK, Landaas ET, Dramsdahl M, et al. Genetic analyses of dopamine related genes in adult ADHD patients suggest an association with the DRD5-microsatellite repeat, but not with DRD4 or SLC6A3 VNTRs. Am J Med Genet B Neuropsychiatr Genet. 2008;147B:1470–5.
35. Lesch KP, Timmesfeld N, Renner TJ, Halperin R, Roser C, Nguyen TT, et al. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. J Neural Transm. 2008;115:1573–85.
36. Ribases M, Ramos-Quiroga JA, Hervas A, Bosch R, Bielsa A, Gastaminza X, et al. Exploration of 19 serotoninergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. Mol Psychiatry. 2009;14:71–85.
37. Sanchez-Mora C, Ramos-Quiroga JA, Bosch R, Corrales M, Garcia-Martinez I, Nogueira M, et al. Case-control genome-wide association study of persistent attention-deficit hyperactivity disorder identifies FBXO33 as a novel susceptibility gene for the disorder. Neuropsychopharmacology. 2015;40:915–26.
38. Thissen AJ, Bralten J, Rommelse NN, Arias-Vasquez A, Greven CU, Heslenfeld D, et al. The role of age in association analyses of ADHD and related neurocognitive functioning: A proof of concept for dopaminergic and serotonergic genes. Am J Med Genet B Neuropsychiatr Genet. 2015;168:471–9.
39. Zayats T, Athanasiu L, Sonderby I, Djurovic S, Westlye LT, Tamnes CK, et al. Genome-wide analysis of attention deficit hyperactivity disorder in Norway. PLoS One. 2015;10:e0122501.
40. Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010;49:884–97.
41. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5:e1000529.
42. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009;10:387–406.
43. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34:816–34.
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
45. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. Bioinformatics. 2010;26:2190–1.
46. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9:1192–212.
47. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7.
48. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
49. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–35.
50. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
51. Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet. 2016;48:803–10.
52. Middeldorp CM, Hammerschlag AR, Ouwens KG, Groen-Blokhuis MM, Pourcain BS, Greven CU, et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. J Am Acad Child Adolesc Psychiatry. 2016;55:896–905. e6.
53. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33:272–9.
54. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238–43.
55. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun. 2018;9:2282.
56. Chen Q, Brikell I, Lichtenstein P, Serlachius E, Kuja-Halkola R, Sandin S, et al. Familial aggregation of attention-deficit/hyperactivity disorder. J Child Psychol Psychiatry. 2017;58:231–9.
57. Biederman J, Faraone SV, Mick E, Spencer T, Wilens T, Kiely K, et al. High risk for attention deficit hyperactivity disorder among children of parents with childhood onset of the disorder: a pilot study. Am J Psychiatry. 1995;152:431–5.
58. Chapman NH, Nato AQ Jr., Bernier R, Ankenman K, Sohi H, Munson J, et al. Whole exome sequencing in extended families with autism spectrum disorder implicates four candidate genes. Hum Genet. 2015;134:1055–68.
59. Shimizu T, Hibi M. Formation and patterning of the forebrain and olfactory system by zinc-finger genes Fezf1 and Fezf2. Dev Growth Differ. 2009;51:221–31.
60. Koukoulas I, Augustine C, Silkenbeumer N, Gunnersen JM, Scott HS, Tan SS. Genomic organisation and nervous system expression of radial spoke protein 3. Gene. 2004;336:15–23.
61. Bonora E, Graziano C, Minopoli F, Bacchelli E, Magini P, Diquigiovanni C, et al. Maternally inherited genetic variants of CADPS2 are present in autism spectrum disorders and intellectual disability patients. EMBO Mol Med. 2014;6:795–809.
62. Hattori K, Tanaka H, Wakabayashi C, Yamamoto N, Uchiyama H, Teraishi T, et al. Expression of Ca(2)(+)-dependent activator protein for secretion 2 is increased in the brains of schizophrenic patients. Prog Neuropsychopharmacol Biol Psychiatry. 2011;35:1738–43.
63. Sadakata T, Shinoda Y, Ishizaki Y, Furuichi T. Analysis of gene expression in Ca(2+)-dependent activator protein for secretion 2 (Cadps2) knockout cerebellum using GeneChip and KEGG pathways. Neurosci Lett. 2017;639:88–93.
64. Velez JI, Chandrasekharappa SC, Henao E, Martinez AF, Harper U, Jones M, et al. Pooling/bootstrap-based GWAS (pbGWAS) identifies new loci modifying the age of onset in PSEN1 p.Glu280Ala Alzheimer's disease. Mol Psychiatry. 2013;18:568–75.
65. Fyfe JC, Madsen M, Hojrup P, Christensen EI, Tanner SM, de la Chapelle A, et al. The functional cobalamin (vitamin B12)-intrinsic factor receptor is a novel complex of cubilin and amnionless. Blood. 2004;103:1573–9.
66. Tanner SM, Aminoff M, Wright FA, Liyanarachchi S, Kuronen M, Saarinen A, et al. Amnionless, essential for mouse gastrulation, is mutated in recessive hereditary megaloblastic anemia. Nat Genet. 2003;33:426–9.
67. Venkatramanan S, Armata IE, Strupp BJ, Finkelstein JL. Vitamin B-12 and cognition in children. Adv Nutr. 2016;7:879–88.
68. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50:912–9.

European Neuropsychopharmacology (2019) 29, 795–802

www.elsevier.com/locate/euroneuro

SHORT COMMUNICATION

Genome-wide analysis of emotional lability in adult attention deficit hyperactivity disorder (ADHD)

Laura Gisbert[a,b,c,1], Laura Vilar[a,d,1], Paula Rovira[a,d], Cristina Sánchez-Mora[a,b,d], Mireia Pagerols[a,d], Iris Garcia-Martínez[a,d,e], Vanesa Richarte[a,b,c], Montserrat Corrales[a,b,c], Miguel Casas[a,b,c,d], Josep Antoni Ramos-Quiroga[a,b,c,d], María Soler Artigas[a,b,d,*], Marta Ribasés[a,b,d,*]

[a] *Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain*
[b] *Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain*
[c] *Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain*
[d] *Psychiatric Genetics Unit, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain*
[e] *Grup de Coagulopaties Congènites, Banc de Sang i Teixits, Barcelona, Catalonia, Spain*

Received 31 July 2018; received in revised form 11 March 2019; accepted 23 April 2019

KEYWORDS
ADHD;
Adult;
Genetics;
GWAS;
Emotional lability

Abstract
Emotional lability is strongly associated with Attention Deficit Hyperactivity Disorder (ADHD), represents a major source of impairment and predicts poor clinical outcome in ADHD. Given that no specific genes with a role in the co-occurrence of both conditions have been described, we conducted a GWAS of emotional lability in 563 adults with ADHD. Despite not reaching genome-wide significance, the results highlighted genes related with neurotransmission, cognitive function and a wide range of psychiatric disorders that have emotional lability as common clinical feature. By constructing polygenic risk scores on mood instability in the UK Biobank sample and assessing their association with emotional lability in our clinical dataset, we found

* Corresponding authors at: Psychiatric Genetics Unit, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain.
E-mail addresses: maria.soler@vhir.org (M. Soler Artigas), marta.ribases@vhir.org (M. Ribasés).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.euroneuro.2019.04.004

suggestive evidence of common genetic variation contributing to emotional lability in general population and in clinically diagnosed ADHD. Although not conclusive, these tentative results are in agreement with previous studies that suggest emotion dysregulation as a transdiagnostic construct and highlight the need for further investigation to disentangle the genetic basis of mood instability in ADHD and co-occurring psychiatric disorders.

1. Introduction

Emotional lability, also known as mood instability, emotional impulsivity, emotional dysregulation, emotional impulsiveness, affective lability, mood lability and deficient emotional self-regulation, is a common clinical feature of a range of psychiatric disorders including major depressive disorder, bipolar disorder, borderline personality disorder or Attention Deficit Hyperactivity Disorder (ADHD) (Childress and Sallee, 2015). Particularly, ADHD diagnostic criteria currently recognizes symptoms of emotional lability as an associated feature of ADHD (Merwood et al., 2014), although the extent of the phenotypic and etiologic associations between emotional lability and symptoms of hyperactivity-impulsivity or inattention remains unclear.

The prevalence of emotional lability symptoms is about 25–45% in children and between 30% and 70% in adults with ADHD (Childress and Sallee, 2015). Moreover, emotional lability has been highlighted as a contributor to the functional impairment in youth and adults with ADHD, it may increase the severity of ADHD symptomatology as well as comorbid disorders and is associated with ADHD persistence and lower quality of life (Shaw et al., 2014).

The nature of the relationship between emotional lability and ADHD is still unclear. Converging evidence in twin and family studies support significant genetic overlap between emotional dysregulation and ADHD symptoms and higher risk of emotional lability in family members of ADHD subjects (Merwood et al., 2014; Riglin et al., 2017). These results are in line with the association between ADHD polygenic risk scores and early-life irritability found in a population-based cohort and in an ADHD clinical sample (Riglin et al., 2017). Pharmacological studies also document a concomitant decline in symptoms of hyperactivity-impulsivity, inattention, and emotional lability in response to methylphenidate and atomoxetine in adults (Marchant et al., 2011). All this evidence suggests that emotional lability and ADHD may arise as a result of a common etiology, point to emotional lability as etiologically relevant to the core ADHD phenotype and support that it may be targeted in clinical intervention (Merwood et al., 2014).

Although ADHD has been the focus of considerable genetic research, to date there is little work focused on the genetic underpinnings of emotional lability or on the genetic basis of the link between them. Both conditions have a complex genetic architecture, with heritability estimates of 74% (Faraone and Larsson, 2018) and 25% (Coccaro et al., 2012) for ADHD and emotional lability, respectively, but the role of specific genes remains still unclear. Although each of the associated variants appears to account for a relatively small proportion of the variance in both traits, SNPs were estimated to account for 10—28% of the heritability of ADHD (Demontis et al., 2019) and 8% of the heritability of mood instability (Ward et al., 2017).

Genetic research on ADHD or emotional lability has mainly focused on common variants through candidate gene or genome-wide association studies. A very recent GWAS meta-analysis in 20,183 ADHD cases and 35,191 controls reported 12 genome-wide significant loci including genes involved in neurodevelopmental processes and evolutionarily conserved genomic regions. Two GWAS on emotional lability have been run so far. The first one identified a genome-wide significant association between the interleukin receptor 2A gene, *IL2RA*, and emotion dysregulation in males, as well as enrichment for genes involved in different psychiatric disorders and in the calcium signaling pathway (Powers et al., 2016). Furthemore, Ward et al. (2017) conducted a GWAS on mood instability in 53,525 cases and 60,443 controls from the UK biobank which revealed four genome-wide significant loci and genetic correlation between mood instability and different psychiatric disorders.

Given that emotional lability is strongly associated with ADHD but no specific genes with a role in the co-occurrence of both conditions have been described, we conducted for the first time a GWAS of emotional lability in adults with ADHD to identify genes and biological pathways underlying this trait that represents a major source of impairment and predicts poor clinical outcome in ADHD.

2. Experimental procedures

2.1. Stage 1: GWAS of emotional lability in ADHD

2.1.1. Participants

The clinical sample comprised 563 adults of European ancestry (67% males; mean age = 33 years; SD = 10.5), who met ADHD diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Exclusion criteria included mental retardation, schizophrenia or other psychotic disorders, symptoms of substance intoxication and withdrawal and neurological or systemic disorders that might explain ADHD symptoms. All subjects were evaluated at Hospital Universitari Vall d'Hebron of Barcelona (Spain) and diagnosis was blind to genotype. The study was approved by the Clinical Research Ethics Committee of our Institution, all methods were performed in accordance with the relevant guidelines and regulations, and written informed consent was obtained from all subjects before inclusion in the study.

2.1.2. Clinical assessment

The evaluation of the ADHD diagnosis was carried out with the Spanish version of the Conners' Adult ADHD Diagnostic Interview for DSM-IV (CAADID parts I and II). Emotional lability was evaluated using the following items from the self-reported Conners' Adult ADHD Rating Scale-long version (CAARS-S:L): "I am irritable", "I have unpredictable moods", "Many things set me off easily", "I have a hot

temper/I lose patience easily", "I still throw tantrums" and "I get frustrated easily". Each item is scored on a four-point Likert scale ranging from 0 to 3 (0= *not at all or never*; 1 = *just a little, once in a while*; 2 = *pretty much, often*; 3 = *very much, very frequently*). Mean score in the CAARS' emotional lability subscale was 8.99 ranging from 0 to 18.

2.1.3. Genome-wide association study

Genomic DNA was isolated from peripheral blood leukocytes by the salting-out procedure. Subjects were genotyped in three different waves using the Illumina HumanOmni1-Quad ($n = 355$), HumanOmni 2.5 ($n = 166$) and the PsychChip ($n = 42$) arrays. Pre-imputation quality control and principal components analysis were implemented with the Ricopili pipeline (https://sites.google.com/a/broadinstitute.org/ricopili/), and ancestry outliers were excluded. Genotype imputation was performed using the European population haplotypes of the 1000 Genomes Project Phase I as the reference panel for waves 1 and 2 and the 1000 Genomes Project Phase III for wave 3 (The 1000 Genomes Project Consortium, 2015). Individuals with >2% genotype missingness were removed, as well as SNPs with low call rate (<0.99), with minor allele frequency (MAF) <0.01, INFO score bellow 0.8 or failing the Hardy–Weinberg equilibrium test ($P<1e{-}06$). Post-imputation best-guess genotype data from a total of 2,777,520 markers available in all three datasets were tested for association with emotional lability through proportional odds logistic regression using the function "polr", from the "MASS" R package with the ologit-gwas script (https://github.com/edm1/ologit-gwas). Age, sex, genotyping waves, ADHD subtype, comorbid psychiatric disorders and the first five principal components were included as covariates. The quantile–quantile and manhattan plots were drawn using the qqman R package (http://cran.r-project.org/web/packages/qqman). Index SNPs were defined based on clumping of variants using the PLINK software with default settings ($p_1 = 0.0001$, $p_2 = 0.01$, $r^2 = 0.5$, kb = 250) (https://www.cog-genomics.org/plink2/). Annotation was performed in accordance with the Human hg19 genome build considering genes within a ±10 kb distance from index SNPs. Locus Zoom interactive web-based visualization tool (http://locuszoom.org/) was used to generate regional plots of the top index SNP with a ±2 Mb flanking distance.

2.1.4. Gene-based and gene-set analyses

The gene-based and gene-set association analyses were conducted using MAGMA (De Leeuw et al., 2015). Gene regions were defined as ±10 kb for each gene according to the UCSC Genome Browser GRCh37/hg19 release (https://genome.ucsc.edu/) and used the 1000 Genomes Project Phase I dataset as reference panel to estimate patterns of LD for each locus (The 1000 Genomes Project Consortium, 2015). For the gene-set analysis the Gene Ontology (GO) and canonical pathways downloaded from MSigDB (http://www.broadinstitute.org/gsea/msigdb) were considered. Correction for multiple testing was applied using false discovery rate (FDR) with a threshold of 5% and 10,000 permutations in the gene-based and gene-set analyses, respectively.

2.2. Stage 2: polygenic risk score analysis based on UK Biobank mood instability GWAS

We generated Polygenic Risk Scores (PRSs) based on the results of the GWAS on mood instability, excluding individuals with psychiatric disorders, run in the UK Biobank sample (Ward et al., 2017) using the Polygenic Risk Score software (PRSice). Quantitative CAARS-S:L scores were dichotomised using a threshold of 12. A logistic regression model was applied to test whether PRS at multiple *P*-value thresholds predicted emotional lability in our ADHD cohort ('target population'). Age, sex, genotyping waves, ADHD subtype, comorbid psychiatric disorders and the first five principal components were included as covariates and 10,000 permutations were computed at the best-fit *P*-value threshold to correct for multiple testing.

3. Results

In stage 1 of the study, and after individual and SNP standard quality control filtering, we conducted a GWAS of emotional lability considering 2,777,520 SNPs in a sample of 563 adults with ADHD. The quantile–quantile plot showed no departure from the null distribution of expected *P*-values, with a genomic inflation factor of $\lambda = 1.08$ (Supplementary Fig. 1).

None of the association signals at SNP or gene level exceeded the genome-wide threshold for significance, with the top hit at rs2165472 located 1.1 Mb upstream from the *LPHN3* gene on chromosome 4 ($P = 3.77e{-}06$; $B = 1.31$; SE = 0.28) (Fig. 1, Table 1). The gene-based association test showed 1016 genes associated with emotional lability ($P < 0.05$), with the top hit in *OR9A4* on chromosome 7 ($P = 1.72e{-}05$) (Table 1). No gene-set was found significant after multiple comparison correction, with a total of 262 GO terms nominally enriched in our gene set ($P<0.05$) and "*Intraciliary Transport Particle*" being the most significant one (GO:0030990, $P = 1.74e{-}04$) (Supplementary Table S1). Moreover, 64 canonical pathways were overrepresented in our gene set, with "*Terpenoid Backbone Biosyntesis*" ($P = 1.19e{-}03$) and "*p75NTR recruits signaling complexes*" ($P = 1.53e{-}03$) among the top signals (Supplementary Table S2).

In stage 2, we constructed PRSs based on mood instability data from the UK Biobank sample (Ward et al., 2017) and assessed their association with emotional lability in our ADHD clinical cohort to test whether emotional lability in a clinical sample of ADHD subjects and in the general population shares common genetic load. We found suggestive evidence of association between PRSs for emotional lability in the general population and emotional lability in clinical diagnosed ADHD, being the most predictive *P*-value threshold set at $P_T = 5e{-}05$ (corrected *P*-value = 0.078; Fig. 2), which explained 0.59% of the variation in emotional lability.

4. Discussion

To our knowledge, this is the first study that investigates the genetic basis of emotional lability in adults with ADHD through a GWAS perspective. Despite not reaching genome-wide significance, our findings show tentative evidence for the involvement of genes relevant in the context of emotional lability, including, cell-substrate adhesion, neurotransmission signaling, neurological diseases and psychiatric disorders.

Emotional lability is a highly prevalent clinical feature in ADHD patients across the lifespan. Although it is not part of the current definition criteria for ADHD diagnosis, emotional dysregulation is present in a subset of patients and represents a major source of functional impairment and poor clinical outcome. About 40% of children and from 35% to 70% of adults with ADHD exhibit emotional dysregulation, with low frustration tolerance, quick anger and explosive

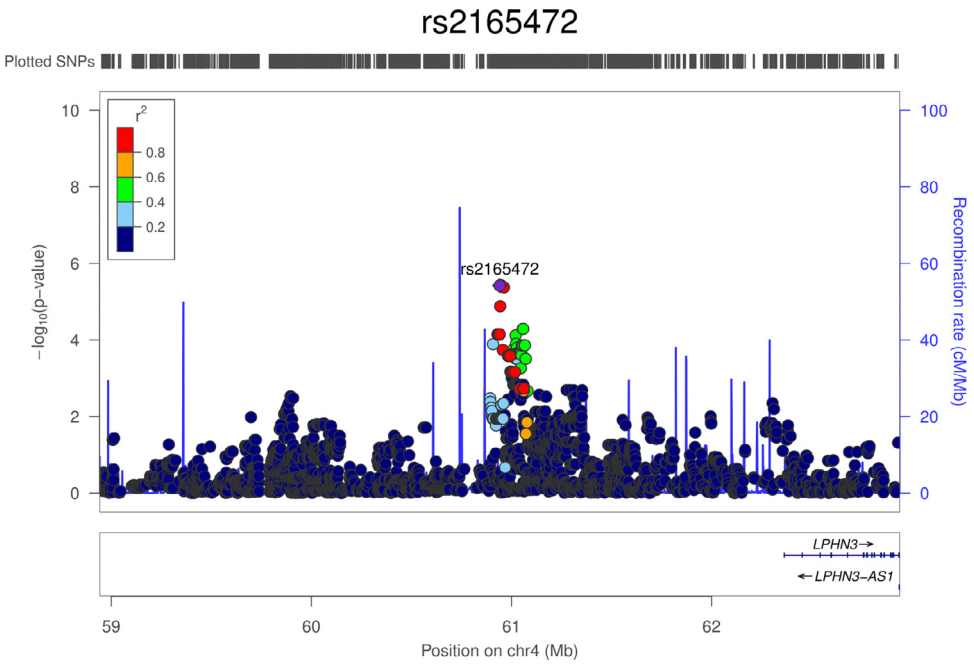


Fig. 1 Regional association plot of top index SNP identified in genome-wide association study and nearby genes.

behavior, regardless of other comorbidities (Shaw et al., 2014). Although the results of the present study suggest potential and interesting genes for emotional lability in ADHD subjects, whether ADHD with emotional dysregulation represents a distinct genetic group or both share a common genetic load remains unknown. The largest GWAS on mood instability performed so far using the UK Biobank sample showed no genetic correlation between both symptom domains that commonly co-exist (Ward et al., 2017), but findings based on twin and familiar co-segregation studies as well as a PRS analysis on early irritability support moderate genetic overlap between ADHD and emotional dysregulation (Riglin et al., 2017; Shaw et al., 2014). Although no statistically significant, the tentative evidence of association between the PRS based on mood instability from the UK Biobank sample and emotional lability in our ADHD clinical cohort, also suggest common genetic background underlying emotional lability in general population and in clinical diagnosed ADHD. These suggestive results emphasize the need for further studies in additional samples to confirm these findings and to understand the genetic underpinnings of mood instability and its link with ADHD.

The top hit from our GWAS, rs2165472, is located 1.1 Mb upstream from the *LPNH3*, which encodes a neuronal adhesion-GPC receptor from the LPHN family that is almost exclusively expressed in brain (Acosta et al., 2016). It plays a role in the development of glutamatergic synapses (O'Sullivan et al., 2014) and has been extensively associated with ADHD, and its severity, long-term outcome, response to treatment and comorbid conditions such as disruptive behaviors and substance use disorder (SUD) (Acosta et al., 2016; Arcos-Burgos et al., 2019). Among the top hits, we also identified SNPs located within, or nearby, other genes of interest for emotional lability including *FOXK1*, expressed in key brain areas for cognitive function (Wijchers et al., 2006); *GABRG3*, which encodes a gamma-aminobutyric acid receptor subunit, or *GRM5*, a glutamate receptor. They are highly involved in neurotransmission and normal brain function and have been widely associated with a variety of psychiatric disorders including anxiety, bipolar mood disorder, SUD, autism or major depressive disorder (Fatemi and Folsom, 2015).

The gene-based and gene-set competitive analyses also highlighted genes and pathways potentially relevant for emotional lability, including genes such as *OR9A4*, previously associated with anorexia nervosa (Wade et al., 2013), *CTBP1*, which is involved in the regulation of gene expression during development and exhibited aberrant blood expression in schizophrenia and bipolar disorder subjects (Tsuang et al., 2005), or *ASS1* which was downregulated in urine samples of subjects with major depressive disorder (Wu et al., 2015). The present study points to a wide range of pathways and cellular processes involved in several psychiatric disorders and neuronal functions. Of particular interest are the "*p75NTR recruits signaling complexes*", involved in survival and formation of neurons (Dechant and Barde, 2002), the "*calcineurin pathway*", essential for synaptic plasticity processes (Xia and Storm, 2005) or "*p38MAPK events*", associated with neuronal death, development and differentiation (Ibrahim et al., 2017).

The results of this study should be viewed in light of several limitations:

- First, our modest sample size is not powered enough to identify genome-wide significant hits and has probably prevented us from detecting variants with modest effects. Despite using a proportional odds logistic regression model to make the most of the data we had, given their ordinal nature, we cannot report any conclusive findings.
- Second, it remains unknown whether the nature of the relationship between ADHD and emotional instability is

Table 1 Top 15 hits from the (a) SNP and (b) gene-based analyses of emotional lability in adult Attention Deficit Hyperactivity Disorder.

(a)

SNP	CHR	Gene (kb distance)	Effect Allel	OR	CI 95%	*P*-value
rs2165472	4	*LPHN3* (−1126 kb)	C	3.71	2.13-6.48	3.77E−06
rs35872837	5	*ISL1* (+300.1 kb)	G	0.49	0.36-0.66	3.81E−06
rs723840	14	*CMTM* (0 kb)	C	1.63	1.33-2.00	3.81E−06
rs2109112	12	*PARP11* (+4.859 kb)	T	1.70	1.36-2.13	3.96E−06
rs117358046	14	*IFT43* (0 kb)	T	4.65	2.42-8.94	3.99E−06
rs3087749	7	*FOXK1* (0 kb)	T	0.61	0.50-0.76	7.58E−06
rs13236432	7	*PRSS37* (+9.559 kb)	C	1.77	1.38-2.27	8.43E−06
rs113365723	5	*ERAP1* (0 kb)	A	5.61	2.62-12.00	8.88E−06
rs1515594	3	*NLGN1* (+278.3 kb)	G	0.55	0.42-0.71	9.69E−06
rs9841241	3	*RYBP* (−20.28 kb)	G	1.63	1.31-2.03	1.13E−05
rs4778109	15	*GABRG3* (0 kb)	A	0.62	0.50-0.77	1.24E−05
rs566277	11	*GRM5* (0 kb)	G	0.26	0.14-0.48	1.36E−05
rs9311047	3	*PDCD6IP* (+335.3 kb)	A	2.03	1.47-2.79	1.50E−05
rs11694790	2	*PNPT1* (+76.83 kb)	T	1.61	1.30-2.00	1.53E−05
rs74870851	2	*ABCB11* (−3.225 kb)	G	0.16	0.07-0.37	1.58E−05

(b)

Gene	CHR	Start	Stop	*n* SNPs	*P*-value
OR9A4	7	141,608,676	141,629,620	29	1.72E−05
PFN1	17	4,838,945	4,862,381	17	6.93E−05
RWDD3	1	95,689,711	95,722,781	54	8.63E−05
BCAS3	17	58,745,172	59,480,199	285	1.00E−04
RNF167	17	4,833,328	4,858,517	21	1.35E−04
HGFAC	4	3,433,702	3,461,214	10	1.36E−04
PAPSS2	10	89,409,476	89,517,462	110	1.49E−04
CTBP1	4	1,195,228	1,252,908	90	1.60E−04
CORO2B	15	68,841,614	69,030,145	107	1.98E−04
HLF	17	53,332,321	53,412,426	36	2.37E−04
ZFP42	4	188,906,925	188,936,199	16	2.37E−04
ASS1	9	133,310,094	133,386,661	16	3.35E−04
GLTSCR1	19	48,101,453	48,216,534	131	3.67E−04
SLC25A11	17	4,830,425	4,853,462	22	3.89E−04
RBM11	21	15,578,466	15,610,693	123	4.13E−04

Note: SNP: Single Nucleotide Polymorphism CHR: Chromosome OR: Odds ratio CI: confidence intervals.
*OR: the odds per effect allele of an increase in the CAARS' emotional lability subscale.

mediated by other comorbid disorders, ADHD subtypes, gender, family history of ADHD or adverse environmental factors (Ward et al., 2017), and their role in the link between ADHD and emotional instability warrants further investigation.

- Third, given the cross-sectional nature of the study, we cannot infer causality or make assertions about the temporal relationship between ADHD and emotional lability. Therefore, prospective, longitudinal studies are required to examine the temporal onset of emotion dysregulation in ADHD subjects.
- Fourth, there are several definitions of emotional lability and different scales to measure the construct. Furthermore, there are certain limitations related to self-report measures of emotion lability. We applied the Conners' definition of emotional lability as irritability, unpredictable moods, setting off easily, hot temper, low frustration tolerance and difficulties in anger management. Although this subscale of the CAARS is a good measure for emotional lability in ADHD subjects and there is evidence supporting that adults with ADHD are reliable informants about symptomatology (Vidal et al., 2014), future research using specific scales of emotional reactivity as well as more thorough and objective measures of this construct is warranted.

In conclusion, to our knowledge, this is the first attempt to assess the genetic background of emotional instability in ADHD patients. Although not conclusive, we found suggestive evidence for genes involved in central nervous system development and function and in a wide range of psychiatric disorders that have emotional lability as common clinical feature. Our results are in line with previous studies supporting a common genetic background underlying emotional lability in the general population and in clinically diagnosed ADHD individuals, suggest emotion dysregulation as a trans-diagnostic construct (Sloan et al., 2017) and highlight the need for further investigation to disentangle the genetic basis of mood instability in ADHD and its role as a source of impairment and clinical outcome.

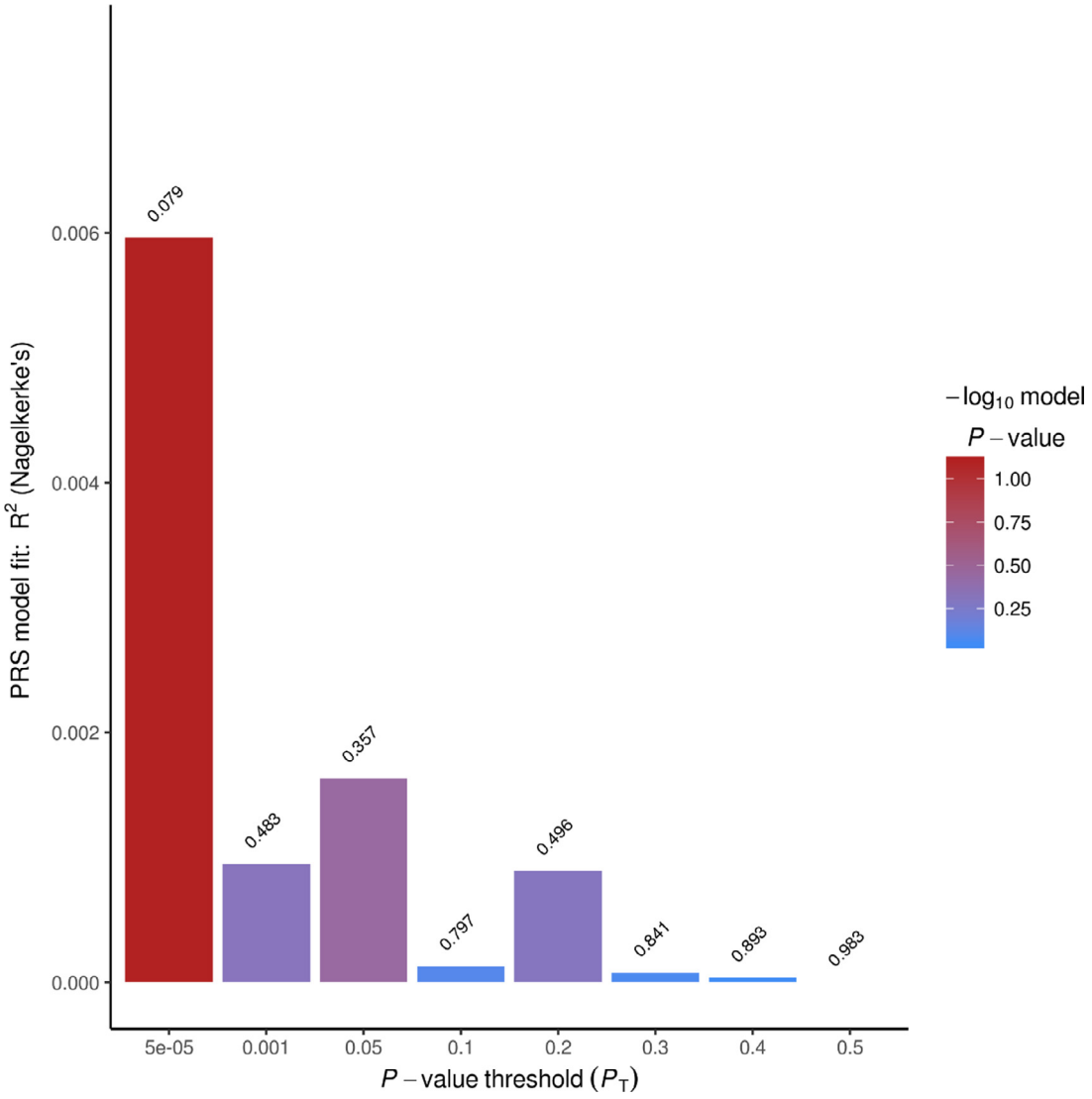


Fig. 2 Bar plot showing results from the PRS analysis based on mood instability data from UK Biobank at broad *P*-value thresholds ($P_T = 0.001$, $P_T = 0.05$, $P_T = 0.1$, $P_T = 0.2$, $P_T = 0.3$, $P_T = 0.4$, $P_T = 0.5$) and at the best-fit PRS ($P_T = 5e{-}05$).

Contributors

P.R., C.S.M., I.G., and M.P. participated in the DNA isolation and preparation of samples. L.G., L.V., P.R., C.S.M., I.G., M.P., M.S.A and M.R., undertook the statistical analyses. L.G., V. R., and M. C., contributed to the clinical assessment and recruitment of patients. M. C. and J. A. R. Q. participated in the study design, clinical assessment and coordination of the clinical research. M. R. conceived the project, wrote the protocol and coordinated the study design and the statistical analyses. J.A.R.Q., M.S.A and M. R. supervised the project and the manuscript preparation. All authors contributed to and have approved the final version.

Conflict of interest

The author L.G. received travel awards for taking part in psychiatric meetings from Shire in the last 3 years.

The author V.R. received travel awards for taking part in psychiatric meetings from Shire in the last 3 years.

The author M.C has received fees to give talks for Janssen-Cilag, Bristol-Mayers Squibb, Ferrer-Brainfarma, Pfizer, Reckitt-Benckiser, Lundbeck, Otsuka, Servier, Lilly, Shire, GSK, Rovi and Adamed. He has received financial compensation for his participation as a member of the Janssen-Cilag, Lilly, Shire, Lundbeck, Otsuka, Ferrer and Rovi board.

The author J.A.R.Q was on the speakers' bureau and/or acted as consultant for Eli-Lilly, Novartis, Shire, Lundbeck, Almirall, BGaze and Rubió in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Rubió, Shire, and Eli-Lilly. The ADHD Program chaired by him received unrestricted educational and research support from the following pharmaceutical companies in the last 3 years: Eli-Lilly, Lundbeck, Janssen-Cilag, Actelion, Shire, and Rubió.

All other authors declare that they have no conflicts of interest.

Role of funding source

This work was funded by the Instituto de Salud Carlos III (PI14/01700, PI15/01789, PI16/01505 and PI17/00289), and co-financed by the European Regional Development Fund (ERDF), Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya, Spain (2014SGR1357 and 2017SGR1461), the European College of Neuropsychopharmacology (ECNP network: 'ADHD across the lifespan'), the Departament de Salut, Generalitat de Catalunya, Spain and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. This project also received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement Nos. 602805 (Agressotype), 667302 (CoCA), 643051 (MiND) and 728018 (Eat2BENice).

L.V. is a recipient of a pre-doctoral fellowship from the Instituto de Salud Carlos III, Spain (FI18/00285). P. R. is a recipient of a pre-doctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI_B 00899). C. S. M. is a recipient of a Sara Borrell contract and a mobility grant from the Spanish Ministerio de Economía y Competitividad, Instituto de Salud Carlos III (CD15/00199 and MV16/00039). M. P. has been a recipient of a pre-doctoral fellowship from the Vall d'Hebron Research Institute (PRED-VHIR-2013) and a research grant from the Deutscher Akademischer Austauschdienst (DAAD), Germany (Research Grants – Short-Term Grants, 2017). I.G.M. has been a recipient of a contract from the 7th Framework Programme for Research, Technological Development and Demonstration, European Commission (AGGRESSOTYPE_FP7HEALTH2013/602805).

M. S. A. is a recipient of a contract from the Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. M. R. is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023).

Acknowledgments

We are grateful to patients from the Hospital Universitari Vall d'Hebron, who kindly participated in this research. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) hosted by SURFsara and financially supported by The Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroneuro.2019.04.004.

References

Acosta, M.T., Swanson, J., Stehli, A., Molina, B.S.G., Martinez, A.F., Arcos-Burgos, M., Muenke, M., 2016. ADGRL3 (LPHN3) variants are associated with a refined phenotype of ADHD in the MTA study. Mol. Genet. Genom. Med. 4 (5), 540–547. doi:10.1002/mgg3.230.

Arcos-Burgos, M., Vélez, J.I., Martinez, A.F., Ribasés, M., Ramos-Quiroga, J.A., Sánchez-Mora, C., Muenke, M., 2019. ADGRL3 (LPHN3) variants predict substance use disorder. Transl. Psychiatry 9 (1), 42. doi:10.1038/s41398-019-0396-7.

Childress, A.C., Sallee, F.R., 2015. Emotional lability in patients with attention-deficit/hyperactivity disorder: impact of pharmacotherapy. CNS Drugs 29, 683–693. doi:10.1007/s40263-015-0264-9.

Coccaro, E.F., Ong, A.D., Seroczynski, A.D., Bergeman, C.S, 2012. Affective intensity and lability: heritability in adult male twins. J. Affect. Disord. 136 (3), 1011–1016. doi:10.1016/j.jad.2011.06.042.

Dechant, G., Barde, Y.-A., 2002. The neurotrophin receptor p75(NTR): novel functions and implications for diseases of the nervous system. Nature Neurosci. 5 (11), 1131–1136. doi:10.1038/nn1102-1131.

De Leeuw, C.A., Mooij, J.M., Heskes, T., Posthuma, D., 2015. MAGMA generalized gene-set analysis GWAS data. PLoS Comput. Biol. 11 (4), e1004219.

Demontis, D., Walters, R.K., Martin, J., M. M., Als, T.D., Agerbo, E., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J., Grove, J., Hansen, C.S., Hauberg, M.E., Hollegaard, M.V., Howrigan, D.P., Huang, H., Maller, J., Martin, A.R., J. M., Benjamin, M.N., 2019. Discovery of the first genome-wide significant risk loci for ADHD. Nat. Genet. 51, 63–75. doi:10.1038/s41588-018-0269-7.

Faraone, S.V., Larsson, H., 2018. Genetics of attention deficit hyperactivity disorder. Mol. Psychiatry doi:10.1038/s41380-018-0070-0.

Fatemi, S.H., Folsom, T.D., 2015. GABA receptor subunit distribution and FMRP-mGluR5 signaling abnormalities in the cerebellum of subjects with schizophrenia, mood disorders, and autism. Schizophr. Res. 167 (1–3), 42–56. doi:10.1016/j.schres.2014.10.010.

Ibrahim, K.E., El-Sherbeeny, A.M., Ahmad, S.F., Al-Harbi, N.O., Attia, S.M., Fardan, A.S., Nadeem, A., 2017. IL-17A causes depression-like symptoms via NFκB and p38MAPK signaling pathways in mice: implications for psoriasis associated depression. Cytokine 97 (February), 14–24. doi:10.1016/j.cyto.2017.05.018.

Marchant, B.K., Reimherr, F.W., Halls, C., Williams, E.D., Strong, R.E., Kondo, D., Robison, R.J., 2011. Long-term open-label response to atomoxetine in adult ADHD: influence of sex, emotional dysregulation, and double-blind response to atomoxetine. Atten. Deficit Hyperact. Disord. 3 (3), 237–244. doi:10.1007/s12402-011-0054-2.

Merwood, A., Chen, W., Rijsdijk, F., Skirrow, C., Larsson, H., Thapar, A., Asherson, P., 2014. Genetic associations between the symptoms of attention-deficit/hyperactivity disorder and emotional lability in child and adolescent twins. J. Am. Acad. Child Adolesc. Psychiatry 53 (2). doi:10.1016/j.jaac.2013.11.006, 209–220.e4.

O'Sullivan, M.L., Martini, F., von Daake, S., Comoletti, D., Ghosh, A., 2014. LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5. Neural Dev. 9 (1), 1–11. doi:10.1186/1749-8104-9-7.

Powers, A., Almli, L., Smith, A., Lori, A., Leveille, J., Ressler, K.J., Bradley, B., 2016. A genome-wide association study of emotion dysregulation: evidence for interleukin 2 receptor alpha. J. Psychiatr. Res. 83, 195–202. doi:10.1016/j.jpsychires.2016.09.006.

Riglin, L., Eyre, O., Cooper, M., Collishaw, S., Martin, J., Langley, K., Thapar, A., 2017. Investigating the genetic underpinnings of early-life irritability. Transl. Psychiatry 7 (9), e1241. doi:10.1038/tp.2017.212.

Shaw, P., Stringaris, A., Nigg, J., Leibenluft, E., 2014. Emotional dysregulation and attention-deficit/hyperactivity disorder. Am. J. Psychiatry 171 (3), 276–293. doi:10.1176/appi.ajp.2013.13070966.

Sloan, E, Hall, K, Moulding, R, Bryce, S, Mildred, H, Staiger, P.K, 2017. Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review. Clin. Psychol. Rev. 57, 141–163. doi:10.1016/j.cpr.2017.09.002.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74. doi:10.1038/nature15393.

Tsuang, M.T., Nossova, N., Yager, T., Tsuang, M.M., Guo, S.C., Kou, G.S., Liew, C.C., 2005. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am. J. Med. Genet. Neuropsychiatr. Genet. 133 B (1), 1–5. doi:10.1002/ajmg.b.30161.

Vidal, R., Valero, S., Nogueira, M., Palomar, G., Corrales, M., Richarte, V., Ramos-Quiroga, J.A., 2014. Emotional lability: the discriminative value in the diagnosis of attention deficit/hyperactivity disorder in adults. Compr. Psychiatry 55 (7), 1712–1719. doi:10.1016/j.comppsych.2014.07.001.

Wade, T.D., Gordon, S., Medland, S., Bulik, C.M., Heath, A.C., Montgomery, G.W., Martin, N.G., 2013. Genetic variants associated with disordered eating. Int. J. Eat. Disord. 46 (6), 594–608. doi:10.1002/eat.22133.

Ward, J., Strawbridge, R.J., Bailey, M.E.S., Graham, N., Ferguson, A., Lyall, D.M., Smith, D.J., 2017. Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. Transl. Psychiatry 7 (11), 1264. doi:10.1038/s41398-017-0012-7.

Wijchers, P.J.E.C., Hoekman, M.F.M., Burbach, J.P.H., Smidt, M.P., 2006. Identification of forkhead transcription factors in cortical and dopaminergic areas of the adult murine brain. Brain Res. 1068 (1), 23–33. doi:10.1016/j.brainres.2005.11.022.

Wu, D., Peng, Y., Zhou, J., Yang, Y.T., Rao, C.L., Bai, S.J., Xie, P., 2015. Identification and validation of argininosuccinate synthase as a candidate urinary biomarker for major depressive disorder. Clin. Chim. Acta 451, 142–148. doi:10.1016/j.cca.2015.09.018.

Xia, Z., Storm, D.R., 2005. The role of calmodulin as a signal integrator for synaptic plasticity. Nat. Rev. Neurosc. 6 (4), 267–276. doi:10.1038/nrn1647.