DNA methylation and its relationship with lifestyle, environmental and cardiovascular risk factors

PhD Candidate: Sergi Sayols Baixeras

DOCTORAL THESIS / 2018

DIRECTORS:

- **Dr. Roberto Elosua Llanos**
- **Dr. Carla Lluís Ganella**

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCE

PhD Programe in Biomedicine 2018

upf. Universitat Pompeu Fabra Barcelona

IMIM Parc de Salut MAR Barcelona

A ma mare i mon pare, per tots els sacrificis que han fet perquè jo arribés fins aquí.

Al meu germà, per ser un referent i el millor dels professors.

Als meus avis, per animar-me sempre a seguir endavant.

A la Marina, per ser aquella persona especial que sempre està al meu costat i és capaç de fer-me riure en qualsevol moment. Pel gran equip que formem i perquè el món se'ns queda petit.

A en Guim, per no haver arribat encara i ja haver-me canviat la visió del món.

‘Las gallinas que entran por las que salen’

Isaac Subirana and Oscar Díaz.

Agraïments

El desenvolupament d'aquesta tesi ha requerit un gran esforç i concentració, però no hauria estat possible sense l'ajuda directa o indirecta de molta gent.
A tots us ho vull agrair amb unes paraules:

A Roberto Elosua, por creer y apostar por mi cuando sabía que podía ser difícil encontrar financiación. Por su grandísima paciencia, por el tiempo que me ha dedicado, por sus ganas interminables de enseñarme y hacer de mí una persona mucho más crítica. Por todas sus bromas, por sus grandes consejos y por sus incansables ganas de ayudarme. Y sobretodo, por ser capaz de motivarme, de hacerme sentir interesado y apasionado por la investigación durante todo el doctorado.

A la Carla Lluís, per introduir-me i guiar-me en els meus inicis de la tesi i en l'anàlisi de grans bases de dades. Per les activitats post-feina i per estar sempre disposada a donar-me un cop de mà.

A l'Isaac Subirana, per les hores de paciència explicant-me com fer els anàlisis estadístics. Per la seva professionalitat i per la seva personalitat disposada a ajudar en tot moment. Sempre l'he considerat com el meu tercer director de tesi.

A la Marina, per estar sempre disposada a discutir resultats i entrebancs que he anat trobant al llarg de la tesi. Per ajudar-me sempre, fins i tot els últims dies amb l'edició final. Per ser capaç de distreure'm i ensenyar-me que hi ha vida més enllà de la tesi.

To Stella Aslibekyan, for accepting me in her group and introducing me in the causal inference world. I really appreciate that she made me feel like home during my USA experience. I also want to acknowledge all the help from her team (Ryan, Berta, H.Tiwari, Vinodh Srinivasasainagendra, Amit, Anh Do and Jin Sha).

A en Gavin, per les seves indicacions clau al començar al grup i el seu optimisme. Però sobretot, per la seva gran revisió de la tesi a través de la seva empresa ThePaperMill.

A l'Alba Fernández, per ser la meva companya predoc, amb qui compartir els camins de la ciència, les alegries i decepcions del dia a dia. Però sobretot per ser la persona amable i comprensiva que és. A l'Albert Prats, per il·luminar-nos en la causalitat i per ser una gran caixa de sorpreses.

Al conjunt del grup REGICOR, especialment a en Jaume Marrugat, per aportar sempre noves idees i una visió crítica de la meva feina. Així com a la Montse Fitó i a en Helmut Schröder. Als companys de despatx, per fer un ambient agradable i fer-me riure cada dia. Al doctor Hernáez, per haver-me acompanyat en aquest camí i haver fet més amens els dies interminables. A l'Oscar, per cadascuna de les converses que hem tingut i per ajudar-me a fer l'estandardització dels valors M de metilació. A la Dolors, per ser la nutricionista de capçalera i la primera persona que em saluda cada dia a la feina. A l'Albert Sanllorente, per ser el focus de distracció del despatx i la persona amb qui

comento l'actualitat i algun sopar de negocis. A la Ruwaida i a l'Anna Funtikova, per proporcionar-nos glucosa i la seva gran amabilitat. A la María Trinidad, per estar disposada sempre a escoltar i ajudar. A la Maria del Mar, per tenir sempre un somriure apunt. A la Gabriela, pels seus comentaris inesperats. A les postdoctorals del grup, per estar sempre disposades a ajudar-me i donar-me bons consells per millorar com a investigador (Irene Roman, María Grau, Betlem Salvador, Olga Castañer i Sara Arranz).

A les secretàries del grup (Yolanda Ferrer, Anna Blasco i Sandra Ferrer). I especialment a l'Esmeralda Gómez, que m'ha resolt en tot moment els embolics administratius. Als estadístics, per ajudar-me a entendre l'estadística fàcilment i per la seva paciència (Isaac Subirana, Joan Vila, Judith Peñafiel i Sílvia Pérez). A les noies de "data management", per ajudar-nos a tenir sempre les bases de dades apunt (Leny Franco, Susanna Tello, Marta Cabañero). Als tècnic de laboratori, pels riures inacabables (Gemma i Sònia) i especialment a en Dani, per ajudar-me en tot moment. A la gran llista de nutricionistes (Laura, Regina, Alejandra, Gal·la, Mireia, ...) i als estudiants que han anat passant, especialment en Santiago, en Sebas i la Maria.

A l'equip de voleibol dels *Vila's People*, per la seva entrega, per millorar els grans resultats en cada torneig, permetent-me desconnectar de tesi.

Als companys de l'empresa Gendiag.exe.

A tot el grup BRGE, dirigit per en Juan Ramón González, que m'ha permès assistir en les seves reunions i enriquir la meva formació.

Als ex-companys de pis, per les hores de distracció, per l'ajuda quan tenia problemes programant (Jordi, Lander, Don Teixidor i Maria Rosa). I algun inquilí inesperat que ens demostra que les lleis de la física són relatives (Robert).
Als RaBiolinis (Albert, Rafel, Adri, Max, Josu, Germàn, Àlex, Fran i Jordi), amb els que em vaig introduir en el món de la biologia. Pel seu recolzament i perquè amb ells tot és possible.

A la Maria Gomis i la Vicky Valls, per la seva inestimable ajuda i suport.
Als amics del poble, perquè sempre han estan al meu costat (Sabeña, Sara, David, Joan, Sandra i Jordi).

A tota la meva família, per haver-me animat i ajudat, especialment al meu germà i als meus pares.

Als participants de les diferents cohorts, els quals ens permeten seguir avançant en la recerca.
A tots aquells que m'he deixat i que em sabrà greu, però és que les presses no són bones conselleres.

Cruïlles, 24 de Setembre de 2018
Sergi Sayols Baixeras

'If I have seen further it is by standing on ye shoulders of Giants'

Isaac Newton. Letter to Robert Hooke (15th February 1676).

Content

Abbreviations

450K: Human Methylation450 BeadChip.
27K: HumanMethylation27 BeadChip.
ABDP: Apolipoprotein-B depleted plasma.
AUC: Area under the curve.
BASICMAR: BASe de datos de Ictus del hospital del MAR.
BMI: Body mass index.
Bp: Base Pair.
CAD: Coronary artery disease.
CARDIA: Coronary Artery Risk Development in Young Adults.
CETP: Cholesterol ester transfer protein.
CpG: Cytosine-phosphate-guanine.
CRFs: Cardiovascular risk factors.
CVD: Cardiovascular disease.
DALY: Disability-adjusted life-years.
dbGAP: Database of Genotypes and Phenotypes.
DMH: Differential methylation hybridization.
DNMT: DNA methyltransferase.
EGA: European Genome-phenome Archieve.
EPIC: HumanMethylationEPIC BeadChip.
EPIC-Italy: European Prospective Investigation into Cancer and nutrition.
ESTHER: Estrogen and Thromboembolism Risk.
EWAS: Epigenome-wide association studies.
EXaC: Exome Aggregation Consortium.
FDR: False discovery rate.
FHS: Framingham Heart Study.
FOS: Framingham Offspring Study.
gnomAD: Genome Aggregation Database.
GOLDN: Genetics of Lipid-Lowering Drugs and Diet Network.
GWAS: Genome-wide association studies.
H2DCF-DA: 2'-7'- dichlorodihydrofluorescein diacetate.
HII: HDL inflammatory index.
HPLC: High-performance liquid chromatography.
ICE FALCON: Inference about Causation through Examination of Familial Confounding.
IDL: Intermediate density lipoproteins.
KORA: Cooperative Health Research in the Augsburg Region.
LDL: Low density lipoproteins.
LDL-C: Low density lipoproteins cholesterol.
LINE: Long interspersed nuclear elements.
LUR: Land use regression.
miRNAs: MicroRNAs.
MeDIP: Methylation DNA immunoprecipitation.

MESA: Multi-Ethnic Study of Atherosclerosis.
mQTL: methylation quantitative trait loci.
MSDK: Methylation-specific digital karyotyping.
MSP: Methylation-sensitive PCR.
NO_2: Nitrogen dioxide.
NOX: Nitrogen oxides.
PCR: Polymorphism Chain Reaction.
PheWAS: Phenome-wide association study.
piRNAs: piwi-interacting RNAs.
PM_{10}: Particular matter with a diameter <10 μm.
PM2.5: Particular matter with a diameter <2.5 μm.
PMcoarse: Difference between PM10 and PM2.5.
PREDIMED: PREvención con Dieta MEDiterránea.
PROGENI: PROgram for GENetic Interaction.
REGICOR: REgistre GIroní del COR.
RLGS: Restriction Landmark Genome Scanning.
RRBS: Reduced representation bisulphite sequencing.
SCARB1: Scavenger receptor class B1.
siRNAs: Small-interfering RNAs.
SNPs: Single nucleotide polymorphisms.
T2D: Type 2 diabetes.
TC: Total Cholesterol.
TG: Triglycerides.
UAB: University of Alabama at Birmingham.
VLDL: Very low density lipoproteins.
WHI: Women's Health Initiative.

Abstract

Coronary artery disease is the first cause of death worldwide, and as a complex disease implies the interplay between genetic and environmental factors. Common genetic variants explain only 15-20% of its heritability. DNA methylation, which regulates the gene expression without altering the DNA sequence, has been proposed as a heritable signature to explain this missing heritability and as a mediator effect of lifestyle and environment on health.

To study the relationship between DNA methylation and smoking, air pollution, and cardiovascular risk factors (obesity, lipid profile, and HDL functionality) we used epigenome-wide association studies, Mendelian randomization, and multi-stage omics integration approaches. We included data from ten studies, but the results were mainly based on the REGICOR (REgistre GIroní del COR) cohort.

We report the association between smoking and 63 methylation sites (CpGs). We did not find association between air pollution and DNA methylation. We identified 94 CpGs associated with obesity and 14 with lipid profile. We observed that methylation at *CPT1A* and *SLC7A11* can modify or be affected by the triglycerides levels, highlighting the complexity of the lipids homeostasis. Finally, we performed the first study showing an association between DNA methylation and HDL functionality.

As conclusions of this research, smoking is strongly associated with a distinctive methylation pattern and there is a relationship between DNA methylation and several cardiovascular risk factors, although its causality is complex.

Resum

La malaltia cardiovascular, que és primera causa de mortalitat al món, com a fenotip complex depèn tant dels factors ambientals com dels genètics. Les variants genètiques comunes expliquen un 15-20% de l'heretabilitat de la malaltia. La metilació de l'ADN, que regula l'expressió gènica sense alterar la seqüència genètica primària, ha estat proposada com a mecanisme per unificar els efectes de l'estil de vida, els factors ambientals i els genètics.

Per estudiar les relacions entre metilació de l'ADN i estil de vida, factors ambientals i factors de risc cardiovascular hem utilitzat estudis d'associació d'epigenoma complet, randomització mendeliana i integració d'òmiques. Hem inclòs dades de 10 estudis diferents, però els resultats han estat basats sobretot en l'estudi REGICOR (REgistre Gironí del COR).

Aquesta tesi demostra l'associació entre 63 llocs de metilació (CpGs) i l'exposició al tabaquisme. No troba associació entre contaminació ambiental i metilació de l'ADN. Identifica 94 CpGs associats a obesitat i 14 als nivells lipídics. Observa que els nivells de triglicèrids en sang poden ser modificats i al mateix temps afectar la metilació en els gens *CPT1A* i *SLC7A11*, evidenciant la complexitat de l'homeostasis dels lípids. A més, s'ha realitzat el primer estudi que mostra l'associació entre la metilació de l'ADN i la funcionalitat de la HDL.

Com a conclusions d'aquesta recerca destaquen la forta associació del tabac a un patró de metilació específic i l'associació entre la metilació de l'ADN i diferents factors de risc cardiovascular amb una relació causal complexa.

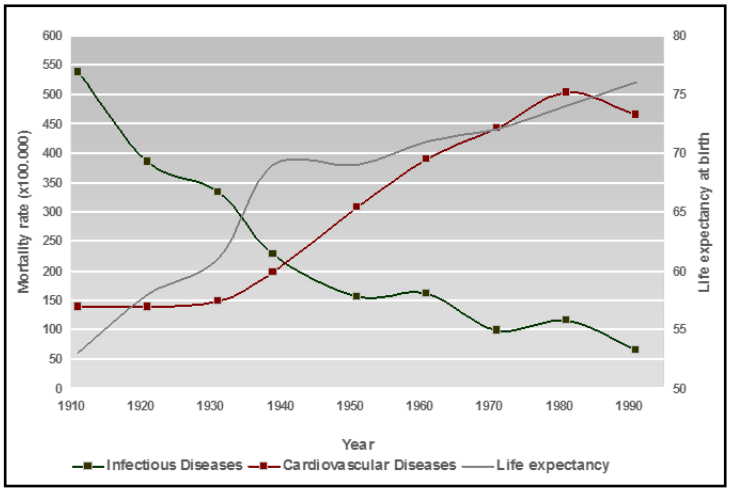
1. Introduction

1.1 CARDIOVASCULAR DISEASES

1.1.1 Increasing population burden of non-communicable complex diseases

In the early 19th century communicable or infectious diseases were the leading cause of death worldwide.[1] Improvements in hygiene and public health measures in the early 20th century, and the discovery and medical use of penicillin in the 1940s dramatically decreased mortality due to infectious diseases[1,2]. Moreover, vaccination campaigns in the United States and Europe further decreased mortality from these diseases during the second part of the 20th century.[2] This decline in infection-related mortality directly caused an increase in life expectancy and in turn an increase in mortality due to non-communicable complex diseases (**Figure 1**).

Figure 1. Epidemiological transition represented by a change in population-wide causes of mortality in the 20th century.



(Image from the national digital archives of datasets)

Non-communicable complex diseases are caused by an interplay between lifestyle, and environmental and genetic factors.[3] Recent international efforts such as the Global Burden of Disease Study are monitoring the frequency of the main diseases worldwide. This study has proposed several indicators for measuring the burden of a given disease in a population, and one of the most commonly used is disability-adjusted life-years (DALY). One DALY represents one lost year of "healthy" life, and is a summary indicator of mortality and morbidity at the population level. Thus, DALY is the sum of the number of years of life lost due to premature mortality, and the number of years during which people in a population have lived with a health condition of interest or its consequences. According to the

2016 Global Burden of Disease Study, the three diseases with the highest impact on DALYs were, coronary artery disease (CAD), cerebrovascular disease, and lower respiratory infection.[4] Therefore, CAD and cerebrovascular diseases, two complex non-communicable cardiovascular diseases, are leading contributors to the global burden of disease.[4]

1.1.2 Cardiovascular diseases

The term cardiovascular disease (CVD) refers to any disease that affects the heart or blood vessels. It is the first cause of death in the world, causing ~3.9 million deaths/year in Europe, and accounting for 45% of all deaths.[5] It also causes an enormous cost to the economy of the European Union economy: €210 billion per year, of which 53% is related to health care (€111 billion), 26% to lost productivity loss (€54 billion), and 21% to the informal care of people with CVD (€45 billion).[5]

Cardiovascular diseases include an extended list of individual diseases, of which CAD and cerebrovascular disease are the most important in epidemiological terms. In this doctoral thesis we will focus on CAD as the leading single cause of mortality in Europe, accounting for 19% (862,000 deaths) and 20% (877,000 deaths) of all deaths among men and women each year, respectively.[5]

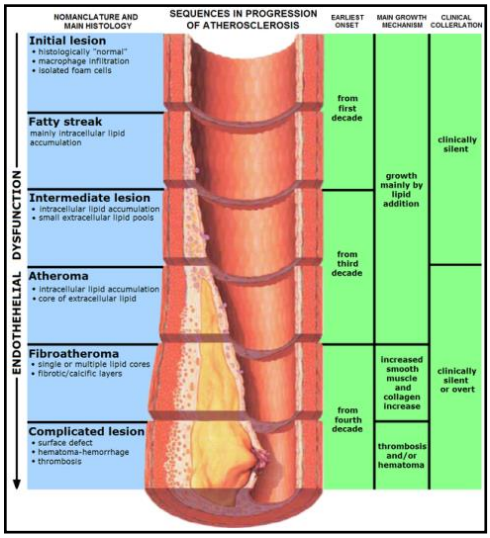
1.1.3 Historical context of cardiovascular epidemiology

Following the decline in mortality due to infectious diseases and the resulting increase in life expectancy, CVD became the leading cause of death worldwide from the middle of the 20th century.[2] In 1932, a seminal study reported an association between different dietary patterns and CAD mortality in different regions.[6] In this context, several epidemiology studies were initiated in the late 1940s to elucidate the causes of cardiovascular diseases.[7] In 1948 the United States Public Health Service started the Framingham Heart Study, carried out in Framingham, Massachusetts to study the causes and mechanisms underlying cardiovascular disease.[8] The Framingham investigators introduced the term "cardiovascular risk factor" in 1961.[9] These epidemiological observations resulted in a paradigm change in our perception of the causes of cardiovascular diseases, and influenced physicians to place greater emphasis on detecting, treating, and preventing risk factors. They also influenced individuals to personally take control of and reduce their risk for heart diseases.[7] The Framingham investigators were also pioneers in developing risk functions for estimating coronary or cardiovascular risk, with the first function being published in 1967.[10] Cardiovascular risk functions are currently used to estimate the probability of presenting a cardiovascular event, usually during the following 10 years, based on the presence or absence of cardiovascular risk factors (CRFs). This risk is used to define the intensity of cardiovascular prevention.[11]

1.1.4 Atherosclerosis: the main etiopathogenic mechanism

The main etiopathogenic mechanism of CAD is atherosclerosis, a complex chronic inflammatory process characterized by remodeling and narrowing of the coronary arteries (**Figure 2**) that supply oxygen to the heart, which can lead to ischemia or necrosis of the myocardium.[3] Key processes in atherosclerosis include lipoprotein oxidation, inflammation and immunity.[12,13] Atherosclerotic plaques progress from the initial lesion characterized by accumulated LDL-cholesterol in the sub-endothelial space of the artery, to complicated, advanced lesions characterized by a high burden of inflammation and a thin fibrous cap that can rupture or erode, triggering the formation of a thrombus and an acute event.[3]

Figure 2. Schematic progression of arteriosclerosis over time.



(Image from commons.wikimedia.org/wiki/File:Endo_dysfunction_Athero.PNG).

This complex process has various clinical cardiac manifestations, including stable angina, acute coronary syndrome, and sudden cardiac death. Like all complex diseases, it is caused by the interplay of environmental factors such as air pollution or lifestyle related factors (diet, smoking or physical activity), and inherited factors, which include genetic or epigenetic factors.[7]

1.1.5 Cardiovascular risk factors

CRFs are traits or measurable characteristics that increase one's probability of suffering a cardiovascular event, and they can provide independent significant predictive information about one's likelihood of having the disease in the future.[7] CRFs can be lifestyles, exposures to environmental factors, genetic variants, biochemical factors, and functional or imaging traits.[14]

More than 200 cardiovascular risk factors have been identified, and we can classify them in two main categories: non-modifiable and modifiable risk factors. The non-modifiable risk factors include age, sex, ethnicity, family history of CVD, and genetics, while the modifiable category includes lipid levels, blood pressure, diabetes, obesity, diet and physical activity. In this chapter I will describe some of the most important ones for estimating cardiovascular risk (age, sex, smoking, hypertension, lipid profile, and diabetes) as well as those that are important for this doctoral thesis (family history of CVD, obesity, HDL (high density lipoprotein) functionality, and air pollution).[7]

1.1.5.1 AGE

Age is the most important non-modifiable CRF. CVD risk exponentially increases with age, independently of all other risk factors. [15–17]

Atherosclerosis is a lifetime process that starts during fetal life, although its clinical manifestations usually appear in middle-aged and older adults.[18] Therefore, even in young individuals with high-risk profile of other CRFs (such as hypertension, obesity, diabetes and hypercholesterolemia), the estimated 10-year CVD risk is usually low, highlighting the importance of age as a predictor of CVD risk.[15]

Aging is a very complex process that depends on sex, lifestyles choices, environment and genetics. Currently, the scientific community is searching for molecular markers of aging (biological age) that will help to predict and improve monitoring of age-associated diseases.[19] Several methods can be used to estimate biological age, such as measuring telomere length, which accelerates or decelerates under different environments,[20] and gene expression,[21] especially of those related to metabolism and DNA repair. Recently the epigenetic clock[19,22] has shown that biological aging (based on DNA methylation markers[1*]) is associated with risk of stroke and mortality, even after adjusting for chronological age.[23,24]

1.1.5.2 SEX

Sex is also a non-modifiable CRF with a very significant impact on CVD risk. When we estimate cardiovascular risk using just age and sex we obtain a high discriminative capacity (C statistic around 0.70-0.75) to differentiate between individuals who will suffer a cardiovascular event in the next 10 years and those who will not.[16] Contrary to the common belief that CVD affects men more than women, CVD mortality is actually higher in women than in men.[5] However, it is true that most fatal cases of CVD are the result of cerebrovascular disease in women and CAD in men. Moreover, the incidence of CVD events is delayed by 5 to 10 years in women with respect to men. These differences between sexes have classically been attributed to a protective effect of sex hormones during

[1*] This thesis includes a section on epigenetic markers (Section 1.3).

reproductive life. Early menopause due to surgical removal of the ovaries is associated with higher CAD risk, with a relative risk of 4.55.[25] However, other factors can also explain this gap in risk between men and women.

Notably, the magnitude of the association between CRF and CAD risk differs between sexes. Young women tend to have a more favorable lipid profile than young men, but the menopause changes women's lipid profile, rising their levels to above those observed in men.[25] While female smokers previously tended to start smoking later in life, and to be exposed to a lower dose of tobacco than men,[26] currently the age of onset of smoking in women is similar to that of men, and women are now exposed to a similar dose of tobacco smoke.[25] Moreover, a meta-analysis showed that the effect of smoking on CAD risk was higher in women than in men (25% greater relative risk),[27] although other studies have shown that the association is only greater in heavy smokers (>20 cigarettes per day).[28] A large body of evidence shows that diabetes increase the relative risk of CAD (44%) and stroke (27%) more in women than in men.[29,30]

1.1.5.3 SMOKING

Smoking is a modifiable CRF. Although it is an independent CRF, it interacts with other major CRFs (HDL-cholesterol (HDL-C), hypertension, and diabetes mellitus). The main physiopathological mechanisms for CVD risk that are associated with smoking are endothelial dysfunction, and induction of a hypercoagulable state, which increase the risk of acute thrombosis.[31]

Prospective epidemiological studies have consistently shown that smoking increases risk of CVD mortality, and a meta-analysis showed a nonlinear dose-response related to daily smoking and relative risk of the disease.[32] Moreover, many studies have demonstrated that former smokers have lower risk of myocardial infarction and CVD death than current smokers.[33–35] Risk decreases very quickly after cessation, by around one-half after one year, although even 10 years after quitting, it remains slightly higher among former smokers than among never smokers.[36]

Secondhand smoking is also associated with increased risk of CAD. A meta-analysis found that individuals exposed to secondhand smoke have 25% to 30% higher risk of CAD than non-exposed individuals.[37] Smoking and second-hand smoking are also associated with other diseases, such as several cancers and respiratory diseases.

Recent surveys report a high prevalence of smoking in the Spanish adult population, around 27.6% in men and 18.6% in women.[5] Therefore, preventive strategies target smoking as a high priority, and several legislative initiatives have been implemented. Smoke-free legislation has been shown to reduce exposure to tobacco smoke among non-current smokers (reducing the prevalence of secondhand smokers).[38,39] Moreover, 2006 partial smoke-free legislation in Spain decreased the incidence and mortality of myocardial infarction by 11%.[40] Smoking control policies should also focus on reducing the use of any type of tobacco product, and on implementing better treatments to help smokers quit smoking.[38,41]

1.1.5.4 HYPERTENSION

Hypertension is a modifiable CRF that is usually asymptomatic. Epidemiological studies have shown a continuous, gradual and independent association between systolic and diastolic blood pressure and CVD risk.[7] Due to the linearity of this association, the definition of hypertension is currently a subject of controversy. For example, higher values of blood pressure, even in the normality range, are also associated with higher risk of CVD. To define hypertension, European guidelines recommend using the classical blood pressure limits (≥140 mmHg for systolic blood pressure or ≥90 mmHg for diastolic blood pressure), while American guidelines have proposed a change in the thresholds (≥130 mmHg for systolic blood pressure or ≥80 mmHg for diastolic blood pressure).[11,42]

The prevalence of hypertension (based on the classical criteria) is high, approximately 19% in the Spanish adult population.[5] In fact, most individuals who have hypertension are not aware of it, and most of those who are aware have not achieved adequate control.

Clinical trials have demonstrated that antihypertensive treatment is associated with a 35-40% decrease in risk of stroke, a 20-25% decrease in risk of myocardial infarction, and >50% decrease in risk of heart failure.[7] Therefore, it is important to increase awareness of this CRF, and improve its control among individuals with hypertension.

1.1.5.5 DIABETES

Type 2 diabetes (T2D) is a complex disease involving an heterogeneous group of chronic metabolic disorders characterized by hyperglycemia.[43] In epidemiological studies T2D is classically defined by a fasting serum glucose concentration of ≥126 mg/dL (or self-reported treatment for diabetes). It is one of the main modifiable risk factors associated with CAD, and increasing risk by 2-3 fold.[7] Women with diabetes have a higher risk than men with diabetes.[44]

Diabetes is also associated with other traditional CRFs, such as hypertriglyceridemia, lower levels of HDL-C, high blood pressure, and obesity, that usually appear before diabetes.[45] These risk factors have a common metabolism alteration which is insulin resistance. A patient presenting a combination of these CRFs is usually diagnosed with metabolic syndrome.[7] Metabolic syndrome has several specific definitions, but the most widely used is the presence of at least three of the following five criteria[46]:

i) Abdominal obesity (waist circumference >102 cm and >88 cm in men and women, respectively).
ii) Hypertriglyceridemia (>150 mg/dL)
iii) Low HDL-C (<40 mg/dL in men and <50 mg/dL in women).
iv) Altered glycaemia (>100 mg/dL).
v) High blood pressure (systolic blood pressure >130 mmHg and diastolic blood pressure >85 mmHg).

There is a controversy about whether the combination of these factors (i.e. the presence of metabolic syndrome) carries greater risk than the sum of the risks carried by the individual components, and thus provides more information about CAD risk than each risk factor separately.[7,47]

1.1.5.6 LIPIDS

Cholesterol and triacylglycerols (or triglycerides, TG) are the most important lipids for CAD. Cholesterol and TG are transported in the blood, and are joined to various proteins (apolipoproteins), thus forming lipoprotein particles. Depending on their molecular weight, these lipoproteins are classified as very low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipoproteins (LDL) or high density lipoproteins (HDL).

Low-density lipoprotein cholesterol (LDL-C) is the main lipoprotein that transports cholesterol to the cells.[48] LDL-C is a modifiable CRF, and is causally associated with CAD. This association is direct and linear, such that the higher the LDL-C the higher the CAD risk. Randomized controlled trials have demonstrated that risk of CAD events decreases following treatment with lipid-lowering therapies, such as statins, ezetimibe and PCSK9 monoclonal antibodies.[48] It has been estimated that a 1-mmol/L reduction in LDL-C leads to a 20% decrease in CAD-related mortality, and a 21-23% decrease in risk of CAD events.[49]

The main role of HDL particles is to remove excess cholesterol from cells and transport it to the liver, a process known as reverse cholesterol transport.[50] HDL-C is a modifiable CRF. Epidemiological studies have consistently showed an inverse association between HDL-C and CAD risk, and that HDL-C is a reliable predictor of CAD events.[51,52] It has been estimated that a 1-mg increase in HDL-C is associated with a 2-3% decrease in CAD risk.[53] However, randomized trials using pharmacological interventions that increase HDL-C concentration did not have any effect on the incidence of CAD events.[54] Moreover, Mendelian randomization studies suggest that the association between HDL-C and CAD is not causal.[55] Although HDL-C does not seem to be causally associated with CAD risk, HDL particles have several antiatherosclerotic functions not reflected by HDL-C concentration.[56] Therefore, research on HDL is now focusing on HDL quality (functionality) and not on HDL quantity (Section 1.1.5.7).

TG is mainly transported by VLDL and is an independent and modifiable risk factor for CAD.[57] Elevated levels of serum TG (>150 mg/dL) increase CAD risk with an OR of 1.5-17, even after adjusting for other associated risk factors.[57] Triglyceride-rich lipoproteins can penetrate the arterial wall and promote atherosclerosis.[58] High TG levels are also associated with the presence of more atherogenic LDL lipoproteins (smaller and more dense) and lower levels of HDL.[57] Although there is limited evidence from clinical trials about the effects of reducing TG, Mendelian randomization studies support a causal association between TG and CAD. A series of genes (*ANGPTL3*, *ANGPTL4* and *LPL*) have recently been found to be associated with TG, and work is now underway to develop drugs that target the proteins encoded by these genes.[59,60]

1.1.5.7 HDL FUNCTIONALITY

As mentioned above, HDL-C is no longer thought to be causally associated with CAD, and current work is now focusing on the functionality of the HDL particle.[56] Two of these functions are well studied: reverse cholesterol transport, and anti-inflammatory capacity.

HDL is the particle responsible for reverse cholesterol transport. This process starts with a key step, cholesterol efflux, in which cholesterol is transferred from the peripheral cell (usually macrophages in laboratory studies) to the HDL particle.[51] In animal models, cholesterol efflux has been causally related to reduced atherosclerosis.[61] In a human population-based cohort, cholesterol efflux capacity was found to be inversely associated with CVD incidence.[56] These authors found that individuals in the highest quartile of cholesterol efflux capacity had 67% lower CVD risk than those in the lowest quartile.[56]

Another key step in reverse cholesterol transport is the activity of the cholesterol ester transfer protein (CETP).[62] This enzyme catalyzes cholesterol transfer from the HDL particles to TG-rich lipoproteins (VLDL), and the transfer of TG from these lipoproteins to HDL.[62] To reduce cholesterol transfer from HDL to VLDL, with the aim of increasing reverse cholesterol transport to the liver, several companies have developed drugs to inhibit CETP activity. These drugs have been shown to increase HDL cholesterol levels, but most have failed to reduce CAD risk. The most striking results are seen with torcetrapib, which significantly increased HDL-C levels, but also increased CVD mortality, forcing the clinical trial to be stopped.[62,63] Anacetrapib is the only CETP inhibitor that has shown positive results for clinical outcomes, although these beneficial effects may be due to reduced LDL-C and not by an increase in HDL-C.[64]

The final step in reverse cholesterol transport is the transfer of cholesterol from HDL to the hepatocytes. This transfer is mediated by several receptors expressed in the hepatocyte membrane, including scavenger receptor class B1 (SCARB1). A recent genetic study reported that loss of functions mutations in *SCARB1* reduced this final transfer, increasing HDL-C levels and CAD risk, which highlights the importance of HDL functionality for levels of HDL-C.[65]

One of the most important antiatherogenic properties of HDL particles is their antioxidant function. This function is important for preventing atherosclerosis because LDL oxidation is a key event in initiating plaque formation.[62] Therefore, a decrease in the antioxidant activity of HDL has been associated with higher risk of CVD events.[66]

1.1.5.8 OBESITY

Obesity is the abnormal accumulation of body fat. It causes health problems and alters one's metabolic profile (lipids, glucose), and respiratory and cardiovascular function and structure.[7,67] There are several indicators for obesity, the most commonly used being body mass index (BMI), which is calculated from weight and height (BMI = Weight in Kg/Height^2 in meters). Obesity is classified

according to BMI, as follows: underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), class I obesity (30.0-34.9 kg/m^2), class II obesity (35.0-39.9 kg/m^2) and class III (≥40 kg/m^2).[67]

Obesity is considered a major public health problem. First, because its prevalence is increasing, currently ~23% of the adult European population is obese.[5,67] Second, it has several adverse health effects, such as higher risk of CVD, diabetes type II, hypertension, cancer, and sleep apnea.[7] Obesity is an independent modifiable risk factor for CAD, increasing risk by 2-3 times in obese adults, and accounting for 5% and 6% of CAD deaths in men and women, respectively.[68] The Framingham Heart Study also reported a dose-dependent pattern, where every 2 additional years lived with obesity corresponds to a 7% increase in the risk of CVD.[69]

Several studies have reported that higher BMI during childhood is associated with higher risk of CVD in adulthood. These data support the idea that atherosclerosis is a continuous process that starts early in life.[7]

1.1.5.9 AIR POLLUTION

Epidemiological studies have shown that exposure to air pollution is associated with higher CVD risk.[70] Air pollution can act both as a chronic factor that promotes atherosclerosis, and as an acute trigger of cardiovascular events.[70]

The mechanisms underlying the chronic and acute effects of air pollution are related to the inhalation and deposition in the alveoli of air pollution particles and other substances that generate a proinflammatory response and release molecules that modify the activity of the vascular tissue.[71] This response leads to an oxidative and inflammatory process that causes endothelial dysfunction, coagulation changes, autonomic imbalance, and vasoconstriction.[71] Hence, short-term exposure to air pollution increases risk of CAD, venous thromboembolism, arrhythmia, and cardiac arrest, and produces a measurable decrease in life-expectancy.[70] Long-term exposure is also associated with higher risk of type 2 diabetes, hypertension, and other atherosclerosis-related diseases.[71] Epidemiological studies have also shown that reduced exposure to air pollution results in a CVD risk decrease. A study estimated that 10 μg/m^3-decrease in the concentration of particles with >2.5μm diameter would result in 1,523 fewer cases of CAD and 3,156 fewer admissions for heart failure per year in US.[72]

The current recommendations for reducing air pollution exposure involve implementing and enforcing regulations, changing to affordable and renewable energy sources, reducing industrial emissions, and empowering individuals to prevent air pollution in their own microenvironments.[71]

1.1.5.10 OTHER RISK FACTORS AND MARKERS

More than 200 CVD risk factors have been described, including family history of CVD; lifestyle factors such as dietary patterns and physical activity; biochemical factors such as Lp(a), C-reactive protein, other inflammatory

biomarkers, and prothrombotic biomarkers; markers of vascular functional such as pulse wave velocity, endothelial function; vascular imaging measures, such as carotid or femoral intima-media thickness or the presence of plaques, and coronary tomography.[7]

1.2 GENETICS

Genetic factors have been suggested to be important for CVD because of the results of several cohort studies showing that a family history of CVD is associated with individual CVD risk.[73]

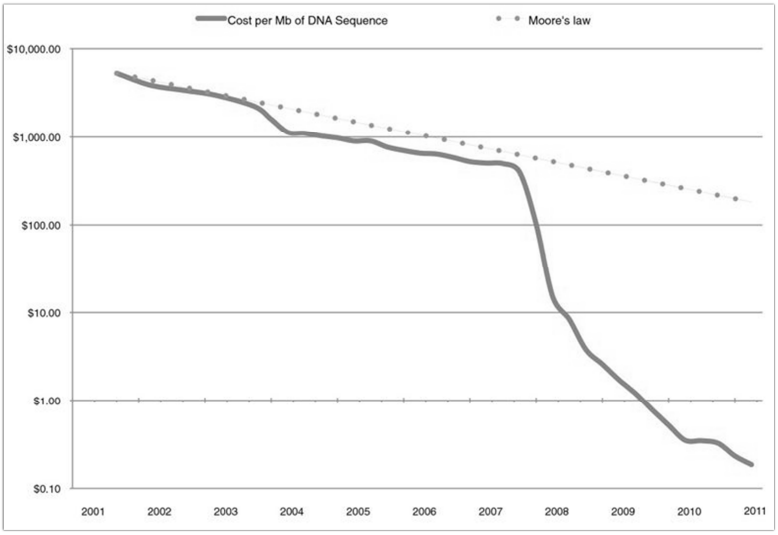
1.2.1 Brief history of genetics

The bases of genetic inheritance were established following the rediscovery in 1900 of the published work of Gregor Mendel describing his famous studies in pea plants in 1865.[74] Mendel's laws were subsequently explained and put in context by advances in our knowledge of cells and chromosomes. In 1902, Garrod observed that alkaptonuria followed the inheritance pattern described by Mendel,[75] and Bateson coined the term genetics in 1905 to describe inheritance. Johannsen created the concept of a gene to define the minimum unit of heredity in Mendelian inheritance, and introduced the concepts of genotype and phenotype in 1909.[76,77] In 1952, results from Hershey and Chase suggested that genes are made of DNA, and the following year Watson and Crick described the structure of the DNA molecule.[78,79]

The 20th century is full of discoveries in genetics that were also driven by the fast evolution of technology. After the discovery in 1975 of a method for sequencing DNA, a great advance in genetics appeared with the invention of PCR (Polymorphism Chain Reaction) in 1986, that allows us to copy and amplify DNA.[80,81] Following the first mapping of the human genome in 1987, the Human Genome Project was launched in 1990, with a draft sequence published in 2001 and a complete sequence in 2003.[82–84] Another two projects that have had a great impact in genetics are the HapMap Project and the 1000 Genomes Project.[85,86] HapMap started in 2002, and it's main goal was to determine common patterns of DNA sequence variations and make this information freely available.[85] The 1000 Genomes Project started in 2008 with the objective of elucidating most of the common and rare human genetic variation. This was the first study to sequence the whole genome of a large number of individuals, and by the time of it ended in 2015, the project had sequenced 2504 participants from 5 continents.[87] This created the most complete catalogue of human genetic variation, and the data were released to the scientific community. Technological advances have reduced the cost of sequencing faster that what was expected according to the Moore's law (**Figure 3**) from the field of electronics.[88] Moore's law states that the number of transistors of

an integrated circuit doubles approximately every 2 years, and is also applicable to other digital electronic devices, including sequencing technologies.[88] These advances have made it possible to create research consortia to sequence the entire exome (the protein coding part of the genome).[89] The Exome Aggregation Consortium (EXaC; http://exac.broadinstitute.org/) has sequenced the exome of 60,706 unrelated individuals from 4 continents, and has made summary data available that include the 1000 Genomes data.[89] This project has evolved into the Genome Aggregation Database (gnomAD; http://gnomad.broadinstitute.org), which has the goal of aggregating and harmonizing all available exome (n=123,136) and genome (n=15,496) sequencing data.[89]

Figure 3. Decreasing cost of DNA sequencing from 2001 to 2011 compared with the expected decrease according to Moore's law.



(Image modified from Sboner et al.[88])

1.2.2 Basic concepts in genetics

The human genome is located in the cell nucleus, and consists of 23 chromosomes containing three billion DNA base-pairs (bp).[90] Humans have a homologous pair of each chromosome, such that individuals have two copies of the same allele, one coming from each of the parents. This information codes 20,719 proteins, according to GENCODE 19 release.[90] However, it is believed that many of these genes do not translate into a functional protein and the total number of functional genes is closer to 19,000 than to 20,000.[91] A gene sequence is composed of many elements, mostly exons (the protein coding part of the genome) and introns (sequences removed by RNA splicing during maturation of the transcribed RNA). Exons represent only 1.5% of the entire DNA sequence.[90]

Genetic diseases are broadly classified into two categories according to their observed patterns of inheritance. One category is monogenic (oligogenic) or Mendelian diseases. Mendelian diseases are those in which one genetic variant in one gene accounts for most or all of the variation in diseases risk.[3] Their inheritance patterns follow Mendelian laws, and may be autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive or Y-linked, among others.[92] Although Mendelian diseases are cases in which a single mutation leads to a disease, their inheritance patterns can be complicated by pleiotropy and reduced penetrance.[93] The other category is complex diseases, which are characterized by complex patterns of inheritance. Disease risk is determined by multiple genetic variants (often with a small effect) combined with environmental factors, and is modulated by their mutual interaction.[3] Therefore, individuals that carry a specific genetic variant present an increased risk of a disease compared to other individuals without the variant, but it does not mean that the carrier individual will develop the disease. For example, in CAD the effects of known genetic variants range from an odds ratio of 1.04 to 1.30 per copy of the risk allele.[94]

1.2.3 Genetics of cardiovascular disease

In 1985 these two fields of knowledge, cardiovascular epidemiology and genetics, joined to report a deletion of several exons in the *LDLR* gene in a patient with homozygous familial hypercholesterolemia. This was the first demonstration of a mutation in a Mendelian cardiovascular phenotype.[95] Historically, four major approaches have been used in genetic research to discover genes for complex diseases (including for CVD, and CVD-related traits).[73]

1. Linkage Studies: Linkage studies are performed in kindreds in which more than one family member carries the disease. The disease is diagnosed in one member (index case), and this family is analyzed for hundreds or thousands of genetic markers distributed through the genome. Then, the transmission or sharing of these markers between related individuals is analyzed to identify where these markers coincide (segregate) with the appearance of the disease.[73] The aim is to detect the genomic region where the gene and the causal genetic variant are located. Once this genetic region is located, a fine mapping genotyping or sequencing can be carried out to identify the gene and causal genetic variant.[73] This kind of approach has been very useful for find causal genes in monogenic and oligogenic diseases. Linkage studies identified a region in the chromosome 1 containing the *PCSK9* gene, which is related to familial hypercholesterolemia.[96] Subsequent sequencing analysis identified variants of *PCSK9* that cause the disease.[97] Currently, monoclonal antibody treatments, known as anti-PCSK9, can be used to block the PCSK9 enzyme, resulting in greater availability of LDLR, and drastically reducing levels of circulating LDL-C.[98] However, linkage studies have been less useful for

studying complex diseases, although they have been used to identify genetic variants in *ALOX5AP* and *MEF2A* associated with CAD.[99,100]

2. Candidate gene approach: In this kind of study one or more genetic variants in a specific gene, usually common variants (allele frequency >5%), are selected *a priori* based on knowledge about the physiopathology of the disease. The researcher then tests whether these variants are more frequent in patients with the disease than in healthy control individuals.[73] This approach has contributed very little to the knowledge of the genetic architecture of CAD. The first candidate gene studied in cardiovascular genetics was *ACE*, in 1992. The authors reported an association between a deletion in the gene and myocardial infarction.[101]

3. Genome-wide association studies (GWAS): Fast technological progress and the HAPMAP study have driven the design of genotyping arrays that include 100,000 to 500,000 genetic variants, which capture a large proportion of common genetic variability.[73] These studies are hypothesis-free, and test for association between all of these genetic variants and the phenotype of interest. They require a population sample for the discovery phase, in which a series of sequence variants are identified as potentially associated with the phenotype. This is then followed by a second independent population sample in which the potential associations are replicated.[73] The simultaneous analysis of a large number of variants represents a large number of multiple comparisons, increasing the alpha error rate.[73] Several methods are available to correct for alpha error, such as the Bonferroni correction and the false discovery rate (FDR).[102]

   This type of study has required international collaborations between groups to assemble the large samples needed to achieve sufficient statistical power. This is because the corrected p-value considered to be statistically significant is very small (usually $<1 \cdot 10^{-8}$), and also because of the very small magnitude of the effects of common genetic variants on risk of complex phenotypes (e.g. the OR for CAD ranges from 1.05 to 1.40).[73] Several GWAS have identified ~176 genetic variants associated with CAD, and >300 associated with CRFs.[3,94,103–112] Most of these variants are located in intergenic regions close to gene promotors, indicating a possible influence on gene expression and epigenetics in determining CAD risk.[73] These variants only explain ~20% of the heritability of CAD (the proportion of the total population variance of the phenotype at a particular time or age that is attributable to genetic variation[3]), and some are also associated with lipid metabolism, blood pressure, and inflammation.[94,103] Due to the small effect of the variants associated with CAD, individually they have limited capacity to predict cardiovascular risk but they can be combined in polygenic risk scores and have better predictive capacity.[73] Using information provided by 6.6 million genetic variants captured by commercial genotyping arrays and imputation, a recent study has developed a new polygenic score that can

identify individuals with 4-fold greater risk of CAD than the population average.[113]

The main benefits of GWAS are the consistency of the results, the creation of international collaborative research networks, and the free availability of genetic data (crude and aggregate) to the scientific community. GWAS databases are available through the European Genome-phenome Archive (EGA) and the database of Genotypes and Phenotypes (dbGaP).[114,115] Aggregate data on associations between genetic variants and CAD, including data for >60,000 patients and >123,000 controls are available from the CARDIoGRAMplusC4D Consortium (www.cardiogramplusc4d.org). The UK Biobank (biobank.ctsu.ox.ac.uk) has also published and made available individual and aggregate data from GWAS of CAD and other phenotypes.

However, the main limitations of GWAS are that they do not identify the genetic variants that are causally associated with the phenotype (the causal associated variant could be in linkage disequilibrium with the identified variant). Also, they do not provide information about the pathophysiological mechanism,[73] which requires functional studies. Moreover, due to international collaboration among many groups, the phenotype definition can be heterogeneous. Also, GWAS are designed to identify common genetic variants, which generally have small effects, but are underpowered to detect rare variants that potentially have larger effects.[73]

4. Genome sequencing studies: Sequencing was classically used to study monogenic and oligogenic diseases.[73] Currently, it is possible to sequence the entire exome (the protein-encoding part of the genome) or the entire genome, which includes ~3.1 billion nucleotides and around 19,000 genes.[73] This type of study has identified genes with rare variants that usually result in a gain- or loss-of-function of the coded protein that is associated with the risk of CAD;[73] the genes identified include *LPL*, *ANGPTL3*, *ANGPTL4*, *SVEP1*, *LDLR* and *APOA5*.[116–119] The OR for the association between these rare genetic variants and CAD ranges between 1.5 and 4.5, which reinforces the idea that there are rare genetic variants with larger effect on the outcome than common variants (rare variants with small effect are also probably present but difficult to identify due to limited statistical power).[73]

Several studies have estimated that the heritability of CAD is between 35 and 55%.[73] However, in addition to genetics, family relationships also share values, lifestyles and environment, which may account for some of this heritability. Despite advances in identifying genetic variants associated with CAD, a large proportion of this heritability remains to be elucidated. This missing heritability is one of the main challenges of the genetics of complex diseases, and it could be accounted for by unknown common genetic variants, rare variants, complex gene-environment interactions, or regulatory changes in DNA that do not modify the DNA sequence (epigenetics).

1.2.4 Applications of genetics

Knowledge of the genetic architecture of CAD has at least three clinical applications:

1.2.4.1 IDENTIFICATION OF NEW THERAPEUTIC TARGETS

Genetics can be a powerful tool for discovering or validating therapeutic targets. Currently, there are three approaches for identifying therapeutic targets:

a) Association studies

Approximately only a third of the CAD-associated loci identified by GWAS are also associated with classical CRFs, which points to new mechanisms or possible therapeutic targets.[73] An analysis of 361 GWAS published up to February 2011 identified 155 genes that were targeted by ongoing drug therapy research.[120] For 63 of these genes the drug was used for treating or preventing the disease trait analyzed by the GWAS, while for 92 genes the drug could be used for treating a different disease.[120] While association studies have had some success, such as the identification of *PCSK9* as a potential therapeutic target, there are many other negative examples, highlighting the need to understand the mechanism underlying the association between the genetic variant and the phenotype.[73] For example, the first GWAS for CAD found a strong and consistent association in genomic region 9p21, in an intergenic region near to a gene cluster containing *CDKN2A* and *CDKN2B*.[121,122] Various hypotheses have been proposed to explain the association between this region and CAD, but the mechanism has still not be elucidated, which has restricted efforts to design new drugs to prevent cardiovascular events.[123]

b) Mendelian randomization studies

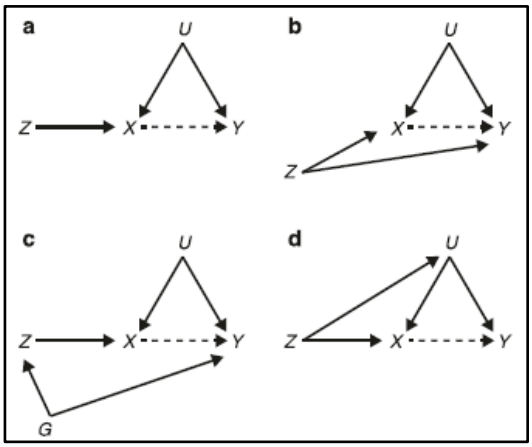
Mendelian randomization is an epidemiological method that uses single nucleotide polymorphisms (SNPs) as instrumental variables to test the causality of associations between a biomarker and a disease.[124] It takes advantage of the natural randomization that occurs at conception and Mendel's Second Law stating that an allele of one gene is transmitted between generations independently of that of another gene.[73] This approach attempts to simulate a randomized clinical trial, in that we can create two groups of individuals who are equal except for the genetic variant of interest associated with a phenotype of interest (biomarker). It is similar to a randomized controlled clinical trial, except that samples are randomized at the moment of conception rather than at the moment of recruitment.[125]

In this type of study, we select a biomarker associated with the disease, one or more genetic variants associated with the biomarker (genetic instrumental variable –GIV–), and we assess whether this GIV is associated with the disease.[126] If the results show a significant association between the GIV and the disease, this suggests a causal relationship between the biomarker and the disease. If there is no

association, this indicates that there is a non-causal relationship between the biomarker and the disease.[126] In this case, the association between the biomarker and the disease could be explained by reverse causation (the diseases cause changes to the biomarker) or that the relationship is mediated by another biomarker or confounder.[126] A Mendelian randomization approach makes three assumptions (**Figure 4**):[125,127,128]

1) There is a reliable association between the genetic instrument (Z) and the biomarker (X) (strength of the GIV) (**Figure 4a**).
2) The association between the genetic instrument (Z) and the clinical outcome (Y) is mediated through the biomarker (X) (**Figure 4b**) and not through a potential confounder (U) (**Figure 4d**).
3) The genetic instrument (Z) should not be associated with other biomarkers or exposures (G) (no pleiotropic effects) (**Figure 4c)**.

Figure 4. Assumptions of Mendelian randomization.



(Image from Latvala and Ollikainen.[127]) a) The genetic instrument (Z) must be associated with the biomarker (X). b) The genetic instrument (Z) must be associated with the outcome (Y) mediated through the biomarker (X). c) The genetic instrument (Z) cannot be associated with other biomarkers or exposures (G). d) The genetic instrument (Z) cannot be associated with confounder variables (U).

The statistical methods used for Mendelian randomization studies are constantly improving to allow the use of aggregate data, to correct for pleiotropic effects, and to control for confounding.[129] This approach has been used to assess the association between CRFs and CAD. It suggests a causal association from LDL-C, TG, interleukin 6, adiposity, and diabetes to CAD.[130–135] In contrast, it suggests that there is no causal association between CAD and HDL-C, C-reactive protein, adiponectin, bilirubin, and vitamin D.[130,131,136–141]

Mendelian randomization has also been used to provide information about the potential side effects of targeting a gene. For example, genetic variants in *PCSK9* that reduce LDL-C levels and risk of CAD also increase glucose levels,

waist circumference and risk of diabetes, suggesting that the drugs used to inhibit *PCSK9* have adverse effects.[142] This application of Mendelian randomization has become more popular since the publication of the first phenome-wide association study (PheWAS).[143] This type of study was initially proposed in 2013 to extend the classical genome-wide approach to electronic medical records, as an unbiased approach for replication and discovery that interrogates relationships between targeted genotypes and multiple phenotypes.[144]

c) Sequencing of loss-of-function variants

Loss-of-function variants are normally rare variants that result in a loss-of-function of the coded protein. This approach requires large studies using sequencing methods to detect genetic variants in specific genes.[73] It has become very useful for identifying and validating some proteins as therapeutic targets.[73] For example, loss-of-function variants in the *NPC1L1* gene encoding the NPC1L1 protein, the target for ezetimibe, were associated with lower risk of CAD. NPC1L1 was subsequently validated as target for CAD prevention by the IMPROVE-IT study, a secondary prevention clinical trial.[145,146] Another example is the discovery of the *ANGPTL4* gene as a new therapeutic target; monoclonal antibodies to the ANGPTL4 protein have been found to markedly reduce triglycerides levels in animal models,[59,118] and clinical trials are now underway to prove safety and efficacy.[59]

1.2.4.2 IMPROVING THE ESTIMATION OF CVD RISK

Individual CVD risk can be estimated using risk functions, which allows cardiologists to define the best preventive strategy.[11] These functions calculate one's risk of suffering a CVD event in the following 10 years according to their exposure to a series of risk factors.[147] However, risk functions have an important limitation in that most events occur among individuals with low or moderate estimated risk (they have low sensitivity).[147] Current research is trying to identify new biomarkers, including genetic variants that can be added to the function to increase this sensitivity. While most studies that have added CVD-associated genetic variants to risk functions have not achieved a better discriminative capacity, they do improve risk classification, especially in the moderate risk category.[148–174]

1.2.4.3 PHARMACOGENOMICS

Genetics can also affect the individual response to a certain drug treatment. Two clear examples in the cardiovascular field are well known:[73] a) Genetic polymorphisms at *SORT1/CELSR2/PSRC1, SLCO1B1, APOE* and *LPA* loci could explain ~5% of inter-individual variability in the reduction of LDL-C by simvastatin, as well as variants in *SLCO1B1* are associated with higher risk of myopathy induced by simvastatin.[175,176] b) In 2010, the US Food and Drug Administration added a warning to the clopidogrel leaflet explaining that genetic characteristics may stop some patients from experiencing the expected benefit.[177]

This drug is metabolized by cytochrome P450, and genetic variants in *CYP2C19* can modify the activity of this enzyme and thereby the efficacy of the drug.[3]

1.3 EPIGENETICS

1.3.1 From epigenesis to epigenetics: a historical perspective

In 1942 Conrad Waddington introduced the word epigenetics to modern biology with a definition related to epigenesis.[178] The epigenesis theory defines development as a gradual process of increasing complexity from initially homogeneous material in the egg following conception.[179] Waddington defined epigenetics as the entire complex of developmental processes that lie between genotype and phenotype.[178,180] His model of the epigenetic landscape highlights the different developmental pathways a cell might take during differentiation depending on the genes that underlie the landscape. Waddington hypothesized that the presence or absent of a gene determines which differentiation path the cell will follow after a certain point of divergence.[180] In 1958, Nanney cited Waddington's work but redefined the word epigenetics. He described two different cellular control systems: the genetic system, which uses the DNA sequence as a template to build a "library of specificities";[181] and an auxiliary, or epigenetic system, which determines the specificities to be expressed in a particular cell (for example the control of gene expression).[181]

In 1961 Jacob and Monod provided the operon model, the first demonstration of gene regulation in bacteria,[182] and this model fitted with Nanney's concept of epigenetics.[180] In 1969 Britten and Davidson proposed the first detailed model of gene regulation in higher organisms.[183] Research into histones and DNA marks started in the 1960s, although these two lines of research developed separately during at least two decades.[180] Modern chromatin and histone research started with the discovery of nucleosomes in 1973-4 by Vicent Allfrey and Alfred Mirsky, who confirmed the inhibitory effect of histones on transcription.[180] In the 1970s several groups reported a correlation between DNA methylation and gene inactivity, and Holliday and Riggs proposed that DNA methylation could regulate gene expression.[180] Cedar and Razin *in vitro* experiments showed that DNA methylation can repress gene expression, and that the methylation pattern is maintained from one cell generation to another.[180] In the 1990s the discovery of imprinted genes in mice and humans indicated that epigenetics is linked to DNA methylation. Holliday, who proposed several definitions of epigenetics during the 1980s and 1990s, defined it as "changes in gene expression through (de-)methylation of DNA".[180] However, after considering that histone modifications are also epigenetics, the definition was changed by Riggs in 1996 to "the study of

mitotically and meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence".[180,184]

1.3.2 Brief concepts in epigenetics

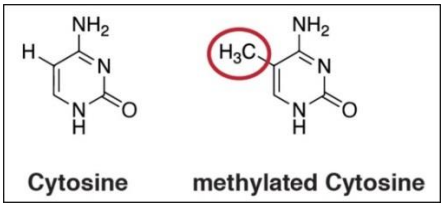
Epigenetics has gained increasing attention as a possible explanation for missing heritability because it is heritable but also modifiable by the environment, thereby representing a gene-environment interaction. During conception, both pro-nuclei undergo changes in epigenetic reprogramming before fusion.[185] Epigenetic reprogramming is a key step in early embryogenesis and gametogenesis that erases and reestablishes epigenetic marks in the embryonic genome during early embryonic development.[186] This phenomenon occurs twice, first during the blastocyst stage and second during development of the primordial germ cell.[185] Some genomic regions such as imprinted genes and most repetitive elements undergo epigenetic reprogramming but under different conditions, because they maintain the epigenetic marks.[185] Imprinting disorders occur because of the loss of imprinting in key imprinted loci.[185] Except in imprinted regions, epigenetic marks are normally erased from one generation to the next; in contrast epigenetic marks are conserved during cell division.[187] Moreover, epigenetic marks can be modified by the environment, and have been proposed as a possible biomarker of the exposome.[188]

Epigenetics is the study of potentially heritable and environmentally induced changes in gene expression without altering the DNA sequence. There are three main mechanisms:

1. **DNA methylation** is the most broadly studied and well-characterized epigenetic modification.[90] DNA methylation is the reversible addition of a methyl group to the 5th position of the pyrimidine of cytosine, forming 5-methylcytosine (**Figure 5**).[189] This reactions is catalyzed by enzymes of the DNMT family (DNA methyltransferase).[189] This addition is a common epigenetic mark in eukaryotes, and represents 3% of the cytosine nucleotides in the human genome, mostly in the context of cytosine-phosphate-guanine (CpG) dinucleotides.[90] In CpG regions, 70% to 80% of cytosines are 5-methylcytosine, which usually silences the gene.[190] CpG sites tend to cluster in regions known as CpG islands, frequently located at the promoter regions of genes.[190] DNA methylation within the body of the gene does not silence the gene, but may stimulate transcription elongation or alter splicing.[191] Thus, the location of methyl groups influences gene expression.[191] Moreover, many methylated sites are localized in repetitive

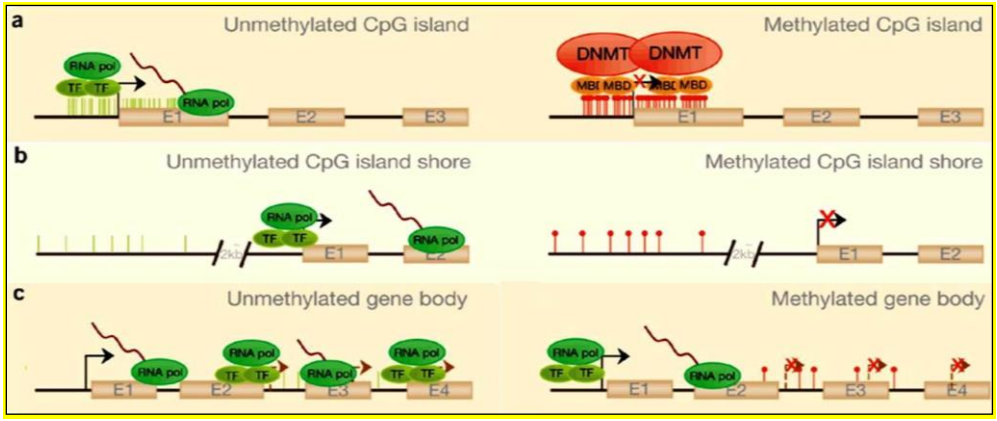
elements, and their function is not linked to transcription regulation, but rather to preserve chromosomal stability and avoid gene disruption[190] (**Figure 6**). Non-CpG methylation sites have also been described in humans. They have the structure CpHG and CpHH, where H is adenine, cytosine or thymine, and seem to be enriched in the gene bodies, and depleted at protein binding sites and enhancers.[190] Non-CpG sites have been identified in stem cells, decrease during differentiation, and are restored in induced pluripotent stem cells, suggesting a role in the origin and maintenance of the pluripotent state.[190] Given the dynamic nature of DNA methylation, epigenetics has been referred to as the interface between genome and environment.[192]

Figure 5. DNA methylation in a cytosine.



H, Hydrogen; N, Nitrogen; O, Oxygen; C; Carbon.
(Image from https://fr.wikipedia.org/wiki/Fichier:La_m%C3%A9thylation_de_l%27ADN_.png)

Figure 6. DNA methylation patterns in the genome.



(Image modified from Portela and Esteller)[190]. a) CpG islands at gene promoters are normally unmethylated, allowing transcription. Aberrant hypermethylation leads to transcriptional inactivation. b) The same pattern is observed when studying island shores, which are located up to 2 kb upstream of the CpG island. c) However, when methylation occurs at the gene body, it facilitates transcription, preventing spurious transcription initiations. In disease, the gene body tends to be demethylated, allowing transcription to be initiated at several incorrect sites.

2. Various **histone modifications** can occur in histone tails, including acetylation, methylation, phosphorylation, ubiqutination, SUMOylation and ADP-ribosylation.[190] Histone modifications have an important role in transcription regulation, DNA repair, DNA replication, alternative splicing and chromosome condensation.[193–195] Histones can be modified at different sites simultaneously; the histone state is not determined by a single histone mark, but rather by the combination of all marks in a nucleosome.[190] The nucleosome is formed by the core histones H2A, H2B, H3 and H4, which group into two H2.A-H2.B dimers and one H3-H4 tetramer. A 147 bp segment of DNA is wrapped in 1.65 turns around the histone octamer and neighboring nucleosomes are separated by around 50 bp of free DNA.[190] Histones changes can regulate gene expression by organizing the genome into formations: euchromatin, which has active and transcription-accessible DNA and is characterized by high levels of acetylation; and heterochromatin or trimethylated H3K4, H3K36 and H3K79, which has inactive and compacted DNA that is less accessible for transcription, and is characterized by low levels of acetylation and high levels of H3K9, H3K27 and H3K79.[190]

3. **Non-coding RNA** is a functional group of diverse RNA molecules that are transcribed from DNA but not translated into proteins.[190] Non-coding RNA can be divided into two groups: (1) Short ncRNAs (<30 nucleotides), including microRNAs (miRNAs), small-interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs); and (2) Long non-coding RNAs (>200 nucleotides),[196] whose function is to regulate gene expression at transcriptional and post-transcriptional level.[196]

1.3.3 DNA methylation and complex diseases

DNA methylation is the most widely studied epigenetic mark. Several studies have shown that methylation patterns could affect the disease phenotype through changes in the global and tissue-specific gene expression.[197] DNA methylation is considered an etiological candidate for complex diseases based on four observations[198]:

1. One of the characteristic traits of complex diseases is discordance between monozygotic twins (genetically identical twins arising from a single zygote), which is as high as 25-45% for diabetes, 50% for schizophrenia, and 40-70% for Alzheimer's disease. This discordance has been suggested to be due to differences in exposure to environmental factors.[199] Studies in twins showed similar levels of DNA methylation during childhood, but different

methylation patterns in adulthood.[200] Therefore, twins' risk of having a given disease can differ due to DNA methylation, reflecting differential exposure to a wide variety of environmental factors and lifestyles.[199]

2. The rapid increase in recent years in the incidence of several diseases such as diabetes and obesity cannot be explained by genetic factors alone. Epidemiological studies have demonstrated that the relationship between diabetes and obesity is at least partly mediated by nutritional and lifestyles factors.[201] These studies highlight the importance of non-genetic factors such as lifestyle and environment in risk of complex diseases.

3. Another peculiarity of complex diseases is sexual dimorphism. Males and females have marked differences in susceptibility to several complex diseases.[199] For example, multiple sclerosis, structural heart disease, osteoporosis and panic disorders are more common in females, whereas autism, alcoholism and Parkinson's disease are more common in men.[202] These diseases are known to be associated with autosomal genes, rather than genes on the sex chromosomes,[203] and there is no specific mechanism to explain these differences, or how sex hormones could predispose to or protect against a disease.[199] We know that sex hormones cannot change the DNA sequence, but they can modify and regulate gene expression through DNA methylation.[199]

4. Epidemiological studies demonstrate the importance of epigenetic reprogramming during embryonic development.[204] Exposure to various factors during the embryonic stage are associated with the appearance of several diseases in adult life.[204] Individuals who were prenatally exposed to the Dutch Hunger Winter (1944-1945) show lower levels of methylation in *IGF2* than non-exposed same-sex siblings.[205] Depending on the duration of their exposure to the famine, these individuals have higher risk of obesity or glucose intolerance. This is the first evidence that the early-life environment can cause epigenetic changes in humans that persist throughout life, and that predispose to disease.[206]

In this context it is very plausible that changes in DNA methylation are also import for cardiovascular disorders and atherosclerosis, and that this is modified by exposure to different environmental factors.

1.3.4 Methods for DNA methylation analysis

PCR is a revolutionary technique in molecular biology that allows us to copy a DNA fragment from a limited starting amount, and to amplify this up to a very large number of copies of this fragment.[207] However, PCR erases all methylation markers from the DNA fragment.[189] Given this, a major advance in

DNA methylation research was the development of sodium bisulfite modification,[208] which converts unmethylated cytosine nucleotides to uracil without modifying the methylated cytosine nucleotides.[208] In combination with PCR, this allows us to distinguish between unmethylated and methylated cytosine nucleotides.[189] The unmethylated cytosine nucleotides are amplified as thymine, and the methylated cytosine nucleotides remains as cytosines.[189] Several methods are available to analyze DNA methylation, and the choice of the most appropriate approach depends on the goals of the study in question.[209]

1.3.4.1 ANALYSIS OF GLOBAL DNA METHYLATION

As mentioned above, 70-80% of all CpG sites are methylated in repetitive elements and regions with low CpG density.[190] Global DNA methylation analysis provides information about most CpG sites, but not about loci in which the methylation status is altered. Thus, it is difficult to assess the relationship between methylation changes and functional outcomes.[209] Two main methods are available to measure global DNA methylation:

a) High-performance liquid chromatography (HPLC) is a highly quantitative and reproducible technique. However, it requires large amounts of high-quality DNA, so is not appropriate for high-throughput analysis.[209]

b) Bisulfite-based PCR methods do not require large amounts of DNA, and can be applied to paraffin-embedded tissues.[209] These methods estimate global DNA methylation by interrogating repetitive elements, mainly *Alu* elements and long interspersed nuclear elements (*LINE*).[209]

1.3.4.2 LOCUS-SPECIFIC ANALYSIS

This is a combination of methods that detects the level of DNA methylation at a specific locus.[209] Two main approaches are available, those that determine DNA methylation at a candidate locus, and those that do so in a genome-wide manner.[209]

1. Candidate locus approach: The candidate locus approach targets a specific locus or a small region, and the most common techniques use PCR methods.[209] The main concern is that methylated and unmethylated DNA sequences sometimes have different amplification efficiencies, leading to so-called PCR bias;[210] thus, PCR bias should be tested in every assay. The most common techniques to perform a candidate locus approach are:

a) Bisulfite and cloning sequencing provides information on allele-specific methylation, and is the gold standard. However, this technique is laborious and very expensive for large numbers of samples.[209]

b) Methylation-sensitive PCR (MSP) is a rapid, highly sensitive technique that can be used in DNA samples of poor quality and limited quantity.[209]

c) Direct bisulfite sequencing provides the average methylation information across all alleles produced after the bisulfite PCR.[209] This method is sensitive and quantitative, but is difficult to apply to many samples.[209]

d) Bisulfite-pyrosequencing is the main extension method for analyzing short to medium length DNA sequences.[209] It quantifies multiple CpG sites in the same reaction with high accuracy, and adds an internal control to the procedure.[209]

e) Bisulfite-PCR followed by MALDI-TOF MS is another high-throughput method that allows one to detect more CpG sites in a single amplicon (maximum of 800 bp); however, it is technically challenging.[209]

2. Genome-wide approach: The rapid advance of technology has changed the analysis of DNA methylation from a locus-specific analysis to a genome-wide analysis, allowing hypothesis-free approaches. The main techniques are:

a) Restriction Landmark Genome Scanning (RLGS) provides methylation information on thousands of loci in a single two-dimensional gel, although it has lower coverage and sensitivity, requiring 30% of methylation to be detected.[209,211]

b) Methylation-specific digital karyotyping (MSDK) has similar limitations to those of RLGS in that it provides low genome coverage and requires large quantities of DNA.[209,212]

c) Methylated DNA immunoprecipitation (MeDIP) technique requires large amounts of DNA and antibody.[209] Moreover, it may have low sensitivity in regions with low-density of CpG sites.[209]

d) The differential methylation hybridization (DMH) method is simple and needs low amount of DNA. It is based on the digestion of DNA using specific enzymes.[209] Therefore, only the CpG sites within the restrictive enzyme site can be analyzed and the results are prone to false positive due to incomplete digestion.[209]

e) Reduced representation bisulphite sequencing (RRBS) is used to sequence some regions after enzyme digestion (BglII or MspI) without selecting specific regions, but with enriched content of CpG sites.[189]

f) Whole genome bisulfite sequencing is the most extensive method, assessing 100% of the coverage. The limitations are the cost and the difficulty in managing the next-generation sequencing data.[213]

g) Bisulfite-PCR arrays methods are currently the gold standard for genome-wide methylation studies. They use previously selected oligonucleotides to quantify the cytosine nucleotides changed to thymine nucleotides at these regions.[209] Although these methods do not assess the whole methylome as whole-genome bisulfite sequencing does, but rather use loci-specific PCR, they interrogate hundreds of thousands of loci simultaneously, and with a lower cost.[209] The principal arrays are commercialized by Illumina.[213]

1.3.5 Illumina Infinium technology: Human Methylation450 Beadchip array

The excessive cost of whole-genome sequencing explains why Illumina's Infinium technology is the most widely used for large-population studies; the Human Methylation450 BeadChip array (450K) has been the most widely commercialized to date.[213,214] Illumina Infinium is a fast and cost-effective technology for analyzing methylation levels genome wide and at single CpG site resolution. The 450K succeeds the HumanMethylation27 BeadChip (27K), and simultaneously is the predecessor of the HumanMethylationEPIC BeadChip (EPIC).[215,216] The 450K platform assesses 485,577 assays (482,421 CpG sites, 3091 non-CpG sites and 65 random SNPs), including 94% of the loci interrogated by the 27K array (which assessed 27,578 CpG sites) with a R^2>0.95.[214] The new EPIC array interrogates 853,307 CpG sites, including 91.1% of the 450K probes with R^2>0.98.[216] In addition, the average correlation between replicates of the 450K array was 99% and the correlation with whole-genome sequencing is around 95%.[214] The 450K array prioritizes extensive coverage across genes and CpG islands,[214] and covers 98.9% of UCSC RefGenes with an average of 17.2 probes per gene region and 96% of CpG islands with an average of 5.63 probes.[214] The content of the array was selected according to the recommendations of a consortium including 19 institutions from around the world.[214]

Each well of the Illumina Infinium arrays contains 50-mer locus-specific sequences linked to silica beads.[217,218] These are randomly assigned to wells in the array such that each oligonucleotide is represented by 15-30 beads, providing repeat measurements.[217,218] This technology uses phosphoramidite chemistry, such that only the beads attached to a target can be decoded.[217,218] Particularly, the 450K array

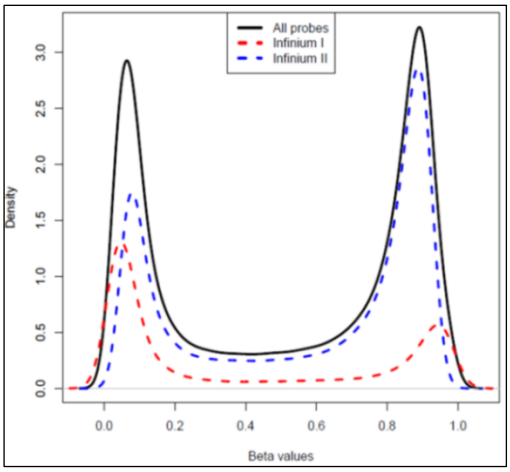
uses two different assays, Infinium I and Infinium II. Infinium I is the assay used in the 27K array and uses two different probes (one for the methylated allele and one for the unmethylated allele) attached to two different bead types and emitting the methylated and unmethylated signal for two different channels; 46,298 probes for the green channel, and 89,203 probes for the red channel.[219] In contrast, Infinium II (implemented for the first time in the 450K array) uses degenerate R-bases, allowing multiple combinations of oligonucleotides to be attached to the bead.[218] Therefore, one probe can hybridize with both the methylated and unmethylated cytosine after bisulfite conversion, emitting in the green and red channels, respectively.[218] It uses a total of 350,076 probes.[214] The main steps performed with this technology are: First, the DNA undergoes bisulfite treatment and is amplified by PCR;[217] second, the amplified DNA is fragmented and hybridized to the array;[217] third, primer extension and immunohistochemistry staining are performed.[217] Infinium I requires two probes to assess a CpG site after allele-specific single-base extension, but Infinium II requires one bead and determines the methylation level at a specific CpG site by single-base extension of the matched bead over the mismatched bead using hapten-labeled terminators.[217] Finally, the methylation levels are estimated from the detected signals by two parameters, using equations a) and b):[220,221]

a) $\beta = \left(\frac{M_i}{M_i + U_i + \alpha} \right)$ b) $M = \log_2 \left(\frac{M_i + \alpha}{U_i + \alpha} \right)$

c) $\beta = \left(\frac{2^{M_i}}{2^{M_i}+1} \right)$ d) $M = \log_2 \left(\frac{\beta_i}{1-\beta_i} \right)$

Where: Mi is the intensity of methylated probes, Ui is the intensity of unmethylated probes, and α is a constant offset with a value of 100 and 1 in equations a and b, respectively.[220] The β-value facilitates biological interpretation, in that it ranges from 0 for completely unmethylated to 1 for completely methylated.[220] M-values have a more difficult interpretation, as M-values close to 0 correspond to half-methylated probes,[220] while positive and negative M-values indicate more and less methylated than unmethylated cytosine nucleotides, respectively.[220] However, the M-value has better statistical properties,[220] in that it has better detection power and a higher true positive rate due to its reduced heteroscedasticity.[220] The β-value can be transformed to the M-value and vice versa using equations c) and d). Several studies have detected a shift in the β-value distribution between Infinium I and II. Infinium II is slightly less accurate than Infinium I at the extreme values.[220] Consequently, the overall distribution from the Infinium I and Infinium II does not reflect the true biological effect, and a series of normalization methods have been developed to solve this problem (**Figure 7**).[222]

Figure 7. Distribution of the β-value for the Infinium I and Infinium II assay and the overall β-distribution.



The normalization step tries to remove the experimental artefacts, random noise and technical and systemic variation caused by both assays and identifies the true biological signal.[222] Most currently normalization methods are classified as i) between-array normalization, which removes technical artefacts between samples on different arrays, and within-array normalization, which corrects for intensity-related dye biases.[223] Therefore, it is essential to normalize the data before analyzing the methylation levels in a region that is interrogated using both Infinium assays.[222] However, normalization is not essential when analyzing single position, although some authors recommend it. Some methods should also be used to account for batch effects.

1.3.6 Epigenome-wide association studies

Advances in technology and especially of the Illumina Infinium technology have allowed us to study associations with DNA methylation in a genome-wide manner.[209] Epigenome-wide association studies (EWAS) assess the relationship between a phenotype of interest and the methylation status of single positions throughout the genome.[224] Similar to GWAS, many different study designs can be used (case-control, monozygotic twins studies, longitudinal cohorts, etc.), but there is no single EWAS design that will be useful for all proposals.[224] The researcher must identify the most adequate to address their hypotheses.

There are some notable differences between EWAS and GWAS:

i. The first difference is the source of tissue from which to extract the DNA. While GWAS can use DNA from any tissue because the DNA sequence is "*equal and immutable*" in all body cells (not entirely true

because of the presence of somatic clonal cells). In contrast, DNA methylation is tissue-specific and dynamic. All tissues are composed of multiple cell types, such that heterogeneity is possible even within a single tissue.[224] For example, there are more than 50 distinct cell types in the blood, so it would not be possible to detect methylation patterns that affect the risk of a disease caused by only one cell type that makes up only a small proportion of cells in the blood. Differences detected between cases and controls may only reflect different cell type compositions, but not real epigenetic variation.[224] However, most studies involve living participants, so the DNA must come from an easily accessible source such as blood, buccal samples, saliva, hair follicles, urine, or faeces; blood is the most widely used.[224] Some authors have suggested that studying methylation patterns of whole blood is a good proxy for the methylation levels at a specific site of action.[225,226]

ii. DNA methylation is dynamic and changes over time according to various environmental and lifestyle.[192] Thus, unlike GWAS data, which can be collected once in a patients' life and remains constant throughout their life, EWAS information is valid for a time point, and thus confers invaluable information as a biomarker.[224] The dynamic of DNA methylation patterns over time affect studies of methylation-phenotype associations. While GWAS support causal associations between genetic variants and the phenotype, EWAS results do not support this causal association, because DNA methylation could be a consequence of the phenotype of interest rather than a cause. The causal direction in EWAS can be examined by using an appropriate study design or by taking advantage of genetic variants (Mendelian randomization approach).[129] It is important to understand the causal relationship in order to assess the usefulness of methylated loci as, for example, markers of disease progression, targets for reversal by treatment, or measures of drug response.[224]

iii. Statistical power is a key determinant of success in genome-wide studies. A EWAS study of a given sample size usually has higher statistical power than a GWAS of the same size because DNA methylation level is a continuous variable, unlike the genetic variables analyzed by GWAS, which can have only 3 possible values (0, 1 or 2). Both GWAS and EWAS involve testing hundreds of thousands of hypotheses and must be corrected for multiple testing to reduce type I error to 5%,[227] using the Bonferroni correction or the FDR correction.[102]

iv. Similar to GWAS, EWAS can also be confounded by systematic differences between cases and controls in the processing of samples (known as the batch effect), or when the ancestry of the cases differs systematically from that of the controls (known as population

structure).[224] However, in EWAS, environmental factors can affect both the epigenotype and phenotype, inflating the type I error and resulting in exaggerated estimates of the effect size.[224] Therefore, all analyses should be adjusted for potential confounders like age, sex and smoking exposure where possible.[224] Further methods using high-throughput data are available to correct for confounding factors such as principal components, surrogate variables, etc.[224,228]

The typical structure for GWAS and EWAS involves a population sample to perform the discovery analyses and an independent population to validate the results.[224] The most commonly used strategy is to initially tolerate some false positive results in the discovery stage and to include and analyze them in the validation stage.[224] The validation phase should ideally include an independent sample studying the same methylation positions using the same trait definition (phenotype or exposure).[229] In EWAS, the DNA should be extracted from the same source due to its tissue-specificity. Moreover, correlations between methylation and genotype to assess causation should be assessed in the replication cohort.[224] To evaluate the relationship between the methylation site and the phenotype, an optimal approach is to combine the statistical tests from the two stages using standard meta-analysis techniques.[224]

The purpose of identifying associations using the EWAS approach is to improve our understanding of disease etiology and provide possible targets to develop therapeutic strategies and new biomarkers for diagnoses or disease prediction.[224]

1.4 JUSTIFICATION OF THIS THESIS

Cardiovascular diseases in general and CAD in particular is the main cause of death worldwide. Several cardiovascular risk factors predisposing to this complex disease have been identified. The risk of presenting CAD depends on a complex interplay between genetics, environmental and lifestyles factors. Several studies have estimated that the heritability of CAD ranges from 35% to 55%. Common genetic variants identified by GWAS explain only around 15-20% of this heritability, and remainder is called missing heritability.

Epigenetic is the study of mitotically and meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence. Some of these epigenetic signatures can be modified by environmental and lifestyles factors. Therefore, studying the relationship between epigenetics and CAD risk could explain some of the missing heritability. Epigenetics could also mediate the relation between environmental/lifestyles factors and CAD risk. DNA methylation is the most stable and widely studied epigenetic mark, and has been proposed as a

mechanism to explain the role of environmental factors and to explain missing heritability.

Thanks to technological advances, the study of DNA methylation using a genome-wide approach is now cost-effective and can be performed in epidemiological studies. Once the DNA signatures related to biological traits are identified, it is important to assess whether these relationships are causal or not. The study of causality will support the potential role of these epigenetic marks as biomarkers and potential therapeutic targets.

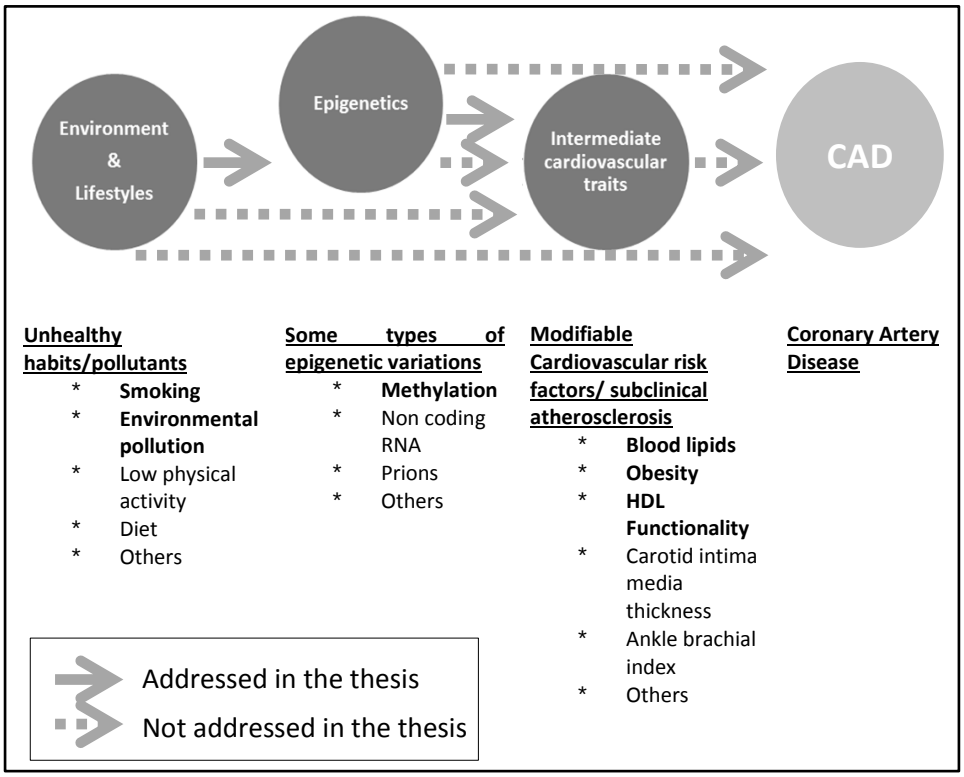
In this thesis I will tackle some of these issues and provide new evidence to expand knowledge on this topic. Specifically, I will address the association between some lifestyles/environmental factors (smoking and air pollution) and DNA methylation, and between DNA methylation and several cardiovascular risk factors (adiposity, lipids, HDL functionality). I will also address the causality of the association between DNA methylation signatures and lipid profile.

2. Hypotheses and Objectives

2.1 HYPOTHESES

The lifestyle and environmental exposures influence DNA methylation signatures, so the DNA methylation changes could partially mediate the effect of lifestyle and environment on the development of CAD (**Figure 8**). Moreover, independently of this mediator effect, DNA methylation could also be associated with intermediate cardiovascular traits, which could partly explain the missing heritability and pathogenesis of CAD.[2*]

Figure 8. Scheme of the thesis' hypotheses.



General hypothesis 1: The lifestyle and environmental exposures influence DNA methylation signatures.

[2*] In this thesis, the associations between DNA methylation and the other factors are written according to the direction of the causality shown in the Figure 8, in which according to the statistical models designed, it is first mentioned the independent and then the dependent variable.

Specific hypotheses:

1.1 Smoking exposure modulates DNA methylation patterns throughout the genome. Smoking is associated with differential levels of DNA methylation at several specific loci:
- Current smokers have a different DNA methylation signature to that in never smokers.
- Former smokers have a different DNA methylation signature to that in current and never smokers.
- Different DNA methylation patterns could be used to discriminate between current and never smokers.
- Different DNA methylation patterns are associated with gene expression and plasma protein levels.

1.2 Different exposures to air pollution can result in different DNA methylation patterns throughout the genome:
- Exposure to traffic related air pollution in residential areas (residential NO2, traffic intensity in the nearest street, and traffic load within 100 meters) is associated with DNA methylation levels.
- Exposure to particulate matter is associated with differential DNA methylation.

General hypothesis 2: DNA methylation could be associated with intermediate cardiovascular traits.

Specific hypotheses:

2.1 DNA methylation signatures are associated with adiposity indicators (body mass index and waist circumference).

2.2 DNA methylation signatures are associated with serum lipid profile (low density lipoprotein cholesterol, high density lipoprotein cholesterol, and triglycerides). Moreover, these different DNA methylation signatures are associated with levels of gene expression.

2.3 Some DNA methylation loci associated with serum lipid levels directly cause this association with lipids, whereas others are a consequence of exposure to these lipid levels (reversal causality).

2.4 DNA methylation signatures are associated with high density lipoprotein function (cholesterol efflux capacity and/or anti-inflammatory capacity).

2.2 OBJECTIVES

General objective 1: To assess the association between lifestyle and environmental factors and DNA methylation.

Specific objectives:

1.1 To determine the association between smoking exposure and DNA methylation levels throughout the genome (Manuscript 1) by:
- Comparing current active smokers with never smokers.
- Comparing former smokers with current and never smokers.
- Analyzing whether DNA methylation patterns can be used to discriminate between smokers and never smokers.
- Exploring the relationship between differential methylation at specific loci and gene expression and plasma protein levels.

1.2 To assess whether traffic-related air pollution and particulate matter in residential areas are related to DNA methylation signatures throughout the genome (Manuscript 2).

General objective 2: To determine the association between DNA methylation and intermediate cardiovascular traits.

Specific objectives:

2.1 To determine the relationship between DNA methylation signatures and adiposity indicators (body mass index or waist) (Manuscript 3).

2.2 To test for association between DNA methylation and serum lipid profile (low density lipoprotein cholesterol, high density lipoprotein cholesterol, and triglycerides) (Manuscript 4) by:
- Analyzing the linear relationship between DNA methylation and serum lipid traits.
- Exploring the relationship between differential methylation at specific loci and gene expression.

2.3 To evaluate causality in the association between DNA methylation and serum lipid levels (Manuscript 5).

2.4 To explore the relationship between DNA methylation signatures and high density lipoprotein functionality (cholesterol efflux capacity or anti-inflammatory capacity) (Manuscript 6).

3. Methods

The six specific objectives, corresponding to five different phenotypes, are addressed one by one in the six manuscripts presented in this doctoral thesis. The methods used for each objective are explained in detail in the corresponding manuscript. In this section I present a summary of the methods.

3.1 STUDY DESIGN

We designed two-stage EWAS to address most of the objectives of this doctoral thesis (1.1, 1.2, 2.1, 2.2; Manuscripts 1 to 4). In the first stage (discovery), we assessed the association between the exposure (smoking, air pollution) or the clinical trait (obesity, lipids) of interest and DNA methylation in a genome-wide manner in the REGICOR study. We selected CpG sites associated with the exposure or clinical trait for subsequent validation, using an arbitrary p-value threshold of $<1 \cdot 10^{-5}$. In the second stage (replication), we evaluated the selected CpG sites in one or more independent populations (**Table 1**), and meta-analyzed the results from these study populations to assess whether the associations selected in the discovery phase were statistically significant. To declare a candidate association as statistically significant we used the Bonferroni correction. This adjusts for the multiple comparisons performed in the analyses, which we established as the number of CpG sites analyzed in the discovery stage (p-value threshold for statistical significance = 0.05/number of CpG sites analyzed in the discovery stage).

To address objective 2.3 (Manuscript 5), we used the Mendelian randomization approach to evaluate the causality of the association between DNA methylation signatures and lipid profile. First, we selected the results reported in Manuscript 4. Second, we replicated the associations in a third database (GAW20 data). Third, we identified genetic variants associated with methylation levels at the CpG site of interest and created a polygenic risk score (genetic instrumental variable) to analyze whether the association from methylation to lipid level was causal. To analyze causality in the opposite direction, i.e. that lipid levels modulate methylation levels, we selected genetic variants associated with the lipid traits of interest, and created polygenic risk scores for these traits. In these analyses we also searched for pleiotropic effects.

To address objective 2.4 (Manuscript 6), we also used an EWAS design, but only included a discovery sample, as no external population with data on methylome and HDL functionality were available to replicate the results.

In Manuscripts 2 and 6, we also used a candidate gene approach to replicate CpG sites in genes previously related to the exposures or traits of interest.

Table 1. Summary of the populations and -omics data used to address the objectives in this thesis.

Study Population			Omic Data						Objectives (O) – Manuscripts (M)					
Cohort Name	**Information**	**Ref.**	**Genetics**	**DNA methylation**		**Transcriptomics**		**Prot.**	**O 1.1 M 1**	**O 1.2 M 2**	**O 2.1 M 3**	**O 2.2 M 4**	**O 2.3 M 5**	**O 2.4 M 6**
				450K	**Pyros.**	**RNAs.**	**µarray**							
REGICOR	Population-based cohort representing Girona Province (n=11,158).	230		● N=648	● N=622				●	●	●	●		●
BASICMAR	Prospective registry of acute ischemic stroke patients within the reference area of Hospital del Mar, Barcelona (n=7,000).	231,232		● N=377					●					
PREDIMED	Large, multicenter randomized, controlled, parallel trial to study the long-term effects of a traditional Mediterranean diet, Spain (n=7,447).	233,234					● N=22		●					
TALAVERA	Population-based survey undertaken in Talavera	235						● N=267	●					
EPIC-Italy	Multi-center prospective study including data from two case-control studies on breast and colorectal cancer, Italy (n=47,749).	236,237		● N=454						●				
FOS	Population-based cohort including 5,209 offspring of participants in the original Framingham Heart Study. Data obtained through dbGAP.	238		● N=2,542							●	●		
GOLDN / GAW20	Families of self-reported European descent with at	239	● N=655	● N=993		● N=98						●	●	

	least two siblings, recruited from two centers, in Minneapolis and Salt Lake City.													
GLGC	Summary statistics for SNPs associated with lipid traits from 45 studies (n=188,577). Aggregated data available at: http://csg.sph.umich.edu/willer/public/lipids2013/	240	● N=188,577									●		
CARDIoGRAMplusC4D	Consortium including the Coronary Artery Disease Genome wide Replication and Meta-analysis (CARDIOGRAM) and the Coronary Artery Disease (C4D) consortia (n=185,000). Aggregated data available at: http://www.cardiogramplusc4d.org/data-downloads/	241	● N=~185,000											●
GIANT	Consortium focused on identifying genetic loci that modulate human body size and shape, including 80 GWAS plus data from 34 additional studies (n=339,224). Aggregated data available at: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files	107	● N=339,224								●			

Ref.: References; 450K: HumanMethylation 450 Beadchip Kit; Piros.: Pyrosequencing technology; RNAs.: RNA sequencing technology; µarray: Microarray technology; Prot.: Proteomics.

3.2 STUDY POPULATIONS

To address the hypotheses of this thesis, we included individual or aggregated data from ten different studies. All cohorts are briefly summarized in **Table 1,** explained in detail in the corresponding manuscripts,[107,230–237,239–241] and described below:

1. **The REGICOR** (REgistre GIroní del COR or Girona Heart Register) study was the main source of population data.[230] It was used in five of the six manuscripts (Objective 1.1, 1.2, 2.1, 2.2 and 2.4) Manuscripts 1, 2, 3, 4, 6) as a discovery cohort, and also used as a replication cohort in Manuscripts 1, 2 and 6. The REGICOR study is a population-based cohort created from three population-based surveys: i) 1,748 individuals aged 25-74 years recruited from 33 villages in 1995; ii) 3,058 individuals aged 25-74 years recruited from 14 villages in 2000; and iii) 6,352 individuals aged 35-79 years recruited from 5 villages in 2005. The participation rate was >70% in all three surveys. The 11,158 participants included are of European ancestry and represents the urban and rural diversity of the province of Girona in Catalonia. All participants were invited to participate in a follow-up visit in 2009-2013. A random subsample of 648 participants recruited in 2005 and attending the follow-up visit was selected to participate in these EWAS as a discovery population, and an additional random subsample of 622 individuals was used as a replication cohort in Objective 1.1, Manuscript 1.[230]

2. **The BASICMAR** (BASe de datos de ICtus del hospital del MAR) study is a hospital-based prospective register of stroke patients since 2005. All stroke patients who are attended at Hospital del Mar, Barcelona, are invited to participate in the register.[231,232] BASICMAR currently includes more than 7,000 patients. A random subsample of 377 patients was selected to analyze genome-wide DNA methylation. This study was a replication population to analyze the association between smoking and DNA methylation (Objective 1.1, Manuscript 1).

3. **The PREDIMED** (PREvención con DIeta MEDiterránea) study was used to study DNA methylation and gene expression data.[233,234] We had access to genome-wide gene expression data for a subsample of 22 individuals. We included this study in our analysis of the association between smoking and epigenetics (Objective 1.1, Manuscript 1).

4. **The TALAVERA** study is a population-based survey performed in an urban area in the province of Toledo, Spain.[235] We included this population to analyze DNA methylation, gene expression and the plasma proteome. We determined the levels of a protein of interest in the plasma of 267 individuals participating in this survey. The data allowed us to perform indirect validation of the

association between smoking and epigenetics and plasma proteome (Objective 1.1, Manuscript 1).

5. **The EPIC-Italy** (European Prospective Investigation into Cancer and nutrition) study is a European collaborative effort to analyze the relationship between nutritional variables and cancer risk in a prospective cohort.[237] EPIC-Italy was used to perform a nested case-control study (cases include 231 individuals with breast cancer and 304 with colorectal cancer) to analyze DNA methylation; EPIC-Italy has also collected data on exposure to air pollution.[242] This study was included as a replication population in our analysis of the association between air pollution and DNA methylation (Objective 1.2, Manuscript 2).

6. **The Framingham Offspring Study** (FOS) is a community-based cohort study.[238] This study was initiated in 1971 and includes the offspring of participants from the initial 1948 cohort. These participants were followed-up and invited to attend to the study clinic every 2-4 years. For participants who attended examination 8 (2005-2008), data are available on DNA methylation pattern and some of the traits of interest for this thesis. This population was used as a replication cohort to address Objectives 2.1 and 2.2 of this thesis (Manuscripts 3 and 4). We obtained these data through the dbGAP (http://dbgap.ncbi.nlm.nih.gov; project number #9047).

7. **The GOLDN** (Genetics of Lipid-Lowering Drugs and Diet Network) study is a family study investigating gene-environment interactions using controlled interventions, as part of PROgram for GENetic Interaction (PROGENI) Network. The aim of this study was to characterize the genetic and epigenetic basis of individual variability in triglycerides in response to three weeks of treatment with fenofibrate.[239,243] This study includes genome-wide data on DNA methylation, expression and genotyping. We used the gene expression data to address the Objective 2.2 (Manuscript 4). The same GOLDN data were released for Genetic Analysis Workshop 20 (GAW20 data), and were used through the GAW20 procedure to address Objective 2.3 (Manuscript 5). This work was conducted during a 6-month stay by the PhD candidate with the GOLDN investigators at the University of Alabama, Birmingham (UAB).

8. Aggregated data from several genome wide association studies were used to test for association between SNPs located in genes of interest and blood lipids (Global Lipid Genetics Consortium), CAD (CARDIoGRAMplusC4D), and body mass index (GIANT).[107,240,241]

The methods used in this thesis adhere to the principles of the Declaration of Helsinki and current legislation in the corresponding countries, and were approved by the local clinical research ethics committees.

3.3 EXPOSURES / PHENOTYPES OF INTEREST

In the REGICOR, BASICMAR, PREDIMED, TALAVERA, EPIC-Italy and GOLDN/GAW20 studies, trained teams performed examinations and administered standardized methods and questionnaires to collect biological samples, sociodemographic and lifestyle data, and other information related to cardiovascular risk factors.[244] For FOS, the relevant information was obtained via dbGAP.

3.3.1 Smoking

Smoking was self-reported and collected using the standardized MONICA questionnaire. Four categories were defined: never smokers, former smokers >5 years, former smokers 1-5 years, and current smokers.[244,245]

3.3.2 Air pollution

REGICOR and EPIC-Italy followed the ESCAPE protocol to evaluate long-term exposure to air pollution.[246,247] All addresses were geocoded at front-door level and we used land use regression (LUR) models to estimate concentrations of nitrogen oxides (NOX), nitrogen dioxide (NO_2), and exposure to particulate matter [aerodynamic diameter of <10μm (PM10), <2.5μm (PM2.5), and PMcoarse (the difference between PM10 and PM2.5)] for each participant.[248,249] Moreover, we estimated traffic intensity in the nearest street and traffic load (sum of traffic intensity multiplied by length of road segment) for all segments in a 100-meter buffer zone, and derived 10-year average values for each participant.[244]

3.3.3 Obesity traits

The weight and height of participants from REGICOR and FOS were measured barefoot and in underwear by nurses using a precision scale of easy calibration. Waist circumference was measured at the umbilicus using standard measuring tape. In FOS, these measures were rounded to the nearest pound and down to the next ¼ inch, respectively, and then converted to SI units (kg and cm). In both studies BMI was calculated as weight divided by the square height (kg/m^2).

3.3.4 Lipid profile

In REGICOR and FOS, TC and TG concentrations were determined enzymatically from serum (Horiba ABX, Montpellier, France); HDL-C concentrations were measured as soluble cholesterol determined from serum using an accelerator-selective detergent method (Horiba ABX). LDL-C was estimated using the Friedewald equation when TG levels were lower than 300 mg/dl.

3.3.5 HDL functionality

All HDL functionality experiments were performed in apolipoprotein-B depleted plasma (ABDP). THP-1 monocyte-derived macrophages were incubated with 0.2 µCI/mL of 1,2-^{3}H(N)-choleseterol (Perkin-Elmer, Waltham, MA, USA), washed, and cultured in the presence or absence (control) of 5% ABDP from the participants. The culture supernatants were obtained and the cell culture lipids extracted. We quantified the radioactivity signal in both the supernatant medium and cell lipids using a beta scintillation Tri-Card 2800TR counter (Perkin-Elmer, Waltham, MA, USA), and the cholesterol-efflux capacity was calculated using the equation e:

e) $$\text{Cholesterol efflux capacity} = \frac{\frac{\text{radioactivity in supernatant}}{(\text{radioactivity in supernatant+cells})} * 100}{\text{Efflux value of the pool}}$$

The other property, HDL anti-oxidant capacity, was measured according to the "HDL inflammatory index" (HII) technique.[250] We measured the capacity of participants' HDL to avoid the oxidative modification of 2'-7'-dichlorodihydrofluorescein diacetate (H2DCF-DA, Life Technologies, Thermo Fisher Scientific, MA, USA) in the presence of oxidized LDLs. Both properties were normalized according to equation f:

f) $$Z = \frac{(X - \bar{X})}{\sqrt{\frac{\sum (X - \bar{X})^2}{(n-1)}}}$$

Where: Z=standardized cholesterol efflux capacity or standardized HDL inflammatory Index, X= cholesterol efflux capacity or HDL inflammatory index for a specific individual, $\bar{X}$= mean cholesterol efflux capacity or mean HDL inflammatory index, and n= sample size.

3.4 DNA METHYLATION ANALYSES

3.4.1 DNA methylome

REGICOR and BASICMAR extracted DNA from peripheral whole blood and EPIC-Italy and FOS extracted DNA from buffy coats. All used standardized methods (Puregen TM; Gentra Systems). GOLDN/GAW20 extracted DNA from CD4+ T-cells. All cohorts assessed methylation levels using the HumanMethylation 450 Beadchip Kit (Illumina, CA, USA) according to the Illumina Infinium HD Methylation protocol. Quality control of DNA methylation data was performed in all cohorts, with each cohort using their own well-defined pipeline.

The analyses were carried out using methylation as β-value and M-value in Manuscript 1 for the REGICOR and BASICMAR cohorts. GOLDN/GAW20 data also used β-value for the analyses in Manuscripts 4 and 5. To reduce batch effects, we developed a standardization process for the β- and M-values (using the equation f; where Z= standardized β-value or standardized M-value, X= β-value or M-value for a specific individual, $\bar{X}$= mean of β-value or mean of M-value, and n= sample size). We used this M-value standardization for the analyses in Manuscripts 2-4 and 6 (**Table 2**).

One of the limitations of the HumanMethylation450 Beadchip is the non-biological variation caused by using different techniques, namely the Infinium I and Infinium II assays in the chip.[222] We used raw data in the analyses to interrogate a single methylation position (Manuscripts 1-4, and 6) because a specific position is always addressed using the same assay (Infinium I or II). However, when we evaluated methylation levels in regions containing different CpG sites addressed with both assays, we normalized the data. We normalized the REGICOR data using dasen[221,223] for the analyses of methylation in regions (Manuscript 1). GOLDN/GAW20 data are normalized using the Combat R package[251] (Manuscripts 4 and 5).

3.5 OTHER OMICS DATA

While, this thesis mainly focuses on DNA methylation using the HumanMethylation450 Beadchip array (Illumina, CA, USA), we also performed analyses using other omics data sources, such as genomics, transcriptomics and proteomics (**Table 1**).

3.5.1 Genome

We downloaded publicly available aggregate GWAS data for lipids, CAD, and BMI. We explored additive associations between genetic variants at several loci of interest and their corresponding phenotypes (Objectives 2.1, 2.2 and 2.4, Manuscripts 3, 4 and 6).

We also used individual genetic data from GAW20 to build polygenic risk scores. These polygenic scores allowed us to use the Mendelian randomization approach to address Objective 2.3 and to explore causality in the association between DNA methylation at specific CpG sites and lipid traits (Objective 2.3, Manuscript 5).

3.5.2 Transcriptome

We used transcriptome data to address Objectives 1.1 and 2.2, Manuscript 1 and Manuscript 4.

In Manuscript 1 we used data from the PREDIMED study. Gene expression was evaluated using a whole transcriptome microarray (Affymetrix Gene Chip Human Genome U133A 2.0). These data were used to study the association between smoking and gene expression, focusing on loci that showed differential methylation related to smoking exposure.

In Manuscript 4 we used data from the GOLDN study. In this study, gene expression levels were assessed using RNA sequencing. We evaluated the relationship between DNA methylation and gene expression, focusing on loci showing differential methylation related to lipid traits.

3.5.3 Proteome

In Objective 1, Manuscript 1 we assessed the plasma concentration of a candidate protein related to a gene showing different methylation in one locus and differential gene expression in relation to smoking exposure. We used the dot-blot technique in serum from participants in the TALAVERA cohort.

3.6 OTHER COVARIATES

We also estimated other important covariates such as cell count proportion or surrogate variables:

- the cell count proportion was calculated using the Houseman algorithm implemented in the minfi R package.[252]
- surrogate variables capturing unmeasured confounder variables were directly constructed from high-throughput data using the sva R package.[228]

3.7 STATISTICAL ANALYSES

Table 2 shows a summary of the main statistical methods used in the thesis.

To address Objective 1.1 and 1.2, we considered smoking and air pollution as independent variables, and DNA methylation as the dependent variable. When analyzing the association between DNA methylation and intermediate cardiovascular traits (Objective 2.1, 2.2 and 2.4), we used DNA methylation as independent variable and the intermediate cardiovascular traits as dependent variables.

To analyze the association between smoking and DNA methylation we fitted a classical multivariate linear regression model (Manuscript 1). In the posterior analyses we used robust multivariate linear regression model to reduce the effects of potential outliers[225] (Manuscripts 2, 3, 4 and 6). To account for family relatedness in the GOLDN/GAW20 cohort, the analyses were performed using multivariate linear mixed models (Manuscript 4 and 5).

Table 2. Statistical models and types of methylation data used in this thesis.

Study	Objectives (O), Manuscripts (M)					
Cohort Name	**O 1.1 M 1**	**O 1.2 M 2**	**O 2.1 M 3**	**O 2.2 M 4**	**O 2.3 M 5**	**O 2.4 M 6**
REGICOR	MLM	RMLM	RMLM	RMLM	-	RMLM
	M and β value	Z-M value	Z-M value	Z-M value	-	Z-M value
	Raw and dasen data	Raw data	Raw data	Raw data	-	Raw data
BASICMAR	MLM	-	-	-	-	-
	M and β value					
	Raw data					
EPIC-Italy	-	RMLM	-	-	-	-
		Z-M value				
		Raw data				
FOS	-	-	RMLM	RMLM	-	-
			Z-M value	Z-M value		
			Raw data	Raw data		
GOLDN / GAW20	-	-	-	LMM	LMM	-
				β value	β value	
				Combat R data	Combat R data	

Ref: References; MLM: Multivariate Linear Model; RMLM: Robust Multivariate Linear Model; M: M-value; β: β-value; Z-M: Standardized M-value.

All analyses were adjusted for potential confounders in a stepwise manner. In most cases, the analysis was initially adjusted for cell count proportion, then for smoking (except in Manuscript 1), and finally for other potential confounder variables, including the surrogate variables.

The main analyses were an epigenome-wide association analysis of single methylation positions (Manuscripts 1, 2, 3, 4, and 6). As explained in the section on study design, we selected for replication the CpG sites with a p-value $<1 \cdot 10^{-05}$, and then meta-analyzed the results of the discovery and replication stages. We declared as statistically significant the CpG sites with p-values >0.05 after Bonferroni correction (p-value <0.05/number of discovery CpG sites).

For the analyses using the candidate loci approach (Manuscripts 2 and 6) we selected the CpG sites according to previous knowledge, and declared as statistically significant those results that fulfilled Bonferroni or FDR criteria.

All analyses were conducted using R.[253]

4. Results

Table 3 shows a brief summary of the main findings of this thesis. Detailed results are shown in the corresponding scientific manuscripts published in high-impact journals.[3*]

[3*] Supplemental information for each of the manuscripts is available at the journal's websites.

Table 3. Summary of the main results of this thesis.

General Objective	Specific Objective - Manuscript	Association	Main Results
To assess the association between lifestyle and environmental factors and DNA methylation.	Objective 1.1 Manuscript 1	Smoking exposure - DNA methylation	We identified 63 CpG sites that are hypomethylated in current smokers vs never smokers.
			We discovered a new CpG in the *LNX2* gene that shows differential methylation in relation to smoking.
			We provide new functional information to two previously reported associations in the *THSB1* and *MTSS1* genes.
			The effect of smoking on DNA methylation seems to be reversible.
			We identified 7 promoter regions, 8 gene body regions, and 4 non-coding RNA regions associated with smoking exposure.
	Objective 1.2 Manuscript 2	Air pollution - DNA methylation	We did not find any association between air pollution exposure and DNA methylation.
To assess the association between DNA methylation and intermediate cardiovascular traits.	Objective 2.1 Manuscript 3	DNA methylation - Obesity traits	We validated 94 CpG sites in 72 genes and 23 intergenic regions associated with BMI, of which 70 CpG sites were new.
			We validated 49 CpG sites in 37 genes and 12 intergenic regions associated with WC, of which 33 CpG sites were new.
			These CpG sites explain approximately 14% and 17% of the variability of BMI and WC in FOS, respectively.
			tagSNPs in 10 of the genes showing differential methylation in relation to obesity were also associated with BMI in the GIANT consortium data.
			Our pathway analysis found that 4 networks, related to neurologic, psychological, endocrine and metabolic pathways, were associated with obesity.
	Objective 2.2 Manuscript 4	DNA methylation - Serum lipid profile	We validated 14 CpG sites in 9 genes and 2 intergenic regions related to TC, HDL-C and TG.
			We replicated previous CpG associations located in *ABCG1*, *SREBF1* and *CPT1A* and 1 intergenic region.
			We discovered 7 new CpG sites associated with TC (located in *SREBF2*), HDL-C (located in *PHOSPHO1*, *SYNGAP1* and an intergenic region on chromosome 2) and TG (*MYLIP*, *TXNIP* and *SLC7A11*).
			These CpG sites explained approximately 1%, 10% and 20% of TC, HDL-C and TG in FOS, respectively.

			Methylation levels of cg16000331 and cg11024682 were inversely associated with expression of the *SREBF2* and *SREBF1* genes, respectively. *SREBF1* expression levels were directly associated with higher levels of HDL-C.
			Genetic variants in *SREBF1*, *PHOSPHO1*, *ABCG1* and *CPT1A* were associated with lipid traits in the GLGC consortium data.
	Objective 2.3 Manuscript 5	Causal relationship between DNA methylation and lipid traits	Methylation levels of cg00574958 (*CPT1A*) and cg06690548 (*SLC7A11*) can be affected by TG.
			Methylation levels of cg00574958 (*CPT1A*) could modify the levels of TG.
	Objective 2.4 Manuscript 6	DNA methylation - HDL functionality	Using an EWAS approach, we discovered 3 CpG sites in *HOXA3*, *PEX5*, *PER3* that are associated with cholesterol efflux capacity.
			Also using EWAS, we discovered a CpG site located in *GABRR1* that is related to HDL inflammatory index.
			Using a candidate gene approach, we found 2 CpG sites located in *CMIP* that are related to cholesterol efflux capacity.
			Genetic variants in *PER3*, *HOXA3* and *GABRR1* showed a possible association with CAD in data from the CARDIoGRAMplusC4D consortium.

4.1 MANUSCRIPT 1

Sayols-Baixeras S, Lluís-Ganella C, Subirana I, Salas LA, Vilahur N, Corella D, Muñoz D, Segura A, Jimenez-Conde J, Moran S, Soriano-Tárraga C, Roquer J, Lopez-Farré A, Marrugat J, Fitó M, Elosua R. **Identification of a new locus and validation of previous reported loci showing differential methylation associated with smoking. The REGICOR study.** Epigenetics. 2015; 10(12): 1156-1165.

Sayols-Baixeras S, Lluís-Ganella C, Subirana I, Salas LA, Vilahur N, Corella D, et al. Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. Epigenetics. 2015;10(12):1156–65. DOI: 10.1080/15592294.2015.1115175

4.2 MANUSCRIPT 2

Sayols-Baixeras S, Fernández-Sanlés A, Prats A, Lluís-Ganella C, Subirana I, EPIC-ITALY, ISGLOBAL 1-2, Marrugat J, Basagaña X, Elosua R. **Association between long-term air pollution exposure and DNA methylation: the REGICOR study.** bioRxiv. 2018. doi: https://doi.org/10.1101/404483

Sayols-Baixeras S, Fernández-Sanlés A, Prats-Uribe A, Subirana I, Plusquin M, Künzli N, et al. Association between long-term air pollution exposure and DNA methylation: The REGICOR study. Environ Res. 2019 Sep;176:108550. DOI: 10.1016/j.envres.2019.108550

4.3 MANUSCRIPT 3

Sayols-Baixeras S, Subirana I, Fernández-Sanlés A, Sentí M, Lluís-Ganella C, Marrugat J, Elosua R. **DNA methylation and obesity traits: An epigenome-wide association study. The REGICOR study.** Epigenetics. 2017; 12(10): 909-916.

Sayols-Baixeras S, Subirana I, Fernández-Sanlés A, Sentí M, Lluís-Ganella C, Marrugat J, et al. DNA methylation and obesity traits: An epigenome-wide association study. The REGICOR study. Epigenetics. 2017;12(10):909–16. DOI: 10.1080/15592294.2017.1363951

4.4 MANUSCRIPT 4

Sayols-Baixeras S, Subirana I, Lluís-Ganella C, Civeira F, Roquer J, Do AN, Absher D, Cenarro A, Muñoz D, Soriano-Tárraga C, Jiménez-Conde J, Ordovas JM, Senti M, Aslibekyan S, Marrugat J, Arnett DK, Elosua R. **Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study.** Hum Mol Genet. 2016; 25(20): 4556-4565.

Sayols-Baixeras S, Subirana I, Lluis-Ganella C, Civeira F, Roquer J, Do AN, et al. Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. Hum Mol Genet. 2016;25(20):4556–65. DOI: 10.1093/hmg/ddw285

4.5 MANUSCRIPT 5

Sayols-Baixeras S, Tiwari HK, Aslibekyan SW. **Disentangling associations between DNA methylation and blood lipids: A Mendelian randomization approach.** BMC Proceedings. 2018; 12 (Suppl 9) :23

Sayols-Baixeras S, Tiwari HK, Aslibekyan SW. Disentangling associations between DNA methylation and blood lipids: a Mendelian randomization approach. BMC Proc. 2018 Sep 17;12(Suppl 9):23. DOI: 10.1186/s12919-018-0119-8

4.6 MANUSCRIPT 6

Sayols-Baixeras S*, Hernáez A*, Subirana I, Lluís-Ganella C, Muñoz D, Fitó M, Marrugat J, Elosua R. **DNA Methylation and High-Density Lipoprotein Functionality –Brief Report. The REGICOR Study.** ATVB. 2017; 37(3): 567-569.

* These authors contributed equally to this manuscript.

Sayols-Baixeras S, Hernáez A, Subirana I, Lluis-Ganella C, Muñoz D, Fitó M, et al. DNA Methylation and High-Density Lipoprotein Functionality-Brief Report: The REGICOR Study (Registre Gironi del Cor). Arterioscler Thromb Vasc Biol. 2017 Mar;37(3):567–9. DOI: 10.1161/ATVBAHA.116.308831

5. Discussion

Epigenetics is a heritable phenomenon that could explain part of the missing heritability of CAD. It has also been proposed as one of the mechanisms that mediates the effect of lifestyles and environmental factors on risk of complex disease. This thesis addresses some of these questions, focusing on: i) the relationship between smoking and air pollution, and DNA methylation; and, ii) the association between DNA methylation and several intermediate cardiovascular traits (obesity, lipid traits and HDL functionality). In addition, to identify new potential therapeutic targets, it is essential to explore the causality of these associations. In this thesis we also evaluate the causality of the relationships we identified between DNA methylation and serum blood lipids.

5.1 SMOKING, AIR POLLUTION, AND DNA METHYLATION

In this work, we replicate previously reported CpG sites associated with smoking, and we identify a new locus associated with smoking (Manuscript 1). We do not identify CpG sites associated with exposure to air pollution (Manuscript 2).

5.1.1 Smoking and DNA methylation

We identified 66 CpG sites showing a differential methylation signature related to smoking exposure, most of which (n=63) were hypomethylated in current smokers with respect to never smokers. This pattern of smoking-associated hypomethylation has been reported in previous studies,[226,254–257] and was also observed in a recent meta-analyses from the CHARGE consortium.[258]

Our top hit, cg05575921 (*AHRR*), was also the CpG that showed the greatest magnitude of association with smoking in most studies. This gene participates in the aryl hydrocarbon receptor signaling cascade, and in regulating cell growth and differentiation.[259] Methylation at this locus could represent both a pathogenic pathway and a biomarker of smoking exposure:

i) A recent study reported an association between hypomethylation of *AHRR* and increased risk of lung cancer, suggesting that methylation at this locus may mediate the carcinogenic effects of smoking exposure[260]
ii) Methylation levels at cg05575921 have been found to accurately discriminate between current smokers and never smokers (area under the curve (AUC) >0.926), and are thus an excellent biomarker of active smoking exposure[261,262]

We also validated two recently observed associations between smoking and CpG sites in the *MSS1* and *THBS1* genes. We further analyzed the functional impact

of smoking on the *THBS1* locus and reported that current smoking was associated with reduced methylation of *THBS1*, and increased expression of *THBS1*, although paradoxically smokers had lower plasma levels of THBS-1 protein. We proposed that this is due to post-transcriptional regulation of THBS-1 protein levels, or complex feedback regulation mechanism of the *THBS1* gene. A decrease in THBS-1 protein levels caused by smoking could be compensated for by hypomethylation of the *THBS1* locus, leading to increased expression of the gene and an increase in concentration of the protein.

As a novel result, we identified and validated the association between smoking and a CpG site located in the *LNX2* gene. *LNX2* encodes the protein ligand of numb-protein X2, an E3 ubiquitin ligase giving specificity to the ubiquitination process by selectively binding substrates. This protein is upregulated during osteoclast differentiation, and has also been associated with adenocarcinoma, epithelial neoplasia, and preeclampsia.[263,264] This result, similar to that for *AHRR*, could indicate a pathogenic mechanism mediating carcinogenic and other deleterious health effect of smoking.

We also identified new genomic regions located in seven genes (*AGAP2-AS1*, *IGHJ3P*, *IGHJ5*, *IGHJ6*, *CLEC16A*, *ZNRF1* and *PRF1*) and in four non-coding RNAs that show differential methylation in relation to smoking, as well as we replicated some of the previously reported loci. These results are interesting because they highlight modifications not only in a specific CpG position, but also at the region level. While we did not validate the results for these genomic regions in a distinct population, several were replicated in other studies after the publication of Manuscript 1.[265,266]

5.1.2 Smoking and DNA methylation: time-dependent association

To analyze the time-dependent association between smoking and DNA methylation we defined four smoking categories: current, former (1-5 years, and >5 years) and never smokers, and analyzed the pattern of association with the 66 identified CpG sites. We also reported a time-dependent relationship between smoking and the methylation signature at the 66 identified CpG sites. These patterns have also been reported in other studies,[254,256,267] suggesting that the relationship between smoking and methylation is reversible.

5.1.3 Causal association between smoking and DNA methylation

Although the cross-sectional design of our study prevents us from disentangling the causal effects of smoking on DNA methylation, there is some evidence smoking modifies DNA methylation rather than *vice versa*:

i) The time-dependent reversible effect.
ii) The experimental evidence in animals and *in vitro* cultures.[265]
iii) The results of cord blood methylation analysis from newborns and maternal smoking in pregnancy.[268,269]
iv) The results from a causal inference study using the method *Inference about Causation through Examination of Familial Confounding* (ICE FALCON), a regression-based methodology for twins. This study suggests that smoking has a causal effect on peripheral blood DNA methylation, but not vice versa.[269]

5.1.4 Air pollution and DNA methylation

In this thesis we also used an epigenome-wide association approach to investigate the association between DNA methylation and air pollution. We did not identify any CpG site showing differential methylation related to air pollution exposure.

Previous studies[270–278] analyzed the association between air pollution and DNA methylation, two of which identified and validated differential methylation signatures related to this environmental factor.[274,277]

We selected these CpG sites as candidate loci to be replicated in our cohort, but we did not find any statistically significant association. Our top hit was associated with $PM_{2.5}$, and was located in an intergenic region on chromosome 1, close to the *CDKN2C* gene (cg10893043, p-value=$6.79 \cdot 10^{-5}$). Expression of this gene was shown to suppress the growth of human cells in animal models, and could have a role in tumorigenesis.[279]

The absence of positive results in our study could be due to several factors:

i) Limited power to detect associations with a small effect size. However, we estimated that our statistical power was similar to that of previous studies.[274,277]

ii) The variability in exposure to air pollution in the REGICOR population is lower than that observed in other studies, which reduces the probability of finding real associations in our study.

iii) In addition, we studied long-term exposure (5 year exposure) whereas other studies have analyzed short-term exposures (2 days to 28 days).[274,277] Nevertheless, a study that analyzed different exposure time points (2 days, 7 days and 28 days) observed that the number of CpG sites associated with air pollution increased with increasing exposure time.[277]

As a summary of objectives 1.1 and 1.2:

- We have demonstrated a reversible time-dependent relationship between smoking and DNA methylation, which validates several previously reported loci plus a new one.

- Although we did not find any statistically significant associations between air pollution and DNA methylation, other studies have reported this relationship. New joint efforts to analyze this association are warranted to increase the sample size, and therefore increase the statistical power and clarify these discrepancies.

5.2 DNA METHYLATION AND INTERMEDIATE CARDIOVASCULAR TRAITS

In this thesis we found different associations between DNA methylation at specific loci and some of the main cardiovascular risk factors (obesity and serum blood lipids), and the two main functions of the HDL particles (cholesterol efflux capacity and anti-inflammatory capacity) (Manuscripts 3, 4 and 6).

5.2.1 DNA methylation and obesity traits

Objective 3 was to assess the relationship between DNA methylation and obesity traits. We identified 94 CpG sites associated with obesity, 70 of which were novel discoveries. All 94 reported CpG sites were associated with BMI and 49 were also associated with waist circumference. These 94 CpG sites explained 14.18% and 16.73% of total variability in BMI and waist circumference in the FOS study, respectively. These percentages are surprisingly high when compared to the variability explained by common genetic variants. Genome-wide estimations suggest that genetic variants should explain >20% of the variation in BMI, but the 97 SNPs described to date explain only around 2.7% of this variation.[107]

A meta-analysis by the CHARGE consortium investigating this relationship also observed that a large number of CpG sites throughout the genome are related to obesity.[280] It is interesting that only a limited number of CpG sites overlapped between the CHARGE study and ours, especially considering that we replicated more CpG sites from previous studies than from the CHARGE consortium.[239,280–290] This lack of replicability with respect to the CHARGE meta-analysis may be partly due to heterogeneity in measuring the phenotypes included in this meta-analysis, in which the data from several studies (using different measurement methodologies) are pooled.[291]

In our results, it is important to remark that ten of the loci identified (*INTU*, *SFRP2*, *RUNX2*, *CTTNBP2, CPT1A, CACNA1C, MCF2L, BCL2L2, MPG* and *KRT16*) also have tagSNPs associated with obesity. This reinforces the importance of these loci for obesity, and suggests a possible direct causal relationship between these loci and the obesity trait.

Note that some of the CpG sites associated with obesity are located in genes with a plausible clinical or functional impact (*CUX1*, *DDAH2*, *CACNA1C*, *PDE1A*, *CLOCK*). *CUX1* has been proposed as a regulator of *FTO* and *RPGR1P1L,* both of which are important loci for obesity.[292] The *DDAH2* gene inhibits nitric oxide synthase, altering the process of vasodilation, respiration, cell migration, immune response and apoptosis.[293,294] Finally, *CACNA1C* and *PDE1A* are involved in intracellular calcium balance, and *CLOCK* is associated with circadian rhythms.[295–298]

We also highlight that some of the loci associated with obesity show differential methylation related to lipid traits (*ABCG1*, *SREBF1*, *SYNGAP1*, *SLC7A11*, *MYLIP*, and *CPT1A*) (we explained these genes in more extension below in the subsection DNA methylation and lipid traits). These results highlight the close relationship between adiposity and lipid metabolism.

The loci that show differential methylation related to obesity traits indicate four main enriched networks that highlight a link between obesity and endocrine and metabolic diseases (diabetes mellitus, metabolic syndrome, hypercholesterolemia, non-alcoholic liver disease), and psychological and neurologic processes (depression, dementia, Alzheimer, schizophrenia, compulsive gambling, bulimia nervosa). These results seem to indicate different pathogenic pathways in the origin of obesity, one more metabolic and other more psychological. Further studies should focus on the value of this epigenetic signature to reclassify obesity into subcategories according to the relevance of these two pathways.

5.2.2 DNA methylation and lipid traits

We identified 14 CpG sites in 9 genes (*SREBF1*, *SREBF2*, *PHOSPHO1*, *SINGAP1*, *ABCG1*, *CPT1A*, *MYLIP*, *TXNIP* and *SLC7A11*) and 2 intergenic regions showing differential methylation associated with at least 1 of the lipid traits of interest: total cholesterol (TC), high density lipoprotein cholesterol (HDL-C), and triglycerides (TG). These CpG sites explained approximately 0.7%, 9.5% and 18.9% of the variability in TC, HDL-C and TG in FOS, respectively. This variability is similar to that explained by common genetic variants, which is 10-12% of the variance in these lipid traits.[240]

Our results replicate previous hits located in three interesting genes, *ABCG1*, *SREBF1* and *CPT1A:*

i) *ABCG1* encodes a transporter involved in macrophage cholesterol and phospholipid transport, and may regulate cellular lipid homeostasis in

other cell types.[299] Higher methylation of this gene has been associated with CAD and aging.[300–302]

ii) *SREBF1* activates the synthesis of fatty acids and the synthesis and uptake of cholesterol.[303,304] We also found an inverse association between methylation in *SREBF1* and its gene expression. The expression of this gene was directly associated with HDL-C levels.

iii) *CPT1A* is a key enzyme in the mitochondrial oxidation of long-chain fatty acids. *CPT1A* deficiency results in a decreased rate of fatty acid beta-oxidation.[239] Methylation levels in this gene have also been associated with BMI, diabetes and plasma adiponectin.[239,282,305,306]

Our work also provides novel findings of association between lipids and methylation levels in *SREBF2*, *PHOSPHO1*, *SYNGAP1*, *TXNIP*, *MYLIP* and *SLC7A11:*

i) *SREBF2* is similar to *SREBF1* and regulates the metabolism of cholesterol and fatty acids.[307] This gene has been related to intima-media thickness through a common genetic variant. We also described an inverse association between methylation levels in this CpG site and *SREBF2* expression, suggesting a potential role in atherosclerosis progression through the regulation of lipid metabolism.

ii) *PHOSPHO1* regulates the vascular smooth muscle calcification.[308,309] These results and previous reports allows us to hypothesize that lower methylation levels is associated with higher *PHOSPHO1* levels, inducing calcification due to lower HDL-C levels and altered glucose homeostasis or other potential mechanisms.

iii) *TXNIP* is involved in redox homeostasis and regulates glucose homeostasis.[310–312] Methylation of *TXNIP* has been associated with diabetes.[305,313–315] We demonstrated a relation between *TXNIP* methylation and TG, independent of diabetes.

iv) Interestingly, we observed an association between *MYLIP* methylation and TG levels. *MYLIP* is an ubiquitin ligase that regulates the LDL receptor.[316] The regulatory mechanism seems similar to that of *PCSK9* in that induction of *MYLIP* leads to the ubiquitination of LDL receptors.[316,317] Thus, RNA silencing downregulates *MYLIP*, which upregulates LDL receptors and enhances LDL uptake.[316] Our results support the notion that increased *MYLIP* methylation decreases *MYLIP* expression and increases triglycerides by decreasing the availability of the LDL receptor, ultimately increasing risk of atherosclerosis.

v) *SYNGAP1* encodes a Ras GTPase activating a protein member of the N-methyl-D-aspartate receptor complex. It regulates synaptic plasticity and neuronal homeostasis. Common genetic variants have been associated with intellectual disability and autism spectrum disorder.[318]

vi) Finally, *SLC7A11* has an important role in protecting cells from oxidative stress and in glutamine metabolism.[319,320] It also regulates the glucose and glutamine dependency in cancer cells.[320]

5.2.3 DNA methylation and HDL functionality

Clinically important blood lipids (LDL and HDL cholesterol, and triglycerides) have an important impact in CAD risk. However, while experimental and Mendelian randomization studies support the causal association between CAD and LDL-cholesterol and TG, these same approaches have questioned the causal relation between CAD and HDL-C levels.[65,321] Some authors suggest that the functionality of HDL rather than its concentration might better explain its protective properties.

In this thesis, we carried out the first study of DNA methylation in relation to the two main functions of the HDL particle, cholesterol efflux capacity, and HDL inflammatory index. We discovered:

i) Five CpG sites associated with cholesterol efflux capacity
   a. Using an EWAS approach, three CpG sites located in three genes
      - *HOXA3*, its expression is associated with abdominal adiposity[322]
      - *PER3*, its expression is altered in atherosclerotic plaques[323]
      - *PEX5*, this is an important element for peroxisome formation[324]
   b. Using a candidate approach, two CpG sites located in the *CMIP* gene[110,240], polymorphisms in the gene are associated with HDL-C levels.

ii) Using an EWAS approach, a CpG site associated with HDL inflammatory index. This CpG site is located in *GABRR1*, which is the main inhibitory neurotransmitter in the central nervous system.[325]

Interestingly, hypomethylation of these loci is associated with better HDL functionality, and could explain the protective role of HDL in arteriosclerotic disease.

Analyzing aggregated data from the CARDIoGRAMplusC4D consortium, we also observed an association between CAD and genetic variants in *PER3*, *HOXA3* and *GABRR1*. These results may suggest that these genes have a causal role in CAD that could be mediated by the functionality of HDL particles.

Our study provides the first evidence to suggest a relationship between HDL functionality and methylation patterns. However, we were unable to replicate these

findings as we were unable to obtain any independent populations with suitable data to confirm these results.

As a summary of objectives 2.1, 2.2 and 2.4:

- We found a significant number of CpG sites associated with obesity in genes that are important for four networks related to endocrine and metabolic diseases and psychological and neurological processes.

- We confirmed previous results for known loci associated with blood lipids, and identified new loci, *PHOSPHO1*, *SYNGAP1*, *TXNIP*, *MYLIP* and *SLC7A11,* showing differential methylation related to blood lipids. In particular, *MYLIP* seems to be a good therapeutic candidate because it has similar regulation of LDL receptors to *PCSK9*.

- Finally, we report evidence of the relationship between DNA methylation and HDL functionality, and highlight the fact that hypomethylation in *HOXA3*, *PER3*, *PEX5*, *CIMP* and *GABRR1* is associated with improved HDL functional capacity.

5.3 CAUSAL INFERENCE OF THESE RELATIONSHIPS

5.3.1 From DNA methylation to lipid traits, or vice versa, or both?

In this thesis we also analyzed the causality of the reported associations between DNA methylation and lipid traits (Manuscript 5). We found evidences that the causal relationship could be bidirectional: DNA methylation modulates lipid levels, lipid levels have a causal effect on DNA methylation, and both effects may occur simultaneously. For example, we found that DNA methylation regulates TG levels and that TG levels can also modify the methylation pattern.

While the cross-sectional design of the analyses performed in this thesis did not allow us to infer causality, we were able to examine causality using Mendelian randomization. We analyzed the results of the associations we observed between DNA methylation and lipid traits (Manuscript 4) in data from the publicly available GAW20 study, which includes data on genome wide genotypes and methylation, and on lipid traits.

In the GAW20 database we replicated 5 CpG sites associated with TG, but none of those associated with HDL-C. We applied the Mendelian randomization using a bidireccional approach to these 5 CpG sites. We implemented an elastic-net strategy (a penalized regression model which shrinks coefficients and sets some coefficients to zero) to select the genetic variants related to the methylation locus of interest and to TG, in order to construct our genetic instrumental variables (Section 1.2.5.2). This approach prevents the inclusion of genetic variants that are in linkage disequilibrium. In addition, instead of adding the variants individually, we created a polygenic score for each methylation locus and also for TG. Our elastic-net method allowed us to assess the causal relationship between methylation and TG in two of the five selected loci. The results of our analyses were as follows:

i. We observed a reversal causal effect from TG to DNA methylation in two loci, cg06690548 and cg00574958, located in *SLC7A11* and *CPT1A*, respectively. That is, TG levels modulate DNA methylation levels at these two loci.
ii. We were unable to find a genetic instrument related to the methylation level at cg06690548, because no SNPs were associated with this methylation locus, to assess its causal effect on TG levels.
iii. We report that the methylation level of the *CPT1A* locus (cg00574958) is causally associated with TG.

These results are partially concordant with those of the largest published Mendelian randomization study assessing the relationship between DNA methylation and blood lipids.[128] This study supported a causal relationship from methylation levels of *CPT1A* to TG, but ruled out the reverse causation that we observed.

Dekkers *et al.* used as a genetic instrument all cis-mQTLs (methylation quantitative trait loci, SNPs that are associated with a CpG site) associated with the methylation loci.[128] We believe this approach could introduce significant noise because some SNPs are likely to be in linkage disequilibrium; this could be avoided with the elastic-net approach we used. Dekkers *et al*. also created a polygenic score to assess the causal association from lipids to methylation, but did not create this type of instrument variable to study the association from methylation to lipids.[128]

Dekkers *et al*.[128] have also reported a causal association from TG to DNA methylation at two loci: cg11024682 – *SREBF1* – and cg06500161 – *ABCG1* –. The CHARGE consortium also provided evidence that methylation level at the *SREBF1* locus (cg11024682) is causal associated with obesity, adiposity and CAD.[326] In this context, we can propose a causal mechanism where an increase in the levels of TG causes an increase in the methylation levels at the *SREBF1* locus, increasing in turn BMI (obesity) and adiposity, and finally leading to higher risk of CAD.

5.3.2 From smoking and air pollution to DNA methylation?

Regarding the association between lifestyle and environmental factors and DNA methylation, above (Section 5.1) we described the indirect evidence supporting the causal association between smoking exposure. If we consider the Bradford Hill criteria,[327] we see that a causal effect is supported by various pieces of evidence such as the strength of the association, its consistency across populations and studies, its temporality (time-dependent association), dose-response gradient, and biological plausibility. We were unable to analyze the causal association between air pollution and DNA methylation due to the lack of positive results in our cohort, and the lack of an instrumental genetic variable related to air pollution.

5.3.3 From DNA methylation to obesity and HDL functionality?

Although we did not assess the causal relationship between DNA methylation and obesity in this thesis, other authors have analyzed the direction of causality of these associations.

Elliot and coauthors[328] assessed the causal relationship between DNA methylation and diabetes, and found that half of the SNPs associated with diabetes can modify methylation levels and could be on a causal pathway to future disease or could be non-causal biomarkers. These authors highlighted one locus on *KCNQ1* that could be in the causal pathway to diabetes.[328]

Since there are no GWAS studies assessing the two main HDL functions, it was not possible to build a genetic instrumental variable, so we were unable to explore the causal association between DNA methylation and HDL functionality.

As a summary of objective 2.3:

- To infer causality using Mendelian randomization, we implemented an elastic-net method to improve the selection of SNPs to enhance the genetic instrument.

- We updated the causal relationship between TG and methylation levels of cg00574958 (located in *CPT1A*). We replicated the previous observation that TG modifies methylation levels at *CPT1A*. We also observed that methylation at that locus can change levels of blood TG, which demonstrates the complexity of lipid homeostasis.

- We observed that TG can modify methylation at *SLC7A11*. However, we were unable to build a good instrument to rule out a causal relationship from methylation levels at this locus to TG.

5.4 STRENGTHS AND LIMITATIONS

The strengths of this work can be summarized in the following highlights:

i) The standardized quality control and analysis pipelines developed and used for DNA methylation studies. The methodology used for the epigenome-wide association analyses has been demonstrated to be powerful and very accurate.
As novel approaches, we standardized the M-values to remove non-biological sources of variation, we removed outlier values that could affect the results, and we used robust multivariate models to reduce even further the potential effect of outliers. Similar to genome-wide studies, we divided our EWAS into two phases (discovery and replication) to minimize type I errors in our observations and findings.

ii) The sample selected for this thesis were large, had very well-defined phenotypes, especially in the cardiovascular area, and the same technology was used to determine DNA methylation in all cases, namely the Illumina 450k array.

iii) The integrated approach used to address some objectives, including additional sources of information from other omics data, such as genetic variants, gene expression and proteomics data, as well as the information obtained from pathway databases (Ingenuity Pathways Analysis).

iv) Following publication of the manuscripts presented in this thesis, most of the relevant discoveries have been replicated and supported by other studies.[258,329,330] The consistency of our results increases their validity and reduces the likelihood that they are false positive results.

This project has several limitations:

i) Methylation markers have been studied in DNA obtained from different sources: whole blood, buffy coat and CD4+ T-cells. Methylation levels in some CpG sites/regions are tissue or cell-specific, so we could miss some signals because we may not have studied the tissue where the phenotypes have the most direct impact.[224,331] However, whole blood is easy and minimally invasive to obtain, and has been proposed as a good proxy for methylation levels at specific sites.[225,226] Although there is an inter-individual variability in the cell composition of whole blood, various algorithms can be used to estimate blood cell count, and thereby to take into account this variability in blood cell type.[252] Differences in the source tissues used assess methylation levels (e.g. CD4+ T-cells vs whole blood)

could explain some of the discrepancies observed in the association between DNA methylation and blood lipids.

ii) In some of the hypotheses tested in this thesis we did not have objective quantitative measurements of the exposure of interest, principally smoking and air pollution. Smoking exposure was self-reported and collected using questionnaires, and we had no data on passive smoking exposure. However, self-report questionnaires have been shown to be a valid instrument, and this is the classical method used in epidemiological studies.[244,245] Regarding air pollution, this exposure was assessed using LUR models, the methodology used in ESCAPE, a European project that has proposed a standardized methodology for estimating exposure to air pollution. The ESCAPE protocol was validated in REGICOR cohort.[249] However there may still be some misclassification of air pollution exposure, which could affect the results.

iii) We detected some heterogeneity between the results observed in the discovery and validation populations. We tried to address the differences by performing a meta-analysis of the observed results. It will be interesting to incorporate other populations to address this heterogeneity and assess the consistency of the results across populations.

iv) In most of the manuscripts (1, 3, 4 and 6) we had large sample-sizes and good statistical power to identify and validate associations between DNA methylation and exposures or phenotypes. However, the analysis of the association between air pollution and DNA methylation had low power due to the small sample size of the replication sample (Manuscript 2).

v) The cross-sectional design used in this thesis has made it more difficult to infer causality. Therefore, we took advantage of genomic variations to disentangle the causal relationships in the associations using the Mendelian randomization approach.[333]

vi) In Manuscript 6 we were not able to replicate the findings observed in the discovery study due to the lack of replication cohorts, mainly because of the great novelty of the research question.

6. Future perspectives

Cardiovascular disease is a complex disease caused by the interplay between genetic, environmental and lifestyle factors. In this thesis we have analyzed the association between some environmental and lifestyles factors (air pollution and smoking exposure), and DNA methylation, and also between DNA methylation and some intermediate cardiovascular traits (obesity, blood lipids and HDL functions). However, other relationships remain to be evaluated:

i) Physical activity and DNA methylation. We are co-leading a project in the CHARGE consortium to explore the association between physical activity and DNA methylation and epigenetic age. We are fitting multivariate linear models in which physical activity is the independent variable and DNA methylation, measured as β-value or epigenetic age, is the dependent variable. This project will include 8 cohorts (Multi-Ethnic Study of Atherosclerosis –MESA–, Coronary Artery Risk Development in Young Adults –CARDIA–, NES, Women's Health Initiative –WHI–, Framingham Heart Study –FHS–, REgistre GIroni del COR –REGICOR–, Genetics Of Lipid Lowering Drugs and Diet –GOLDN– and the Rotterdam study).

In addition to our contribution to the CHARGE consortium, to further explore this association we have designed a two-stage EWAS using REGICOR as the discovery study and FOS as a replication study. We are implementing different methods to solve the non-symmetrical distribution of physical activity with inflation at zero.[333]

ii) Diet and DNA methylation. We are also implementing a two-stage EWAS to explore the association between dietary patterns and DNA methylation.

iii) Sex and DNA methylation. We have identified and validated a group of 3241 CpG sites that have different methylation levels in men and women, most being hypomethylated in women. We have used an elastic-net approach to build an epigenetic score of 50 relevant CpG sites, allowing us to discriminate between men and women (AUC=0.99) and to transform the categorical variable sex into a continuous trait. We will explore the association between this epigenetic sex score and several diseases that have different prevalence in men and women, with the aim of explaining these sex-related differences. Currently, we are analyzing the relationship between this epigenetic sex score and CAD in several cohorts (REGICOR, FOS, ESTHER– Estrogen and Thromboembolism Risk–, KORA– Cooperative Health Research in the Augsburg Region–, MESA– Multi-Ethnic Study of Atherosclerosis – and WHI –Women Health Initiative–).

iv) DNA methylation and other intermediate cardiovascular traits (diabetes, blood pressure, subclinical atherosclerosis). We will use the methodology described in this thesis to expand our current research to other cardiovascular risk factors and markers.

v) DNA methylation and clinical cardiovascular disease. A clear extension of this work will be to explore the association between DNA methylation and CAD or other cardiovascular outcomes. This work is ongoing in the REGICOR study.

vi) The results of this thesis highlight some mechanisms that could help us to better understanding the pathogenesis of cardiovascular disease. We used Mendelian randomization to infer causality in the observed association. Disentangling causality is essential for selecting candidate therapeutic targets. In this sense, several studies are planned:

a. Smoking and CAD. We will use public databases to implement a Mendelian randomization approach to determine whether smoking is causally associated with CAD. Due to the ethical issues, no experimental human study can be performed to show whether the association between smoking and CAD is causal. Only a recent study that used the *Inference about Causation through Examination of Familial Confounding* method, disentangled the causal network.[269] We will also analyze the potential mediating role of DNA methylation in this association. We will use summary GWAS data from the UK Biobank (n=337,030) and CARDIOGRAM4D (n_{cases}= 60,801; $n_{controls}$= 123,504). We will then assess whether any CpG site are causally associated with smoking. We will use the mQTLdb resource[334] (http://www.mqtldb.org/) to create a genetic instrument for the CpG sites, and we will use the summary GWAS data from the UK biobank to create a genetic instrument for the smoking exposure.

b. *MYLIP* and CAD. As described above (Section 5.2), *MYLIP* seems to have a similar regulation mechanism to that of *PCSK9*. We will further study the causal association between *MYLIP* methylation and TG and CAD by creating a genetic instrument for *MYLIP* methylation based on mQTLdb data.[334] The genetic instrument for TG and CAD will be created from the specific GWAS summary data of those phenotypes.[240,241]

vii) In addition to the importance of these causal associations between the methylation markers and the phenotypes of interest, and their possible value as pharmaceutical targets, we will assess the utility of these biomarkers for improving the classification of complex syndromes such as obesity.
We will use the 94 CpG sites that explain 14.18% and 16.73% of the variability in BMI and waist circumference, respectively, and the genes where those CpG sites were located. These genes are enriched in pathways related to endocrine and metabolic diseases and psychological and

neurological processes. It may be possible to subcategorize obese individuals in a way that better represents the underlying metabolic pathways or psychological mechanisms. Using these 94 CpG sites, we will apply multidimensional scaling to assess whether these two pathogenic clusters can be identified in individuals with obesity, to differentiate those with a more psychological background from those with a more metabolic background. This clustering could greatly impact therapeutic strategies.

viii) The scientific community has identified some genetic and epigenetics effects on cardiovascular risk factors. Integrating different omics data in the same analysis, such as genetics, epigenetics, transcriptomics, proteomics and metabolomics could improve our understanding of the pathogenesis of complex diseases. The project I really would like to carry out would be an observational study of approximately 2000 participants with baseline samples collected in order to study genetics, epigenetics, transcriptomics, proteomics and metabolomics. I would like to follow them for 10 years and obtain samples and data for all of these omics platforms (except genetics) after 1, 2, 5, 8 and 10 years. In terms of genetics, I would like to perform whole genome sequencing to capture all of the genetic information (rare and common) in those individuals. I would like to observe the changes in the epigenetic data (methylation of the DNA and RNA, non-coding RNA), transcriptomic, proteomics and metabolomics during those years. I would assess the association with phenotypic data, and the most exciting part would be to bring together all of the signals for different omics platforms to try to explain the mechanisms of cardiovascular disease, taking into account genetics and the effect of the environment through other mechanisms.

7. Conclusions

General conclusion 1: We identified significant associations between smoking and DNA methylation.

Specific conclusions:

1.1 Regarding the associations between smoking and DNA methylation level across the genome:

- We report that smoking is associated with a strong DNA methylation signature: 63 CpG sites are hypomethylated and 3 CpG sites are hypermethylated in current active smokers with respect to never smokers.
- We also report a time-dependent association, suggesting that the effect of smoking on DNA methylation could be reversible.
- Our top hit, a methylation site located in *AHRR*, discriminates very accurately between current smokers and never smokers, (AUC>0.926) and could be used as a biomarker for smoking exposure.
- We show that current smokers have reduced methylation levels at the *THBS1* locus, increased expression of the *THBS1* gene, but reduced plasma levels of THBS-1 protein.

1.2 We did not detect any significant association between traffic-related residential air pollution or particulate matter concentration and genome-wide DNA methylation.

General conclusion 2: We identified several loci showing differential methylation related to intermediate cardiovascular traits, especially obesity, blood lipids and the main functional properties of HDL particles.

Specific conclusions:

2.1 We identified 94 CpG sites associated with obesity traits, 70 of which were novel discoveries. These CpG sites explaining approximately 15% of the variability in BMI and waist circumference. These loci are enriched in four metabolic networks related to endocrine and metabolic diseases, and psychological and neurological processes, highlighting the importance of metabolic and psychological pathways in the pathogenesis of obesity.

2.2 We report differential methylation signatures in 14 CpG sites, located in 9 genes and 2 intergenic regions, associated with three lipid traits (TC, HDL-C and TG). Five of these loci were new discoveries (*PHOSPHO1*, *SYNGAP1*, *TXNIP*, *MYLIP* and *SLC7A11*). Globally, these methylation markers explain around 1%, 10% and 19% of the variability in TC, HDL-C and TG, respectively.

2.3 We observed a causal effect from TG to DNA methylation at two loci located in *CPT1A* and *SLC7A11*. We also report a causal relationship from methylation at *CPT1A* locus to TG, which demonstrates the complex relationship between DNA methylation and serum lipids.

2.4 We identified for the first time hypomethylation in 4 genes associated with better cholesterol efflux capacity (*HOXA3*, *PER3*, *PEX5*, *CIMP*) and better anti-inflammatory capacity (*GABRR1*), the two main functional properties of HDL particles.

8. Bibliography

1. Tejada Vera B, Bastian B, Arias E, Al. E. Mortality trends in the United States, 1900-2015. Natl Cent Heal Stat. 2017.
2. CDC. Achievements in Public Health, 1900-1999: Changes in the Public Health System. MMWR MorbMortalWklyRep. 1999;48(50):1141–7.
3. Sayols-Baixeras S, Lluis-Ganella C, Lucas G, Elosua R. Pathogenesis of coronary artery disease: focus on genetic risk factors and identification of genetic variants. Appl Clin Genet. 2014; 7:15–32.
4. Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulkader RS, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1260–344.
5. Wilkins E, Wilson L, Wichramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, et al. European Cardiovascular Disease Statistics 2017. Eur Heart Network, Brussels. 2017.
6. Raab W. Alimentäre faktoren in der entstehung von arteriosklerose und hypertonie. Med Klin. 1932;28:487–521.
7. O'Donnell CJ, Elosua R. [Cardiovascular risk factors. Insights from Framingham Heart Study]. Rev.Esp.Cardiol. 2008; 61:299–310.
8. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. American journal of public health. 1951; 41:279–81.
9. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. Ann Intern Med. 1961;55:33–50.
10. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. J Chronic Dis. 2005;20: 511–24.
11. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. European Heart Journal. 2016;37:2315–81.
12. Rouwet E, Lutgens E. 2016 Jeffrey M. Hoeg Award Lecture: Immune Checkpoints in Atherosclerosis: Toward Immunotherapy for Atheroprotection. Aterioscler Thromb Vasc Biol. 2018.
13. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature; 473 (7347):317–25.
14. Lewandowski R, Gracey C. Chapter 2: Classification of Cardiovascular Disease Risk and Cardiovascular Disease Prevention in Women. In: Management of Cardiovascular Disease in Women. 2014:13–42.
15. Dhingra R, Vasan RS. Age as a Cardiovascular Risk Factor. Med Clin North Am. 2012;96(1):87–91.
16. Sniderman AD, Furberg CD. Age as a modifiable risk factor for cardiovascular disease. Lancet. 2008;371(9623):1547–9.
17. Leritz EC, McGlinchey RE, Kellison I, Rudolph JL, Milberg WP. Cardiovascular Disease Risk Factors and Cognition in the Elderly. Current Cardiovascular Risk Reports. 2011;5:407–12.
18. Singh RB, Mengi SA, Xu YJ, Arneja AS, Dhalla NS. Pathogenesis of atherosclerosis: A multifactorial process. Experimental and Clinical Cardiology.

2002;7:40–53.
19. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV, et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol Cell. 2013;49(2):359–67.
20. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. Nature. 1990;345(6274):458–60.
21. de Magalhães JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 2009;25(7):875–81.
22. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.
23. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, et al. DNA methylation-based measures of biological age: Meta-analysis predicting time to death. Aging (Albany NY). 2016;8(9):1844–65.
24. Soriano-Tárraga C, Giralt-Steinhauer E, Mola-Caminal M, Vivanco-Hidalgo RM, Ois A, Rodríguez-Campello A, et al. Ischemic stroke patients are biologically older than their chronological age. Aging (Albany NY). 2016;8(11):2655–66.
25. Appelman Y, van Rijn BB, ten Haaf ME, Boersma E, Peters SAE. Sex differences in cardiovascular risk factors and disease prevention. Atherosclerosis. 2015;241(1):211–8.
26. Giovino GA, Mirza SA, Samet JM, Gupta PC, Jarvis MJ, Bhala N, et al. Tobacco use in 3 billion individuals from 16 countries: An analysis of nationally representative cross-sectional household surveys. Lancet. 2012;380(9842):668–79.
27. Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. Lancet. 2011;378(9799):1297–305.
28. Woodward M, Lam TH, Barzi F, Patel A, Gu D, Rodgers A, et al. Smoking, quitting, and the risk of cardiovascular disease among women and men in the Asia-Pacific region. Int J Epidemiol. 2005;34(5):1036–45.
29. Huxley R, Barzi F, Woodward M. Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies. BMJ. 2006;332(7533):73–8.
30. Peters S a E, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775,385 individuals and 12,539 strokes. Lancet. 2014;383(9933):1973–80.
31. Jonas MA, Oates JA, Ockene JK, Hennekens CH. Statement on smoking and cardiovascular disease for health care professionals. Circulation. 1992;86:1664–9.
32. Law MR, Wald NJ. Environmental tobacco smoke and ischemic heart disease. Progress in Cardiovascular Diseases. 2003;46:31–8.
33. Gordon T, Kannel WB, Mcgee D, Dawber TR. Death and coronary attacks in men after giving up cigarette smoking. A Report from the Framingham Study. Lancet. 1974;304(7893):1345–8.
34. Åberg A, Bergstrand R, Johansson S, Ulvenstam G, Vedin A, Wedel H, et al.

Cessation of smoking after myocardial infarction. Effects on mortality after 10 years. Heart. 1983;49(5):416–22.
35. Kuller LH, Ockene JK, Meilahn E, Wentworth DN, Svendsen KH, Neaton JD. Cigarette smoking and mortality. MRFIT Research Group. Prev Med (Baltim). 1991;20(5):638–54.
36. Lightwood JM, Glantz SA. Short-term Economic and Health Benefits of Smoking Cessation. Circulation. 1997;96(4):1089–96.
37. Barnoya J, Glantz SA. Cardiovascular effects of secondhand smoke: Nearly as large as smoking. Circulation. 2005;111:2684–98.
38. Lidón-Moyano C, Fu M, Pérez-Ortuño R, Ballbè M, Feliu A, Martín-Sánchez JC, et al. Impact of the Spanish smoking legislation among smokers: A longitudinal study with biomarkers in Barcelona (Spain). Addict Behav. 2018;87:101–8.
39. Fernández E, Fu M, Pérez-Ríos M, Schiaffino A, Sureda X, López MJ. Changes in secondhand smoke exposure after smoke-free legislation (Spain, 2006-2011). Nicotine Tob Res. 2017;19(11):1390–4.
40. Agüero F, Dégano IR, Subirana I, Grau M, Zamora A, Sala J, et al. Impact of a Partial Smoke-Free Legislation on Myocardial Infarction Incidence, Mortality and Case-Fatality in a Population-Based Registry: The REGICOR Study. PLoS ONE. 2013;8:e53722.
41. Frazer K, Callinan J, McHugh J, van Baarsel S, Clarke A, Doherty K, et al. Legislative smoking bans for reducing harms fromsecondhand smoke exposure, smoking prevalence and tobacco consumption. Cochrane Database Syst Rev. 2016;2:CD005992.
42. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American college of cardiology/American heart association task force on practice guidelines. Circulation. 2014;129.
43. Rawshani A, Rawshani A, Franzén S, Eliasson B, Svensson A-M, Miftaraj M, et al. Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. N Engl J Med. 2017;376(15):1407–18.
44. Rydén L, Grant PJ, Anker SD, Berne C, Cosentino F, Danchin N, et al. ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J. 2013;34(39):3035–87.
45. Wilson PW, McGee DL, Kannel WB. Obesity, very low density lipoproteins, and glucose intolerance over fourteen years: The Framingham Study. Am J Epidemiol. 1981;114(5):697–704.
46. Simmons RK, Alberti KGMM, Gale EAM, Colagiuri S, Tuomilehto J, Qiao Q, et al. The metabolic syndrome: useful concept or clinical tool? Report of a WHO Expert Consultation. Diabetologia. 2010;53(4):600–5.
47. Elosua R, Lluís-Ganella C, Subirana I, Havulinna A, Läll K, Lucas G, et al. Cardiovascular Risk Factors and Ischemic Heart Disease: Is the Confluence of Risk Factors Greater Than the Parts? A Genetic Approach. Circ Cardiovasc Genet. 2016;9(3):279–86.
48. Baris G, Mach F. Lipid management in ACS: Should we go lower faster? Atherosclerosis. 2018;275:368–75.
49. Hadjiphilippou S, Ray KK. Lipids and Lipoproteins in Risk Prediction.

Cardiology Clinics. 2018;36:213–20.
50. Maranhão RC, Filho CA, Sigal GA, Chagas ACP, da Luz PL. Chapter 22: HDL and Endothelium. In: Endothelium and Cardiovascular Diseases: Vascular Biology and Clinical Syndromes. 2018:297–317.
51. Siddiqi HK, Kiss D, Rader D. HDL-cholesterol and cardiovascular disease: Rethinking our approach. Current Opinion in Cardiology. 2015;30:536–42.
52. Rader DJ, Hovingh GK. HDL and cardiovascular disease. Lancet. 2014;384(9943):618–25.
53. Brewer HB. Increasing HDL Cholesterol Levels. N Engl J Med. 2004;350(15):1491–4.
54. Vitali C, Khetarpal SA, Rader DJ. HDL Cholesterol Metabolism and the Risk of CHD: New Insights from Human Genetics. Current Cardiology Reports. 2017;19.
55. White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et al. Association of lipid fractions with risks for coronary artery disease and diabetes. JAMA Cardiol. 2016;1(6):692–9.
56. Rohatgi A, Khera A, Berry JD, Givens EG, Ayers CR, Wedin KE, et al. HDL Cholesterol Efflux Capacity and Incident Cardiovascular Events. N Engl J Med. 2014;371(25):2383–93.
57. McBride P. Triglycerides and risk for coronary artery disease. Curr Atheroscler Rep. 2008;10(5):386–90.
58. Saeed A, Feofanova E V., Yu B, Sun W, Virani SS, Nambi V, et al. Remnant-Like Particle Cholesterol, Low-Density Lipoprotein Triglycerides, and Incident Cardiovascular Disease. J Am Coll Cardiol. 2018;72(2):156–69.
59. Dewey FE, Gusarova V, O'Dushlaine C, Gottesman O, Trejos J, Hunt C, et al. Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. N Engl J Med. 2016;374(12):1123–33.
60. Dewey FE, Gusarova V, Dunbar RL, O'Dushlaine C, Schurmann C, Gottesman O, et al. Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. N Engl J Med. 2017;377(3):211–21.
61. Rader DJ, Alexander ET, Weibel GL, Billheimer J, Rothblat GH. The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. J Lipid Res. 2009;50(Supplement):S189–94.
62. Hernaez A. Effects of the mediterranean diet and virgin olive oil on the function of high-density lipoproteins and the atherogeneicity of low-density lipoproteins in humans. University of Barcelona; 2016.
63. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJP, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. N Engl J Med. 2007;357(21):2109–22.
64. HPS3/TIMI55-REVEAL Collaborative Group. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. N Engl J Med. 2017;377:1217-27.
65. Zanoni P, Khetarpal SA, Larach DB, Hancock-Cerutti WF, Millar JS, Cuchel M, et al. Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. Science. 2016;351(6278):1166–71.
66. Wilson Tang WH, Wu Y, Mann S, Pepoy M, Shrestha K, Borowski AG, et al. Diminished antioxidant activity of high-density lipoprotein-associated proteins

in systolic heart failure. Circ Hear Fail. 2011;4(1):59–64.
67. Parto P, Lavie CJ. Obesity and CardiovascularDiseases. Curr Probl Cardiol. 2017;42(11):376–94.
68. Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health: erratum. Int J surgery Oncol. 2017;2(9):e42.
69. Elagizi A, Kachur S, Lavie CJ, Carbone S, Pandey A, Ortega F, et al. An Overview and Update on Obesity and the Obesity Paradox in Cardiovascular Diseases. Prog Cardiovasc Dis. 2018;61:142-50.
70. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux A V, et al. Particulate Matter Air Pollution and Cardiovascular Disease. Circulation. 2010;121(21):2331–78.
71. Hadley MB, Vedanthan R, Fuster V. Air pollution and cardiovascular disease: A window of opportunity. Vol. 15, Nature Reviews Cardiology. 2018. p. 193–4.
72. Dominici F, Peng R, Bell M. Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. JAMA. 2006;10(295):1127–34.
73. Elosua R, Sayols-Baixeras S. The Genetics of Ischemic Heart Disease: From Current Knowledge to Clinical Implications. Rev Esp Cardiol. 2017;70(9):754–62.
74. Gayon J. From Mendel to epigenetics: History of genetics. C R Biol. 2016;339(7-8):225–30.
75. Garrod AE. The incidence of alcaptonuria: a study in chemical individuality. Lancet. 1902;160:616–20.
76. Keynes M, Cox TM. William Bateson, the rediscoverer of Mendel. Journal of the Royal Society of Medicine. 2008;101:104.
77. Portin P, Wilkins A. The evolving definition of the term “Gene.” Genetics. 2017;205(4):1353–64.
78. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. 1952;36(1):39–56.
79. Watson JD, CRICK FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171:737–8.
80. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016;107:1–8.
81. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51:263–73.
82. Cooper G. The Cell: A Molecular Approach. 2nd Editio. Sinauer Associates; 2000.
83. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291:1304–51.
84. International Human Genome Sequencing Consortium, Human I, Sequencing G. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
85. Consortium IH. The International HapMap Project. Nature. 2003;426:789–96.
86. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68–74.
87. Birney E, Soranzo N. Human genomics: The end of the start for population

sequencing. Nature. 2015;526:52–3.
88. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: Higher than you think! Genome Biol. 2011;12(8).
89. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91.
90. Mola-Caminal M. Genetic contribution to functional Outcome and disability after stroke. Universitat de Barcelona; 2017.
91. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. Hum Mol Genet. 2014;23(22):5866–78.
92. Genetic Alliance TNY-M-AC for G and NSS. Understanding Genetics. Genetic Alliance; 2009.
93. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. Cell. 2012;148:1242–57.
94. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature genetics. 2013;45:25–33.
95. Lehrman MA, Schneider WJ, Südhof TC, Brown MS, Goldstein JL, Russell DW. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science. 1985;227(4683):140–6.
96. Varret M, Rabes JP, Saint-Jore B, Cenarro A, Marinoni JC, Civeira F, et al. A third major locus for autosomal dominant hypercholesterolemia maps to 1p34.1-p32. Vol. 64, Am J Hum Genet. 1999;64(5):1378–87.
97. Abifadel M, Varret M, Rabès J-P, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature genetics. 2003;34:154–6.
98. Latimer J, Batty JA, Neely RDG, Kunadian V. PCSK9 inhibitors in the prevention of cardiovascular disease. Journal of Thrombosis and Thrombolysis. 2016;42:405–19.
99. Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nature genetics. 2004;36:233–9.
100. Wang L, Fan C, Topol SE, Topol EJ, Wang Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. Science. 2003;302:1578–81.
101. Cambien F, Poirier O, Lecerf L, Evans A, Cambou JP, Arveiler D, et al. Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature. 1992;359(6396):641–4.
102. Noble WS. How does multiple testing correction work? Nat Biotechnol. 2009;27:1135–7.
103. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49(9):1385–91.
104. van der Harst P, Verweij N. The Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery

Disease. Circ Res. 2017;CIRCRESAHA.117.312086.
105. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010; 42:937–48.
106. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42:949–60.
107. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.
108. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:05–16.
109. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42:579–89.
110. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466:707–13.
111. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478:103–9.
112. Wain L V, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet. 2011;43: 1005–11.
113. Kera A, Chaffin M, Aragam K, Hass M, Roselli C, Choi S, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-24.
114. Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, Ur-Rehman S, et al. The European Genome-phenome Archive of human data consented for biomedical research. Nat Genet. 2015;47:692–5.
115. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet;39:1181–6.
116. Khera AV, Won H-H, Peloso GM, O'Dushlaine C, Liu D, Stitziel NO, et al. Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. JAMA. 2017;317(9):937-46.
117. Stitziel NO, Khera A V., Wang X, Bierhals AJ, Vourakis AC, Sperry AE, et al. ANGPTL3 Deficiency and Protection Against Coronary Artery Disease. J Am Coll Cardiol. 2017;69(16):2054–63.
118. Myocardial Infarction Genetics and CARDIoGRAM, Exome Consortia Investigators. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. N Engl J Med. 2016;12374:1134–44.
119. Do R, Stitziel NO, Won HH, Jørgensen AB, Duga S, Merlini PA, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature. 2015;518(7537):102–6.
120. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al.

Use of genome-wide association studies for drug repositioning. Vol. 30, Nat Biotechnol. 2012;30:317–20.
121. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007;316:1488–91.
122. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Vol. 316, Science. 2007;316:1491–3.
123. McPherson R. Chromosome 9p21.3 Locus for CAD: How little we know. JAmCollCardiol. 2013;62:1382–3.
124. Burgess S, Thompson SG, Genetics Collaboration CC. Methods for meta-analysis of individual participant data from Mendelian randomisation studies with binary outcomes. Stat Methods Med Res. 2012;1–22.
125. CA E, AV K, Kathiresan S. Mendelian randomization. JAMA. 2017;318(19):1925–6.
126. Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018;362.
127. Latvala A, Ollikainen M. Mendelian randomization in (epi)genetic epidemiology: an effective tool to be handled with care. Genome Biol. 2016;17(1):156.
128. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol. 2016;17(1):138.
129. 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.
130. Jørgensen AB, Frikke-Schmidt R, West AS, Grande P, Nordestgaard BG, Tybjærg-Hansen A. Genetically elevated non-fasting triglycerides and calculated remnant cholesterol as causal risk factors for myocardial infarction. Eur Heart J. 2013;34(24):1826-33.
131. Varbo A, Benn M, Tybjaerg-Hansen A, Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG. Remnant cholesterol as a causal risk factor for ischemic heart disease. Journal of the American College of Cardiology. 2013;61:427–36.
132. Holmes M V., Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015;36(9):539–50.
133. Interleukin-6 Receptor Mendelian Randomisation Analysis C. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet. 2012;379(9822):1214–24.
134. Hägg S, Fall T, Ploner A, Mägi R, Fischer K, Draisma HHM, et al. Adiposity as a cause of cardiovascular disease: A Mendelian randomization study. Int J Epidemiol. 2015;44(2):578–86.
135. Ahmad OS, Morris JA, Mujammami M, Forgetta V, Leong A, Li R, et al. A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. Nat Commun. 2015;6.
136. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: A

mendelian randomisation study. Lancet. 2012;380(9841):572–80.
137. Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA. 2009;302(1):37–48.
138. C Reactive Protein Coronary Heart Disease Genetics Collaboration. Association between C reactive protein and coronary heart disease: Mendelian randomisation analysis based on individual participant data. BMJ. 2011;342:d548.
139. Borges MC, Lawlor D a, de Oliveira C, White J, Horta B, Barros AJ. The Role of Adiponectin in Coronary Heart Disease Risk: A Mendelian Randomization Study. Circ Res. 2016;119(3)491-9.
140. Stender S, Frikke-Schmidt R, Nordestgaard BG, Grande P, Tybjærg-Hansen A. Genetically elevated bilirubin and risk of ischaemic heart disease: Three Mendelian randomization studies and a meta-analysis. J Intern Med. 2013;273(1):59–68.
141. Brøndum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: A Mendelian randomization study. Int J Epidemiol. 2015;44(2):651–61.
142. Schmidt AF, Swerdlow DI, Holmes M V, Patel RS, Fairhurst-Hunter Z, Lyall DM, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2017;5(2):97–105.
143. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010;26(9):1205–10.
144. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102–10.
145. Myocardial Infarction Genetics Consortium Investigators, Stitziel NO, Won H-H, Morrison AC, Peloso GM, Do R, et al. Inactivating Mutations in *NPC1L1* and Protection from Coronary Heart Disease. N Engl J Med. 2014;371(22):2072–82.
146. Cannon CP, Blazing M a, Giugliano RP, McCagg A, White JA, Theroux P, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. N Engl J Med. 2015;372(25):2387–97.
147. Marrugat J, Vila J, Baena-Diez JM, Grau M, Sala J, Ramos R, et al. [Relative validity of the 10-year cardiovascular risk estimate in a population cohort of the REGICOR study]. Rev Esp Cardiol;64:385–94.
148. Davies RW, Dandona S, Stewart AFR, Chen L, Ellis SG, Tang WHW, et al. Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. Circulation: Cardiovascular Genetics. 2010;3:468–74.
149. Anderson JL, Horne BD, Camp NJ, Muhlestein JB, Hopkins PN, Cannon-Albright L a., et al. Joint effects of common genetic variants from multiple genes and pathways on the risk of premature coronary artery disease. American Heart Journal. 2010;160:250–6.
150. Qi L, Ma J, Qi Q, Hartiala J, Allayee H, Campos H. Genetic risk score and risk of myocardial infarction in Hispanics. Circulation. 2011;123:374–80.

151. Qi L, Parast L, Cai T, Powers C, Gervino E V, Hauser TH, et al. Genetic susceptibility to coronary heart disease in type 2 diabetes: 3 independent studies. J Am Coll Cardiol. 2011;58:2675–82.
152. Lv X, Zhang Y, Rao S, Qiu J, Wang M, Luo X, et al. Joint effects of genetic variants in multiple loci on the risk of coronary artery disease in Chinese Han subjects. Circ J. 2012;76:1987–92.
153. Patel RS, Sun Y V, Hartiala J, Veledar E, Su S, Sher S, et al. Association of a genetic risk score with prevalent and incident myocardial infarction in subjects undergoing coronary angiography. Circ Cardiovasc Genet. 2012;5:441–9.
154. Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, Klopp N, et al. Genetic markers enhance coronary risk prediction in men: The MORGAM prospective cohorts. PLoS ONE. 2012;7:p. e40922.
155. Vaarhorst AA, Lu Y, Heijmans BT, Dolle ME, Bohringer S, Putter H, et al. Literature-based genetic risk scores for coronary heart disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. Circ Cardiovasc Genet. 2012;5:202–9.
156. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. Am J Epidemiol. 2007;166:28–35.
157. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240–9.
158. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD, et al. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin Chem. 2008;54:467–74.
159. Brautbar A, Ballantyne CM, Lawson K, Nambi V, Chambless L, Folsom AR, et al. Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the atherosclerosis risk in communities study. Circulation: Cardiovascular Genetics. 2009;2:279–85.
160. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. Ann Intern Med. 2009;150:65–72.
161. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010;303:631–7.
162. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010;376:1393–400.
163. Shiffman D, O'Meara ES, Rowland CM, Louie JZ, Cushman M, Tracy RP, et al. The contribution of a 9p21.3 variant, a KIF6 variant, and C-reactive protein to predicting risk of myocardial infarction in a prospective study. BMC Cardiovasc Disord. 2011;11:10.
164. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, et al. A Genetic Risk Score is Associated with Incident Cardiovascular Disease and Coronary Artery Calcium: The Framingham Heart Study. Circ Cardiovasc Genet. 2015;5:113–21.

165. Lluis-Ganella C, Subirana I, Lucas G, Tomas M, Munoz D, Senti M, et al. Assessment of the value of a genetic risk score in improving the estimation of coronary risk. Atherosclerosis. 2012;222:456–63.
166. Gränsbo K, Almgren P, Sjögren M, Smith JG, Engström G, Hedblad B, et al. Chromosome 9p21 genetic variation explains 13% of cardiovascular disease incidence but does not improve risk prediction. Journal of Internal Medicine. 2013;274:233–40.
167. Isaacs A, Willems SM, Bos D, Dehghan A, Hofman A, Arfan Ikram M, et al. Risk scores of common genetic variants for lipid levels influence atherosclerosis and incident coronary heart disease. Aterioscler Thromb Vasc Biol. 2013;33:2233–9.
168. Ganna A, Magnusson PKE, Pedersen NL, De Faire U, Reilly M, Ärnlöv J, et al. Multilocus genetic risk scores for coronary heart disease prediction. Aterioscler Thromb Vasc Biol. 2013;33:2267–72.
169. Tikkanen E, Havulinna AS, Palotie A, Salomaa V, Ripatti S. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. Aterioscler Thromb Vasc Biol. 2013;33:2261–6.
170. Goldstein BA, Knowles JW, Salfati E, Ioannidis JPA, Assimes TL. Simple, standardized incorporation of genetic risk into non-genetic risk prediction tools for complex traits: Coronary heart disease as an example. Front Genet. 2014;5:254.
171. Krarup NT, Borglykke A, Allin KH, Sandholt CH, Justesen JM, Andersson EA, et al. A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 danish individuals. Atherosclerosis. 2015;240(2):305–10.
172. De Vries PS, Kavousi M, Ligthart S, Uitterlinden AG, Hofman A, Franco OH, et al. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: The Rotterdam Study. Int J Epidemiol. 2015;44(2):682–8.
173. Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. Eur Hear J. 2016;37(6):561–7.
174. Iribarren C, Lu M, Jorgenson E, Martínez M, Lluis-Ganella C, Subirana I, et al. Clinical Utility of Multimarker Genetic Risk Scores for Prediction of Incident Coronary Heart Disease: A cohort study among over 51 thousand individuals of European Ancestry. Circ Cardiovasc Genet. 2016;9(6):531–40.
175. Postmus I, Trompet S, Deshmukh HA, Barnes MR, Li X, Warren HR, et al. Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. Nat Commun. 2014;5.
176. Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. N Engl J Med. 2008;359789–99.
177. US Department of Health and Human Services. Information on Clopidogrel Bisulfate. 2007 [cited 2017 Jan 12]. Available from: http://www.fda.gov/
178. Holliday R. Epigenetics: A historical overview. Epigenetics. 2006;1:76–80.
179. Goldberg B. Epigenesis and the rationality of nature in William Harvey and Margaret Cavendish. Hist phylosophy life Sci. 2017;39(2):8.

180. Deichmann U. Epigenetics: The origins and evolution of a fashionable topic. Dev Biol. 2016;416(1):249–54.
181. Nanney D. Epigenetic control systems. PNAS. 1958;44:712–7.
182. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. 1961;3:318–56.
183. Britten RJ, Davidson EH. Gene regulation for higher cells: A theory. Science. 1969;165(3891):349–57.
184. Riggs A, Russo VEA, Martienssen R. Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press. 1996;32.
185. Elhamamsy AR. Role of DNA methylation in imprinting disorders: an updated review. Journal of Assisted Reproduction and Genetics. 2017;34:549–62.
186. Bajrami E, Spiroski M. Genomic imprinting. Maced J Med Sci. 2016;4(1):181-184.
187. Budhavarapu VN, Chavez M, Tyler JK. How is epigenetic information maintained through DNA replication? Epigenetics and Chromatin. 2013;6.
188. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010;465:721–7.
189. Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet. 2010;11(3):191–203.
190. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28(10):1057–68.
191. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92.
192. Feil R, Fraga MF. Epigenetics and the environment: Emerging patterns and implications. Nat Rev. 2012;13:97–109.
193. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128:693–705.
194. Luco R, Pan Q, Tominaga K. Regulation of alternative splicing by histone modifications. Science. 2010;327(5968):996–1000.
195. Huertas D, Sendra R, Muñoz P. Chromatin dynamics coupled to DNA repair. Epigenetics. 2009;4:31–42.
196. Dong Y, Xu S, Liu J, Ponnusamy M, Zhao Y, Zhang Y, et al. Non-coding RNA-linked epigenetic regulation in cardiac hypertrophy. Int J Biol Sci. 2018;14(9):1133-41.
197. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009;10(3):184–94.
198. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. 2010;20(4):434–9.
199. Ptak C, Petronis A. Epigenetics and Complex Disease: From Etiology to New Therapeutics. Annu Rev Pharmacol Toxicol. 2008;48(1):257–76.
200. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. From The Cover: Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci. 2005;102(30):10604–9.
201. Permutt MA, Wasson J, Cox N. Genetic epidemiology of diabetes. Journal of Clinical Investigation. 2005;115:1431–9.
202. Seeman M V. Psychopathology in women and men: Focus on female hormones.

American Journal of Psychiatry. 1997;154:1641–7.
203. Kaminsky Z, Wang SC, Petronis A. Complex disease, gender and epigenetics. Annals of Medicine. 2006;38:530–44.
204. Tang WY, Ho SM. Epigenetic reprogramming and imprinting in origins of disease. Reviews in Endocrine and Metabolic Disorders. 2007;8:173–82.
205. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci. 2008;105(44):17046–9.
206. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhausler BS. Epigenetics and human obesity. Int J Obes. 2015;39:85–97.
207. Rahman MT, Uddin MS, Sultana R, Moue A, Setu M. Polymerase Chain Reaction (PCR): A Short Review. Anwer Khan Mod Med Coll J. 2013;4(1):30-36.
208. Susan JCi, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 1994;22(15):2990–7.
209. Shen L, Waterland RA. Methods of DNA methylation analysis. Curr Opin Clin Nutr Metab Care. 2007;10(5):576–81.
210. Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 1997;25(21):4422–6.
211. Costello J, Smiraglia D, Plass C. Restriction landmark genome scanning. Methods. 2002;27:144–9.
212. Hu M, Yao J, Polyak K. Methylation-specific digital karyotyping. Nat Protoc. 2006;1(3):1621–36.
213. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel). 2016;5(1):3.
214. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98:288–95.
215. Sandoval J, Heyn HA, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692–702.
216. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–99.
217. Bibikova M. DNA Methylation Microarrays. In: Epigenomics in Health and Disease. Academic Press; 2015;19–46.
218. Bibikova M, Fan J, Gunderson KL. DNA methylation profiling using Illumina BeadArray platform. In: Epigenomics: From Chromatin Biology to Therapeutics. Cambridge University Press; 2012. p. 235–48.
219. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3(6):771–84.
220. Du P, Zhang X, Huang C-C, Jafari N, Kibbe W a, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.BMC bioinformatics. 2010;11:587.
221. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A

comprehensive overview of Infinium HumanMethylation450 data processing. Briefings in bioinformatics. 2014;15(6):929-41.
222. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, et al. Review of processing and analysis methods for DNA methylation array data. Br J Cancer;109:1394–402.
223. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14:293.
224. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529–41.
225. Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environmental Health Perspectives. 2012;120:1425–31.
226. Shenker NS, Polidoro S, Van VK, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22:843–51.
227. Sedgwick P. Multiple hypothesis testing and Bonferroni's correction. BMJ. 2014;349.
228. Jeffrey T. Leek and W. Evan Johnson and Hilary S. Parker and Andrew E. Jaffe and John D. Storey. sva: Surrogate Variable Analysis. R package version 3.10.0.
229. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007;447:655–60.
230. Grau M, Subirana I, Elosua R, Solanas P, Ramos R, Masiá R, et al. Trends in cardiovascular risk factor prevalence (1995-2000-2005) in northeastern Spain. Vol. 14, European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2007;14(5):653–9.
231. Ois A, Cuadrado-Godia E, Jimenez-Conde J, Gomis M, Rodriguez-Campello A, Martinez-Rodriguez J, et al. Early Arterial Study in the Prediction of Mortality After Acute Ischemic Stroke. Stroke. 2007;38:2085–9.
232. Roquer J, Rodríguez-Campello A, Gomis M, Jiménez-Conde J, Cuadrado-Godia E, Vivanco R, et al. Acute stroke unit care and early neurological deterioration in ischemic stroke. J Neurol. 2008;255(7):1012–7.
233. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, et al. Primary prevention of cardiovascular disease with a mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med. 2018;378(25).
234. Castañer O, Corella D, Covas MI, Sorlí J V., Subirana I, Flores-Mateo G, et al. In vivo transcriptomic profile after a Mediterranean diet in high-cardiovascular risk patients: A randomized controlled trial. American Journal of Clinical Nutrition. 2013;98:845–53.
235. Marrugat J, Subirana I, Ramos R, Vila J, Marin-Ibanez A, Guembe MJ, et al. Derivation and validation of a set of 10-year cardiovascular risk predictive functions in Spain: the FRESCO Study. Prev Med. 2014;61:66–74.
236. Beulens JWJ, Monninkhof EM, Monique Verschuren WM, van der Schouw YT,

Smit J, Ocke MC, et al. Cohort profile: The EPIC-NL study. Int J Epidemiol. 2010;39(5):1170–8.
237. Palli D, Berrino F, Vineis P, Tumino R, Panico S, Masala G, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. Tumori. 2003;89(0300-8916):586–93.
238. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. American journal of epidemiology. 1979;110:281–90.
239. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. Obesity (Silver Spring). 2015;23(7):1493–501.
240. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45:1274–83.
241. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121–30.
242. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, et al. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clin Epigenetics. 2015;7(1):67.
243. Wojczynski MK, Gao G, Borecki I, Hopkins PN, Parnell L, Lai C-Q, et al. ApoB genetic variants modify the response to fenofibrate: a GOLDN study. J Lipid Res. 2010;526(11):520–6.
244. Manual of The MONICA Project [Manual on the Internet]. Geneva, World Health Organisation; 2000. Available from: http://www.ktl.fi/publications/monica/manual/index.htm
245. Baena-Díez JM, Alzamora-Sas MT, Grau M, Subirana I, Vila J, Torán P, et al. [Validity of the MONICA cardiovascular questionnaire compared with clinical records]. Gac Sanit. 2010;23(6):519–25.
246. Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, et al. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. Atmos Environ. 2013;72:10–23.
247. Eeftens M, Beelen R, De Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. Development of land use regression models for PM2.5, PM 2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project. Environ Sci Technol. 2012;46(20):11195–205.
248. Rivera M, Basagaña X, Aguilera I, Foraster M, Agis D, de Groot E, et al. Association between long-term exposure to traffic-related air pollution and subclinical atherosclerosis: The REGICOR study. Environmental Health Perspectives. 2013;121:223–30.
249. Basagaña X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J, et al. Measurement error in epidemiologic studies of air pollution based on land-use regression models. Am J Epidemiol. 2013;178(8):1342–6.
250. Patel PJ, Khera A V., Jafri K, Wilensky RL, Rader DJ. The anti-oxidative capacity of high-density lipoprotein is reduced in acute coronary syndrome but

not in stable coronary artery disease. J Am Coll Cardiol. 2011;58(20):2068–75.
251. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.
252. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
253. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2010. Available from: http://www.R – project.org/.
254. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8:e63812.
255. Besingi W, Johansson Å. Smoke-related DNA methylation changes in the etiology of human disease. Human Molecular Genetics. 2014;23:2290–7.
256. Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, et al. Dynamics of Smoking-Induced Genome-Wide Methylation Changes with Time Since Smoking Cessation. Hum Mol Genet. 2015;24:2349–59.
257. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics. 2014;9:1382–96.
258. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9(5):436–47.
259. Ono Y, Torii K, Fritsche E, Shintani Y, Nishida E, Nakamura M, et al. Role of the aryl hydrocarbon receptor in tobacco smoke extract-induced matrix metalloproteinase-1 expression. Exp Dermatol. 2013:22(5):349–53.
260. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015;6:10192.
261. Philibert R, Hollenbeck N, Andersen E, Osborn T, Gerrard M, Gibbons FX, et al. A quantitative epigenetic approach for the assessment of cigarette consumption. Front Psychol. 2015;6:1–8.
262. Philibert R, Todorov A. Methods and compositions for predicting tobacco use. World; WO2015155660A1, 2015.
263. D'Agostino M, Tornillo G, Caporaso MG, Barone M V., Ghigo E, Bonatti S, et al. Ligand of Numb proteins LNX1p80 and LNX2 interact with the human glycoprotein CD8 and promote its ubiquitylation and endocytosis. Journal of Cell Science. 2011;124:3545–56.
264. Zhou J, Fujiwara T, Ye S, Li X, Zhao H. Ubiquitin E3 Ligase LNX2 is Critical for Osteoclastogenesis In Vitro by Regulating M-CSF/RANKL Signaling and Notch2. Calcif Tissue Int. 2015;96(5):465–75.
265. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8(5):599–618.
266. Lee MK, Hong Y, Kim SY, London SJ, Kim WJ. DNA methylation and smoking in Korean adults: epigenome-wide association study. Clin Epigenetics. 2016;8(1):103.

267. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum Mol Genet. 2009;21:3073–82.
268. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680–96.
269. Li S, Wong EM, Bui M, Nguyen TL, Joo JHE, Stone J, et al. Causal effect of smoking on DNA methylation in peripheral blood: A twin and family study. Clin Epigenetics. 2018;10(1):18.
270. de F.C. Lichtenfels AJ, van der Plaat DA, de Jong K, van Diemen CC, Postma DS, Nedeljkovic I, et al. Long-term Air Pollution Exposure, Genome-wide DNA Methylation and Lung Function in the LifeLines Cohort Study. Environ Health Perspect. 2018;126(02):027004.
271. Plusquin M, Guida F, Polidoro S, Vermeulen R, Raaschou-Nielsen O, Campanella G, et al. DNA methylation and exposure to ambient air pollution in two prospective cohorts. Environ Int. 2017;108:127–36.
272. Zhong J, Karlsson O, Wang G, Li J, Guo Y, Lin X, et al. B vitamins attenuate the epigenetic effects of ambient fine particles in a pilot human intervention trial. Proc Natl Acad Sci. 2017;114(13):3503–8.
273. Chi GC, Liu Y, MacDonald JW, Barr RG, Donohue KM, Hensley MD, et al. Long-term outdoor air pollution and DNA methylation in circulating monocytes: results from the Multi-Ethnic Study of Atherosclerosis (MESA). Environ Heal. 2016;15(1):119.
274. Gruzieva O, Xu CJ, Breton C V., Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-wide meta-analysis of methylation in children related to prenatal NO2air pollution exposure. Environ Health Perspect. 2017;125(1):104–10.
275. Goodrich JM, Reddy P, Naidoo RN, Asharam K, Batterman S, Dolinoy DC. Prenatal exposures and DNA methylation in newborns: a pilot study in Durban, South Africa. Environ Sci Process Impacts. 2016;18(7):908–17.
276. Breton C, Gao L, Yao J, Siegmund K, Lurmann F, Gilliland F. Particulate matter, the newborn methylome, and cardio-respiratory health outcomes in childhood. Env Epigenet. 2016;2(2):dvw005.
277. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, et al. Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. Environ Health Perspect. 2016;124(7):983–90.
278. Jiang R, Jones MJ, Sava F, Kobor MS, Carlsten C. Short-term diesel exhaust inhalation in a controlled human crossover study is associated with changes in DNA methylation of circulating mononuclear cells in asthmatics. Part Fibre Toxicol. 2014;11:71.
279. van Veelen W, Klompmaker R, Gloerich M, van Gasteren CJR, Kalkhoven E, Berger R, et al. P18 is a tumor suppressor gene involved in human medullary thyroid carcinoma and pheochromocytoma development. Int J cancer. 2009;124(2):339–45.
280. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, Tan S, Fiorito G, Franke L, Guarrera S, Kasela S, Kriebel J,

Richmond RC, Adamo M, Afzal U, Ala-Korpela M, Albetti B, Ammerpohl O, Apperley JF, Beekman M, Bertazzi PA, CJ. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541(7635):81–6.

281. Arner P, Sinha I, Thorell A, Rydén M, Dahlman-Wright K, Dahlman I. The epigenetic signature of subcutaneous fat cells is linked to altered expression of genes implicated in lipid metabolism in obese women. Clin Epigenetics. 2015;7(1):93.
282. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou Y-H, et al. Epigenome-wide Association Study (EWAS) of BMI, BMI Change, and Waist Circumference in African American Adults Identifies Multiple Replicated Loci. Hum Mol Genet. 2015;24(15):4464–79.
283. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA methylation and body-mass index: A genome-wide analysis. Lancet. 2014;383(9933):1990–8.
284. Almen MS, Nilsson EK, Jacobsson JA, Kalnina I, Klovins J, Fredriksson R, et al. Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. Gene. 2014;548(1):61–7.
285. Ollikainen M, Ismail K, Gervin K, Kyllönen A, Hakkarainen A, Lundbom J, et al. Genome-wide blood DNA methylation alterations at regulatory elements and heterochromatic regions in monozygotic twins discordant for obesity and liver fat. Clin Epigenetics. 2015;7(1):39.
286. Voisin S, Almén MS, Zheleznyakova GY, Lundberg L, Zarei S, Castillo S, et al. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. Genome Med. 2015;7(1):103.
287. Mansego ML, Milagro FI, Zulet MA, Moreno-Aliaga MJ, Martinez JA. Differential DNA Methylation in Relation to Age and Health Risks of Obesity. Int J Mol Sci. 2015;16(8):16816–32.
288. Rönn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen A, et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet. 2015;24(13):3792–813.
289. Al Muftah W, Al-Shafai M, Zaghlool S, Visconti A, Tsai P, Kumar P, et al. Epigenetic associations of type 2 diabetes and BMI in an Arab population. Clin Epigenetics. 2016;8(13):1–10.
290. Huang RC, Garratt ES, Pan H, Wu Y, Davis EA, Barton SJ, et al. Genome-wide methylation analysis identifies differentially methylated CpG loci associated with severe obesity in childhood. Epigenetics. 2015;10(11):995–1005.
291. Burmeister M, McInnis MG, Zöllner S. Psychiatric genetics: progress amid controversy. Nat Rev Genet. 2008;9(7):527–40.
292. Stratigopoulos G, Leduc C a, Cremona ML, Chung WK, Leibel RL. Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated (FTO) and retinitis pigmentosa gtpase regulator interacting protein-1 like (RPGRIPIL) genes, and co-ordinates leptin receptor signaling. J Biol Chem. 2010;1:1–26.
293. Leiper J, Nandi M. The therapeutic potential of targeting endogenous inhibitors of nitric oxide synthesis. Nat Rev Drug Discov. 2011;10(4):277–91.

294. Khor GH, Froemming GRA, Zain RB, Thomas Abraham M, Omar E, Tan SK, et al. DNA methylation profiling revealed promoter hypermethylation-induced silencing of p16, DDAH2 and DUSP1 in primary oral squamous cell carcinoma. Int J Med Sci. 2013;10(12):1727–39.
295. Gaber EM, Jayaprakash P, Qureshi MA, Parekh K, Oz M, Adrian TE, et al. Effects of a sucrose-enriched diet on the pattern of gene expression, contraction and Ca2+ transport in Goto-Kakizaki type 2 diabetic rat heart. Exp Physiol. 2014;99(6):881–93.
296. Howarth FC, Qureshi M a, Hassan Z, Al Kury LT, Isaev D, Parekh K, et al. Changing pattern of gene expression is associated with ventricular myocyte dysfunction and altered mechanisms of Ca2+ signalling in young type 2 Zucker diabetic fatty rat heart. Exp Physiol. 2011;96(3):325–37.
297. Nagel DJ, Aizawa T, Jeon K-I, Liu W, Mohan A, Wei H, et al. Role of nuclear Ca2+/calmodulin-stimulated phosphodiesterase 1A in vascular smooth muscle cell growth and survival. Circ Res. 2006;98(6):777–84.
298. Milagro FI, Gómez-Abellán P, Campión J, Martínez JA, Ordovás JM, Garaulet M. CLOCK, PER2 and BMAL1 DNA Methylation: Association with Obesity and Metabolic Syndrome Characteristics and Monounsaturated Fat Intake. Chronobiol Int. 2012;29(9):1180–94.
299. Gelissen IC, Harris M, Rye KA, Quinn C, Brown AJ, Kockx M, et al. ABCA1 and ABCG1 synergize to mediate cholesterol export to ApoA-I. Arterioscler Thromb Vasc Biol. 2006;26(3):534–40.
300. Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA methylation of lipid-related genes affects blood lipid levels. Circ Cardiovasc Genet. 2015;8(1942-3268):334–42.
301. Guay SP, Brisson D, Munger J, Lamarche B, Gaudet D, Bouchard L. ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia. Epigenetics. 2012;7(5):464–72.
302. Peng P, Wang L, Yang X, Huang X, BA Y, Chen X, GUO J, Lian J ZJ. A preliminary study of the relationship between promoter methylation of the ABCG1, GALNT2 and HMGCR genes and coronary heart disease. PLoS One. 2014;9(8):e102265.
303. Horton JD. Sterol regulatory element-binding proteins: transcriptional activators of lipid synthesis. Biochem Soc Trans. 2002;30(6):1091–5.
304. Edwards P a., Tabor D, Kast HR, Venkateswaran A. Regulation of gene expression by SREBP and SCAP. Biochim Biophys Acta - Mol Cell Biol Lipids. 2000;1529(1-3):103–13.
305. Kulkarni H, Kos MZ, Neary J, Dyer TD, Göring HHH, Cole SA, et al. Novel epigenetic determinants of type 2 diabetes in mexican american families. Hum Mol Genet. 2015;24(18)
306. Aslibekyan S, Do AN, Xu H, Li S, Irvin MR, Zhi D, et al. CPT1A methylation is associated with plasma adiponectin. Nutr Metab Cardiovasc Dis. 2017;27(3):225–33.
307. Ye J, DeBose-Boyd RA. Regulation of cholesterol and fatty acid synthesis. Cold Spring Harbor Perspectives in Biology. 2011;3:1–13.
308. Kiffer-Moreira T, Yadav MC, Zhu D, Narisawa S, Sheen C, Stec B, et al.

Pharmacological inhibition of PHOSPHO1 suppresses vascular smooth muscle cell calcification. J bone Miner Res. 2013;28(1):81–91.
309. Bobryshev YV, Orekhov AN, Sobenin I CD. Role of bone-type tissue-nonspecific alkaline phosphatase and PHOSPO1 in vascular calcification. Curr Pharm Des. 2014;20(37):5821–8.
310. Parikh H, Carlsson E, Chutkow WA, Johansson LE, Storgaard H, Poulsen P, et al. TXNIP regulates peripheral glucose metabolism in humans. PLoS Med. 2007;4(5):0868–79.
311. Zhou J, Chng WJ. Roles of thioredoxin binding protein (TXNIP) in oxidative stress, apoptosis and cancer. Mitochondrion. 2013;13:163–9.
312. Turturro F, Friday E, Welbourne T. Hyperglycemia regulates thioredoxin-ROS activity through induction of thioredoxin-interacting protein (TXNIP) in metastatic breast cancer-derived cells MDA-MB-231. BMC Cancer. 2007;7:96.
313. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. lancet Diabetes Endocrinol . 2015;3(7):526–34.
314. Florath I, Butterbach K, Heiss J, Bewerunge-Hudler M, Zhang Y, Schöttker B BH. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. Diabetologia. 2016;59(1):130–8.
315. Soriano-Tarraga C, Jimenez-Conde J, Giralt-Steinhauer E, Mola-Caminal M, Vivanco-Hidalgo R, Ois A, et al.. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. Hum Mol Genet. 2016:25(3):609-19.
316. Lindholm D, Bornhauser BC, Korhonen L. Mylip makes an Idol turn into regulation of LDL receptor. Cell Mol Life Sci. 2009;66:3399–402.
317. Raal FJ, Stein EA, Dufour R, Turner T, Civeira F, Burgess L, et al. PCSK9 inhibition with evolocumab (AMG 145) in heterozygous familial hypercholesterolaemia (RUTHERFORD-2): a randomised, double-blind, placebo-controlled trial. Lancet. 2015;385(9965):331–40.
318. Prchalova D, Havlovicova M, Sterbova K, Stranecky V, Hancarova M, Sedlacek Z. Analysis of 31-year-old patient with SYNGAP1 gene defect points to importance of variants in broader splice regions and reveals developmental trajectory of SYNGAP1-associated phenotype: Case report. BMC Med Genet. 2017;18(1):62.
319. Mandal PK, Seiler A, Perisic T, Kölle P, Canak AB, Förster H, et al. System xc\n- and thioredoxin reductase 1 cooperatively rescue glutathione deficiency. J Biol Chem. 2010;285(29):22244–53.
320. Koppula P, Zhang Y, Zhuang L, Gan B. Amino acid transporter SLC7A11/xCT at the crossroads of regulating redox homeostasis and nutrient dependency of cancer. Cancer Commun. 2018;38(1):12.
321. Pirillo A, Tibolla G, Norata GD C AL. HDL: to treat or not to treat? Curr Atheroscler Rep. 2014;16(8):429.
322. Karpe F, Pinnick KE. Biology of upper-body and lower-body adipose tissue—link to whole-body phenotypes. Nat Rev Endocrinol. 2014;11(2):90–100.
323. Dijk DJ, Archer SN. PERIOD3, circadian phenotypes, and sleep homeostasis. Sleep Medicine Reviews. 2010;14:151–60.

324. Van Veldhoven PP. Biochemistry and genetics of inherited disorders of peroxisomal fatty acid metabolism. J Lipid Res. 2010;51(10):2863–95.
325. García-Martín E, Martínez C, Serrador M, Alonso-Navarro H, Navacerrada F, Esguevillas G, et al. Gamma-Aminobutyric Acid (Gaba) Receptors Rho (Gabrr) Gene Polymorphisms and Risk for Migraine. Headache. 2017;57(7):1118–35.
326. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease. PLoS Med. 2017;14(1):e1002215.
327. Höfler M. The Bradford Hill considerations on causality: A counterfactual perspective. Emerging Themes in Epidemiology. 2005;2:11.
328. Elliott HR, Shihab HA, Lockett GA, Holloway JW, McRae AF, Smith GD, et al. The Role of DNA Methylation in Type 2 Diabetes Aetiology - Using Genotype as a Causal Anchor. Diabetes. 2017;66(6):db160874.
329. Hedman ÅK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin MR, et al. Epigenetic Patterns in Blood Associated with Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results from Genome-Wide Association Studies. Circ Cardiovasc Genet. 2017;10(1).
330. Geurts YM, Dugué PA, Joo JE, Makalic E, Jung CH, Guan W, et al. Novel associations between blood DNA methylation and body mass index in middle-Aged and older adults. Int J Obes. 2018;42(4):887–96.
331. Lowe R, Slodkowicz G, Goldman N, Rakyan VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics. 2015;10(4):274:81.
332. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89–98.
333. Lim S, Wyker B, Bartley K, Eisenhower D. Measurement error of self-reported physical activity levels in new York City: Assessment and correction. Am J Epidemiol. 2015;181(9):648–55.
334. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016;17:61.