Human genetic disorders: linking Mendelian and complex diseases.

Nino Spataro

TESI DOCTORAL UPF / ANY 2016

DIRECTORA DE LA TESI

Dra. Elena Bosch Fusté

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

For my family and the people who cares about me

ACKNOWLEDGMENTS

Ho cominciato questo dottorato con l’entusiasmo di ricongiungermi con te Natalia e finalmente vivere insieme nello stesso posto e dimenticare i lunghi periodi di lontananza. Finalmente abbiamo trovato il nostro posto e personalmente ho trovato il mio all’IBE. Grazie mille per la pazienza e il supporto durante questi anni, sei sempre stata la motivazione principale per cominciare un dottorato, anche se “gli esperti del settore” avrebbero preferito immaginare che diventare un ricercatore fosse il mio sogno sin da bambino. Ma l’entusiasmo dovuto all'inizio di questa nuova esperienza era misto a un po di tristezza per quello che mi lasciavo alle spalle, per la mia famiglia e i miei amici che anno dopo anno sono lontani nello spazio ma sempre vicini nei pensieri. Un grazie speciale va a tutti loro, in particolar modo ai miei genitori che hanno sempre creduto in me inondandomi di tutto l’affetto che ogni figlio dovrebbe ricevere. Non potró mai ripagarvi per quello che avete fatto per me e spero di rendervi orgogliosi cosí come meritate. Un pensiero speciale a te Nicola, spero sia possibile in futuro essere cosí vicini come quando da bambini passavamo intere giornate insieme giocando e condividendo tutto. Grazie alla vasta lista di amici ho avuto l’onore e il piacere di conoscere e alla famiglia allargata che Cattolica rappresenta per me. Grazie Vincenzo per la tua esistenza, il mondo ha bisogno di gente con il tuo atteggiamento positivo misto a pazza genialitá. Grazie Zazá, mi fará sempre piacere ascoltare i tuoi argomenti seri e semplicistici sulla societá e il “sistema” per provare a convincermi che un mondo diverso è possibile. Grazie Matteo per la diametralmente opposta visione e per ricordare a tutti i Zazá del mondo che il duro lavoro e un po di “spirtizza” muove tutto. Grazie Santino, per il tuo talento silenzioso, e tutti i Militi Ignoti, tra i quali Fregapane e la sua “cuetanza” e Giovanni. Ricordate! Sono sempre stato il primo dei vostri fan e il primo dei vostri detrattori...e uno dei fondatori. Se un giorno la musica vi renderá ricchi pretenderó la mia parte! “O semu a causa”! Grazie Peppe Jo e Ciccio, siete le persone sulle quali si puó sempre contare e sono sinceramente onorato della vostro amicizia. Grazie Gery per tutte le risate dovute al tuo sarcasmo e alla tua repulsione per il politicamente corretto. Grazie Amedeo, sai perfettamente che tutti ti vogliamo bene cosí come sei e sono immensamente felice che tutto sia tornato come prima. Un grazie infinito a tutto il gruppo di amici, includendo in particolar modo Tiziana e il suo folle carattere e Margherita per le lunghe discussioni durante i viaggi in autobus. Grazie Anna, Laura Terrasi e Laura Vaccaro, Irene, Rosanna, Noemi, Marta, Liliana e Rossella. Insieme abbiamo condiviso un sacco di momenti meravigliosi, se gli amici sono la famiglia che uno si sceglie, siete di gran lunga la scelta migliore della mia vita. Un abbraccio affettuoso a tutti i ragazzi dell’esperienza romana, a Giorgio e Laura, Vincenzo “u rasta”, William, Lorenzo, Aurelio, Emiliano, Silvia, Alice e tutto il gruppo di Sogene. Grazie Peppe Montanaro e Francesco alias “Uomo French”. Vivere insieme è stato un piacere e i momenti condivisi nella cittá piú bella al mondo saranno sempre nel mio cuore. Grazie ai ragazzi dell'Erasmus e alla famiglia di Hardy Farm. Che posto meraviglioso in una cittá cosí orribile! Dopotutto cara odiata Manchester grazie anche a te. Sono arrivato senza sapere una parola e mi ritrovo dopo qualche anno a scrivere un'intera tesi in Inglese! Li ho conosciuto te Natalia e non potró mai dimenticare i giorni passati insieme e con il resto del nostro "block", in particolare con Neila e Shae-Ree. Grazie ai membri del laboratorio della professoressa Helmer-Citterich, in particolar modo Luca per la sua supervisione, Valerio e Iolanda. Vi auguro il meglio per le vostre carriere e spero poter essere un buon bioinformatico cosi come lo siete voi.

Let's pass to my Spanish (...sorry Catalan) period. Thank you to Jordi Planas (most of you right now will think: "who the hell is Jordi Planas?") and the people of the University of Vic. If I have been working on science it is thanks to the trust you placed in me to make me start a PhD with your group. Your tolerance and understanding of my feelings made this thesis possible. A very special thanks to my supervisor Elena Bosch. Thank you for always being present whenever I needed it, I really enjoyed working with you and I hope to have the possibility to collaborate in the future. You have been the supervisor I needed, you always gave me the scientific freedom to explore what I was interested in, but at the same time you were always close with your supervision and expertise. I never had a boss, but a careful person supervising and teaching me as a mentor should. I hope we made a good team as I

think we did. An enormous thank you to “los compañeros” Elena Carnero and Johannes Engelken. What wonderful people! Thank you for the trust when reading my first e-mail! I am pretty sure that if different people had read it, I would be finishing my PhD in Vic. Only people with your sincere goodness could understand what I meant. Thank you Elena for your happiness and Johannes for your pure and authentic scientific passion, I consider myself a privileged person to have had the possibility to know both of you. A thank you to the 419 family and to the former and current members. This is not an office, but a space where people collaborate and share scientific and not scientific opinions. Thanks to Urko (and his turtles...) for its encyclopedic knowledge and irony, a discussion with you can never be boring or not interesting. Thank you to Ferran and Oscar, we miss all of you in the office. Thank you Juan to be as you are, Diego and Marina for your passion and your long scientific discussions, I miss hearing your voices in the background while working. Thank you David for all your contributions and help, your point of view is never ordinary. How the hell did you learn all the stuff you know? Thank you Josephine and the new family from office 421. I will really miss the time spent together in the space that people simply call “office”. Thank you to the rest of BioEvo and IBE members, for the countless “chocolate now” for any possible reason and for the wonderful retreats during these years. A special mention is for the “Tupper Club” members, probably the only successful experiment I will ever be involved in my whole career. Quoting from a previous thesis: “Tupper Club is the best of my PhD...by far!”. Lunch time could have never been more interesting than what it has been during these years. I am really proud of being part of this social experiment: different people, cultures and sensibilities sharing food and points of view. A mix of differences complementing each other and creating something beautiful. “Sí se puede!”. Thank you very much for the tasty food you have been cooking, when all of this will finish I will really miss reading: “Tupperos a las 1:45 a buscar sitio”. Thank you to the founding fathers: Diego, Elena, Fede, Johannes, Marc, Maria and Pierre for letting the “nuevos” join in. Thanks to the rest of the former and current members: Adriano, Alicia, Alex, Arturo, Juan, Ilaria (is she real or not?), Lara, Lisa, Marco, Maria, Marina, Miruna, Veronica, Vicky. Tupper club is not a group of people eating together, but a space of the soul. A final thank you to the people “del fútbol de los miércoles”, I thought it was crazy to play football at 9 AM but I really enjoyed it and I had a lot of fun with all of you. Thank you to the people “del volley de los jueves”, I am sure next year we will win the Christmas tournament! Thank you to the people that love playing beach volley and to the PRBB staff for the wonderful championship and party they organize every year that make me feel part of a community.

Many thanks to all of you dear people, I consider myself a privileged person and that’s thanks to you. Thank you so much!

Nino Spataro Barcelona, Spain 07/01/2016

ABSTRACT

From Darwin's "On the Origin of Species", many years elapsed before human diseases were considered in an evolutionary framework. Besides theoretical and empirical advances, we are far from the complete understanding of disease aetiology. Highly penetrant disorders with Mendelian inheritance are mostly explained by the mutation-selection balance model, which is insufficient to describe the selective pressures acting on the full set of alleles related to diseases. We show in the first two papers that Next Generation Sequencing (NGS) technologies provide a unique opportunity to investigate variation and contribute to the understanding of the genetic architecture of disease. Besides exploring the role of rare and copy number variants in Parkinson's disease (PD), we demonstrate the functional relation between Mendelian and idiopathic PD. In the last paper, we report that variation in genes previously related to Mendelian disorders has a more important role in driving complex disease susceptibility than genes associated only to complex diseases.

RIASSUNTO

Dall' *"Origine delle Specie"* di Darwin sono passati molti anni prima che le malattie umane fossero considerate in un contesto evolutivo. Nonostante i progressi teorici ed empirici, siamo ancora ben lontani dalla completa comprensione della loro eziologia. Le patologie altamente penetranti ad ereditá Mendeliana sono in larga parte spiegate da un modello basato sull'equilibrio mutazione-selezione, che é insufficiente per descrivere la pressione selettiva che agisce sulla totalitá degli alleli associati a malattie. Nei primi due articoli dimostriamo che le tecnologie di sequenziamento di nuova generazione (NGS) forniscono una possibilitá unica per lo studio della variazione genetica e contribuiscono alla comprensione dell'architettura genetica delle malattie. Oltre a esplorare il ruolo di varianti rare e varianti nel numero di copia nel morbo di Parkinson (PD), dimostriamo la relazione funzionale tra la forma Mendeliana e idiopatica di PD. Nell'ultimo articolo, riportiamo che la variazione nei geni precedentemente relazionati a patologie Mendeliane ha un ruolo piú rilevante nel regolare la suscettibilitá per le malattie complesse rispetto alla variazione osservata nei geni associati unicamente a malattie complesse.

PREFACE

> Fatti non foste a viver come bruti, ma
> per seguir virtute e canoscenza.
>
> *Canto XVI, Inferno. The Divine Comedy.* DANTE ALIGHIERI

The masterpiece “*The Origin of the Species*” of Darwin, published in 1859, represents the cornerstone of life sciences and the beginning of evolutionary biology. The properties of living organisms at any scale were explained on the basis of their environmental context and a precise framework was described to explain how species interact among them and evolve in time. Simultaneously, Mendel was working in hybridization experiments and described the general rules of heredity, establishing the theoretical bases of genetics. Only in 1918, with the work of Fisher and the subsequent contributions of Haldane and Wright between 1936 and 1947, it was theoretically and mathematically demonstrated that Mendelian inheritance was consistent with natural selection, and the so-called “modern synthesis” became to a large extent the current paradigm of evolutionary biology. The discovery of the biochemical nature of genes and the description of the DNA features between 1940 and 1960 gave insights into the mechanical processes underlying species evolution and provided a wonderful possibility to understand the biological processes occurring in living organisms, including genetic diseases.

Alleles underlying human diseases are introduced in the genetic pool of populations by random mutations and eventually eliminated by purifying selection, as this is the evolutionary force that keeps deleterious alleles at low frequencies within populations. This simple mechanism seems to be well suited for strong and highly penetrant single gene disorders with Mendelian inheritance, but it is insufficient to explain the whole spectrum of genetic diseases. Many authors contributed in the last decades to develop alternative theoretical frameworks for the evolutionary explanation of genetic diseases. The recent technological developments and the advent of the so-called *“-omics”* sciences furnished an additional opportunity to describe the genetic variants underlying human diseases and understand their pathogenesis. Even if most of risk factors remain to be detected in most of the cases, in the last decade Genome Wide Association Studies (GWAS) strongly contributed to the understanding of the genetic architecture of complex polygenic diseases. Currently, Next Generation Sequencing (NGS) technologies represent an extraordinary attempt to bridge the gap between genotype and phenotype and address part of the “missing heritability” problem. Each single disease and phenotype can be studied at an unexpected level of resolution compared to the past, providing additional knowledge for the understanding of human genetic disorders.

In the current thesis, I procure a further contribution to the field using a bottom-up approach. First, we use Parkinson’s disease (PD) as an example and we demonstrate that rare functionally relevant variants play a role in complex diseases. In addition, we demonstrate that familial and sporadic forms of PD are functionally related and that rare variants may be at the basis of this relation. Furthermore, we provide evidence that prediction tools based on depth of coverage data correctly identify copy number variants (CNVs) related to PD. The combined analysis of single nucleotide polymorphism (SNPs), smalls indels and CNVs substantially improves the detection rate of PD cases with causal genetic alterations and demonstrates the validity of sequencing experiments to describe different types of variants related to human diseases. In the last part, we provide evidence for the widespread functional interconnection between Mendelian and complex diseases. In particular, we demonstrate that genes harboring causal variants for strongly deleterious disorders with Mendelian inheritance contribute to the aetiology of complex diseases beyond random expectation.

A special thought to all the researchers curating the resources we took advantage of and providing the theoretical and technological tools needed for this work. This modest contribution would not be possible without all of them.

CONTENTS

ACKNOWLEDGMENTS...v

ABSTRACT..xi

RIASSUNTO..xiii

PREFACE...xv

PART I: INTRODUCTION..1

Chapter 1: BACKGROUND...3
1.1 Rise and history of *Homo sapiens*..3
1.2 Human genetic variation...6
1.3 Theoretical framework for genetic diseases..........................11
1.4 Classification of human diseases..22
    1.4.1 Mendelian diseases..23
    1.4.2 Complex diseases...26
1.5 Biological properties of genes related to human disease......34
1.6 The human disease network...37
    1.6.1 Relations among human diseases.................................39
    1.6.2 Blurring the boundaries between complex and Mendelian diseases..43
1.7 Parkinson’s disease: a case study...45
    1.7.1 Mendelian forms of Parkinson’s disease......................48
    1.7.2 Sporadic form of Parkinson’s disease..........................54

Chapter 2: OBJECTIVES..65

PART II: RESULTS..69

CHAPTER 3:
NEXT GENERATION SEQUENCING ON CANDIDATE GENES RELATED TO PARKINSON’S DISEASE..............71

CHAPTER 4:
COPY NUMBER VARIATION ON PARKINSON’S DISEASE...85

CHAPTER 5:
PROPERTIES OF HUMAN DISEASE GENES..................135

PART III: DISCUSSION...185

CHAPTER 6: DISCUSSION..187
6.1 Rare variants in complex diseases......................................187
6.2 Copy number variation on Parkinson’s disease...................191
6.3 The role of Mendelian genes in complex diseases.............193
6.4 Limitations...196
6.5 What's next?...199

CHAPTER 7: CONCLUSIONS...207

PART IV: APPENDIX...211

LIST OF PUBLICATIONS...213

SUPPLEMENTARY MATERIALS.......................................217
Next generation sequencing on candidate genes related to Parkinson’s disease...217
Copy Number Variation on Parkinson’s disease.......................249
Properties of human disease genes..263

BIBLIOGRAPHY...309

ELECTRONIC APPENDIX

Electronic Appendix 1: Supplementary Information for Chapter 3
Electronic Appendix 2: Supplementary Information for Chapter 5

PART I

INTRODUCTION

CHAPTER 1

BACKGROUND

> Doubt is the origin of wisdom.
> RENÉ DESCARTES

1.1 Rise and history of *Homo sapiens.*

According to paleontological and genetic records, the *Homo* genus originated in Africa 2,000-2,500 KYA and gave rise to many extinct lineages. Anatomically modern humans (*Homo sapiens*) emerged ~200 KYA in Eastern Africa and represent the only surviving species. Several recent genetic analyses demonstrate that our past history is far more complicated than was initially believed and the relationships and the extent of admixture among hominin lineages remain debated (Lalueza-Fox and Gilbert 2011). The advent of next generation sequencing technologies and the improvement of DNA isolation techniques from ancient samples allowed the publication of the first Neanderthal genome in 2010 (Green et al. 2010). Subsequent works demonstrate that our close relatives, Neanderthals and Denisovans, diverged from modern humans around 500-600 KYA and spread through the Middle East, Europe and Central Asia (Prufer et al. 2014). Moreover, the fact that no

Homo sapiens fossil records older than 50 KYA exist outside Africa, suggests that modern humans left Africa ~50 KYA and spread slowly worldwide (Tattersall 2009). Notably, several emerging evidences indicate that inter-species breeding occurred and traces of these admixture events remain in the current genetic pool of present-day human populations (Stoneking and Krause 2011). For example, it has been calculated that ~1.5-4% of the European and Asian populations genomes have a Neanderthal origin (Green et al. 2010), while an estimated 4-6% of the genome of Melanesians is derived from Denisovans (Reich et al. 2010).

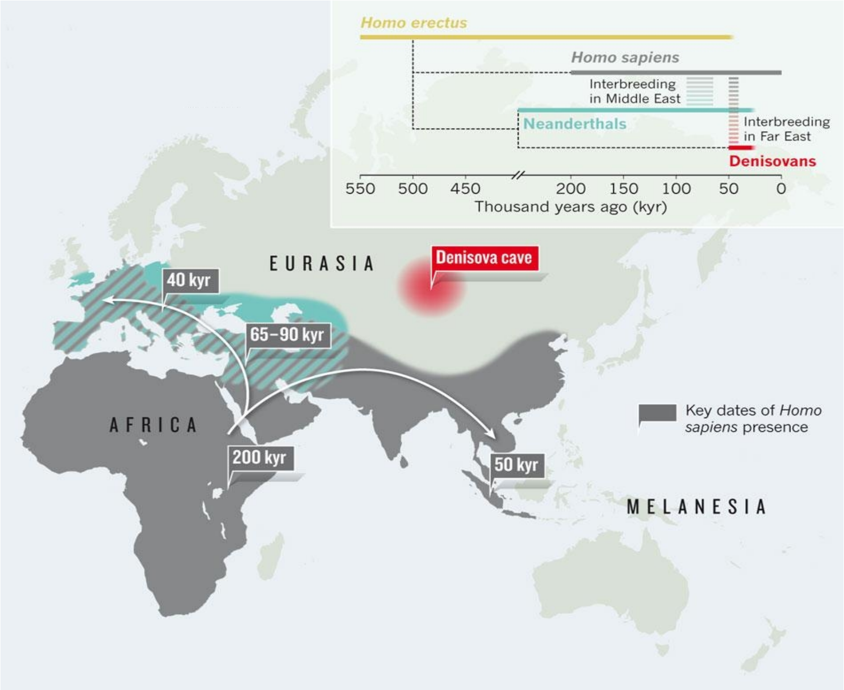

Homo erectus
Homo sapiens
Interbreeding in Middle East
Neanderthals
Interbreeding in Far East
Denisovans
550
500
450
200
150
100
50
0
Thousand years ago (kyr)
EURASIA
Denisova cave
40 kyr
65–90 kyr
AFRICA
200 kyr
50 kyr
Key dates of *Homo sapiens* presence
MELANESIA

Figure 1.1. Geographical distribution and temporal divergence of modern and archaic humans. From Callaway 2011.

Interestingly, genetic variants inherited from Neanderthal increase the risk for depression, mood disorders and actinic keratosis (Simonti et al. 2016), as well as for systemic lupus erythematosus, primary biliary cirrhosis, Crohn’s disease and diabetes mellitus type 2 (Sankararaman et al. 2014), indicating that our past demographic history is also relevant for the understanding of our susceptibility to disease. Conversely, recent studies show that immunity genes and innate immunity genes carry haplotypes that appear to have introgressed from archaic hominin populations and Neanderthals (Abi-Rached et al. 2011; Deschamps et al. 2016) and that may protect against some infectious diseases. Around 10 KYA, with the Mesolithic-Neolithic transition, the advent of agricultural practices and animal domestication implicated a huge technological revolution and a dramatic lifestyle change from nomadic hunter-gatherer communities to sedentary agriculturalist ones. As a result of the subsequent higher population density, the close contact with domesticated animals and the use of inadequate sanitary practices, infectious diseases started to spread easily. Together with the dietary shift, the new conditions help to explain the rise in sickness observed with the Neolithic revolution (Wolfe, Dunavan, and Diamond 2007). These and additional changed life style conditions could also have an effect on human disease risk, as suggested by the “thrifty genotype” (Neel 1962) and the “sodium retention” (Wilson 1986) hypotheses among others. All these hypotheses will be collectively described in detail in the following sections and represent an effort to explain human illness from an evolutionary point of view. Indeed, to have a proper understanding of their incidence and persistence in current populations, human diseases should not be considered just as the result of occasional events that disrupt relevant biological functions, but under a broader evolutionary perspective.

1.2 Human genetic variation.

Genetic variation originates from random errors occurring during DNA replication and only when they involve germ line cells the new mutations are transmitted to the offspring (Drake 1970). According to its size, genetic variation can be classified as point mutations and structural variants. The former represent the most abundant type of variation in the human genome and are generally known as single nucleotide polymorphisms (SNPs); the latter represent an heterogeneous class ranging from small insertions/deletions/inversions of few base pairs to huge structural variants encompassing few megabases and chromosomal rearrangements.

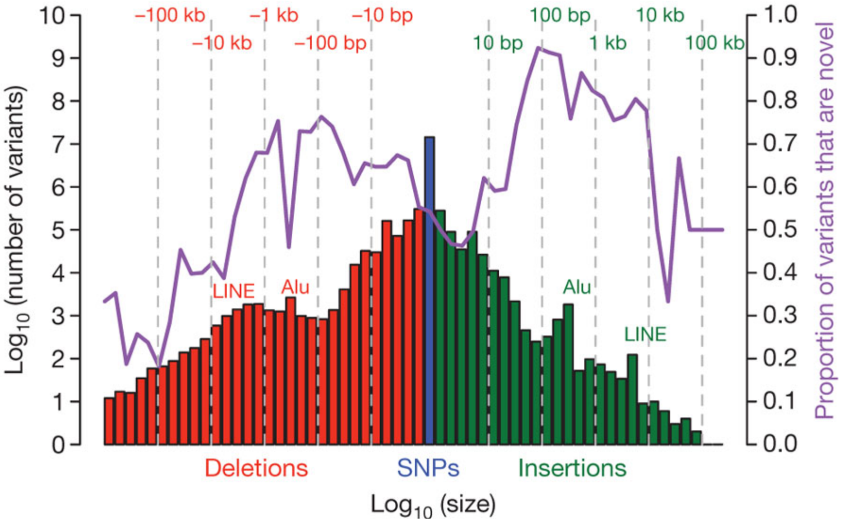
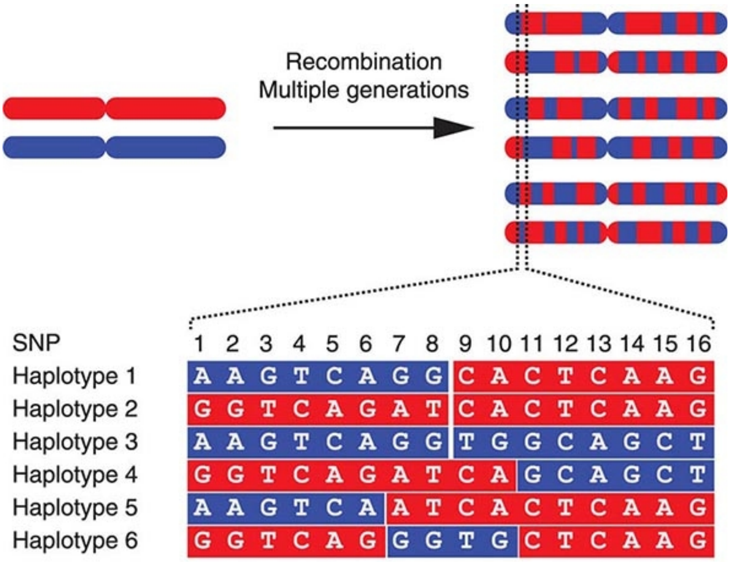


Figure 1.2. Size distribution and novelty of variants discovered in the pilot phase of the 1000 Genomes Project. From Durbin et al. 2010.

Besides mutations, an additional source of variation in our genome is originated by recombination. While mutations generate variation *de novo*, recombination refers to the process in which during meiosis homologous chromosomes from paternal and maternal origin align and exchange chunks of consecutive sequences (Morgan 1911; Creighton and McClintock 1931), rearranging different combinations of existing variants in the same chromosome (haplotypes).



SNP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Haplotype 1	A	A	G	T	C	A	G	G	C	A	C	T	C	A	A	G
Haplotype 2	G	G	T	C	A	G	A	T	C	A	C	T	C	A	A	G
Haplotype 3	A	A	G	T	C	A	G	G	T	G	G	C	A	G	C	T
Haplotype 4	G	G	T	C	A	G	A	T	C	A	G	C	A	G	C	T
Haplotype 5	A	A	G	T	C	A	A	T	C	A	C	T	C	A	A	G
Haplotype 6	G	G	T	C	A	G	G	G	T	G	C	T	C	A	A	G

Figure 1.3. Mechanism describing how different haplotypes are generated by recombination. From Jameson and Kopp 2014.

The publication of the first human genome sequence in 2001 represented a cornerstone in the scientific history and probably marked most molecular biological research on the current century (Lander et al. 2001; Venter et al. 2001). It not only symbolized a great scientific achievement, but also prepared the ground for a new era on biology, establishing the basis of the many “*-omics*” sciences. In spite of the success of the Human Genome Project, it is only with advent of the high-throughput genotyping technologies that it started to be feasible to reveal genome wide patterns of human variation at population level (Botstein and Risch 2003). These methodological advances led to the organization of the International Haplotype Project, whose aim was to obtain and describe a high resolution haplotype map across different populations (International HapMap Consortium 2005). Simultaneously, Genome Wide Association Studies (GWAS) became the standard methodology to investigate the association between genotype and phenotype and facilitated the identification of susceptibility variants for many different human diseases and polygenic traits. All the variants significantly associated to human phenotypes are stored in public databases (Hindorff et al. 2009; Welter et al. 2014) and represent an exceptional resource for physicians and biologists to bridge the gap between genotype and phenotype. Even though GWAS have several well-known limitations, such as the type of variants investigated (mostly limited to SNPs) or the lack of direct functional evidences to explain the disease aetiology, the GWAS era identified many different new susceptibility loci and provided important insights on the genetic architecture of human complex diseases (Visscher et al. 2012). With the publication of the pilot release of the 1000 Genomes Project in 2010, the first complete description of human genetic variation within and among ethnic groups was available (Durbin et al. 2010). This impressive project, involving thousands of researchers worldwide, was aimed to discover >95% of the human variation with minor allele frequency as low as 1%. Due to the technological and methodological challenges, the 1000 Genomes Project not only represented a huge resource of biological data, but also provided methodological standards and novel bioinformatic tools that the whole scientific community is taking advantage of (Clarke et al. 2012). The new catalog of genetic variation is currently having a strong impact on many different aspects of human biology. For instance, many different studies have used these data to make inferences about the selective and demographic processes affecting the re-sequenced populations (The 1000 Genomes Project Consortium 2015). The discovered catalog of variants indicates that most of the variation in the genome is rare and that rare variants are enriched of functionally relevant alleles (Zhu et al. 2011; Tennessen et al. 2012). Moreover, it has been observed that each single person carries on average around 2,500 non-synonymous variants at conserved positions, 250-300 loss of function variants and 50-100 variants previously described to be associated to human disorders (The 1000 Genomes Project Consortium 2012).

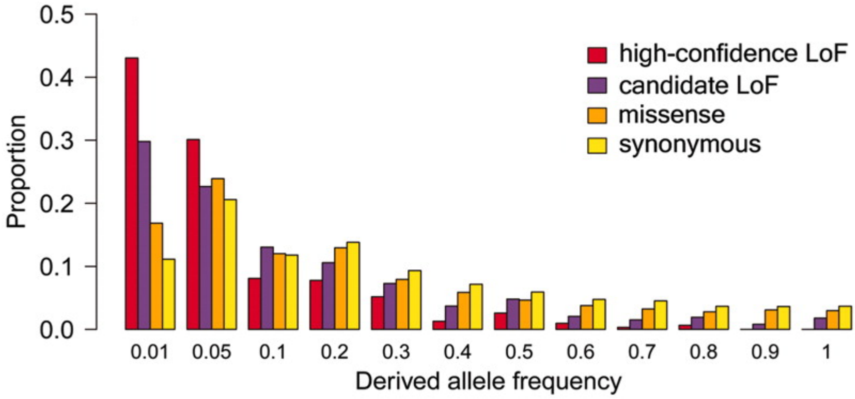


Figure 1.4. Unfolded site frequency spectrum of different functional classes of SNPs. From MacArthur et al. 2012.

Finally, the constant dropping of re-sequencing costs allow researchers to investigate variation in specific groups of samples in search of the functional “smoking gun” for human traits of special interest. This dispersed and chaotic generation of human variation data, deposited in databases such as the European Genome-Phenome Archive (EGA) (Lappalainen et al. 2015) and the Sequence Read Archive (SRA) (Leinonen et al. 2011), furnishes an extraordinary amount of biological information. However, the interpretation of data originated under specific experimental conditions and technologies poses novel theoretical challenges that the scientific community needs to face.

1.3 Theoretical framework for genetic diseases.

The first theoretical effort to explain the occurrence of human diseases, and more specifically infectious pathologies, dates back to the ancient Greek world and in particular to the physician Galen of Pergamon (129-216 AD). His “miasmatic theory” held that diseases were caused by “bad air”, originated from rotten organic matter, and remained a valid explanation until the first part of the 19th century.

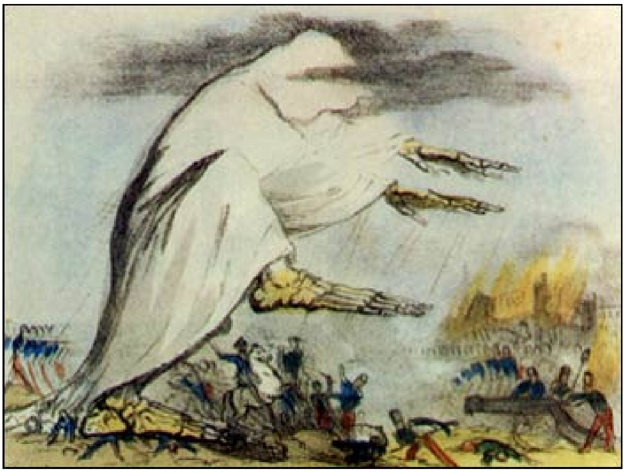


Figure 1.5. Representation by Robert Seymour depicting the spread of cholera outbreak of the 19th century in the form of poisonous air.

Only with the epidemiological work of Snow during the cholera outbreak in London in 1854 (Parkes 2013) and with the experimental work of Pasteur and Koch (Pasteur 1881; Koch 1890), the “germ theory of diseases” was proposed and widely accepted, leading to the “golden age” of bacteriology and to the identification of many relevant infectious agents.

The whole spectrum of human diseases not only includes pathologies mediated by infectious agents, but also disorders that “run in the family”. The pioneering theory of Maupertius in 1751 regarding the “particulate concept of heredity” (Glass 1947), the subsequent observation in 1753 of polydactyly in the Ruhe family and the first estimation of the likelihood for it to be hereditary

(Glass 1947), could be considered the dawn of medical genetics. In the following decades, the works of Dalton, Martin and Otto regarding daltonism, hereditary blindness and haemophilia, respectively, gave a further determinant contribution to the understanding of the familial clustering of many human diseases (Hunt et al. 1995; Martin 1809; Parrish 1845). Only in 1814, the first general statements regarding the mechanisms and features of hereditary diseases were made in a broader context. In his work, Adams formulates the concepts of “family” and “hereditary” disorders, corresponding to autosomal dominant and recessive diseases. Additionally, he described the general features of these classes of diseases and he also suggested that particular demographic processes, such as inbreeding and the founder effect, could be at the basis of the familial and geographical clustering of certain diseases (Adams 1816). Unfortunately, due to the lack of scientific evidence, these ideas were largely ignored by his contemporaries. During the whole 19th century a great amount of new evidences contributed to the understanding that heredity is behind many human diseases. Lamentably, the enormous implications of Mendel work between 1851 and 1863 on pea plants and the establishment of the general laws of heredity (Mendel 1951) were neglected by the scientific community. Only at the beginning of the 20th century, with the re-discovery of Mendel’s work (Vries 1901), the first proof of a human disease transmitted according to Mendelian rules of inheritance was provided by Garrod in 1902. In his work, the British physician confirmed that alkaptonuria is caused by a recessive mutation (Garrod 1902), which was later demonstrated to be located in the *HGD* gene and to produce an alteration of the phenylalanine and tyrosine processing (Fernandez-Canon et al. 1996; La Du et al. 1958). With the formulation of the “Chromosome theory of heredity” by Sutton and Boveri (Sutton 1903; Boveri 1904) and the experimental observations of Morgan regarding sex-linked traits on *Drosophila melanogaster* (Morgan 1910), modern medical genetics arose as a new science. Meanwhile, Hardy and Weinberg described and elaborated the first mathematical explanations for evolutionary processes (Hardy 1908; Weinberg 1908), while Fisher, Haldane and Wright with their “modern synthesis” revolutionized evolutionary biology, reconciling Mendelian genetics with gradual evolution by means of natural selection (Fisher 1930; Haldane 1934; Wright 1932).

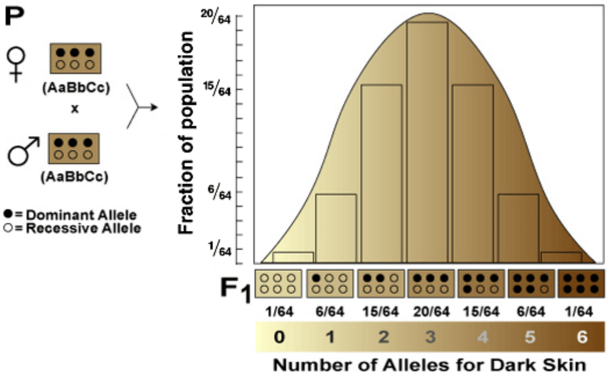


Figure 1.6. Example of polygenic Mendelian inheritance for a quantitative trait, under the assumption that loci have a cumulative effect and each locus contributes equally to the phenotype.

These discoveries led to the formulation of a general theory to explain the presence and persistence of deleterious variants in the human genome. On the basis of the "mutation-selection balance" model, alleles underlying human diseases are introduced in the genetic pool of variation by random mutations and eventually eliminated by purifying selection, as this is the evolutionary force that keeps deleterious alleles at low frequencies within populations (Haldane 1949b).

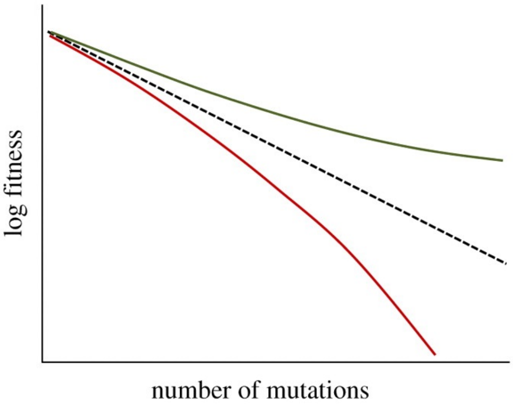


Figure 1.7. Relationship between fitness and mutation load. Dashed line indicate absence of epistasis among mutations; green and red lines represent examples of positive and negative epistasis, respectively. From de Visser, Cooper, and Elena 2011.

A proof of this process is given by the high allelic heterogeneity and the low disease prevalence characterizing many single gene disorders with Mendelian inheritance (Reich and Lander 2001). Even if valid for a wide range of diseases, the "mutation-selection balance" model is not able to explain the whole spectrum of human genetic disorders. Already in the 40s, Haldane introduced the concept of "selective shadow" to explain why the dominant mutation causing Huntington's disease was not eliminated by the action of natural selection. Since the causal allele has a detrimental effect after the reproductive age, it would be invisible to the action of purifying selection, persisting in the population and eventually increasing in frequency by drift (Haldane 1941).

In 1949, the molecular bases of sickle cell anemia were discovered by Pauling (Pauling et al. 1949) and Haldane suggested an evolutionary explanation for the persistence of the corresponding causal alleles (Haldane 1949a). In 1954, Allison demonstrated that the frequency of sickle cell anemia mapped out very similarly to the distribution map for the most severe forms of malaria, proving that the heterozygous advantage against malaria infection was maintaining causal alleles for sickle cell anemia at high frequency (Allison 1954).

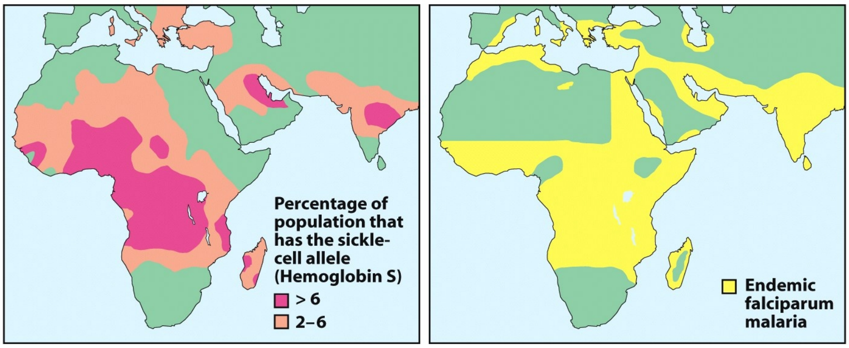

Percentage of population that has the sickle-cell allele (Hemoglobin S)
> 6
2–6
Endemic falciparum malaria

Figure 1.8. Geographical correlation between sickle cell anemia prevalence and endemic Malaria infection.

These evidences had enormous theoretical implications for the understanding of the evolutionary forces driving the frequencies of alleles associated to disease. Adaptive processes could also target strongly deleterious variants, suggesting that alternative models to the “mutation-selection balance” could better explain the observed prevalence of some human genetic disorders.

For complex polygenic disease traits, it is even more difficult to conceive a general evolutionary model. Usually, little is known on epistasis, that is, how the different genetic components interact among them (Moore 2003); on the contribution of environmental factors and on their interaction with genotype (Hunter 2005). The first attempts to explain complex diseases in an evolutionary perspective were proposed by Medawar and Williams in the 50’s (Medawar 1952; Williams 1957). The “mutation accumulation” theory of Medawar properly formulated the concept of selective shadow already proposed by Haldane a decade before (Haldane 1941). The decline of the action of natural selection after the end of reproductive age causes an accumulation of variation with detrimental effects on late stages of life, resulting in a deterioration of health conditions in old age. In its theory of senescence, William proposed an adaptive framework to explain senescence and the high prevalence of polygenic diseases after the reproductive period ends. The “antagonistic pleiotropy” hypothesis states that senescence could be due to alleles that are advantageous earlier in life and thus positively selected, but detrimental after the reproductive age. While Medawar suggested that late onset complex diseases could be mainly driven by drift, William proposed adaptation as the major force to explain the high prevalence of such class of diseases.

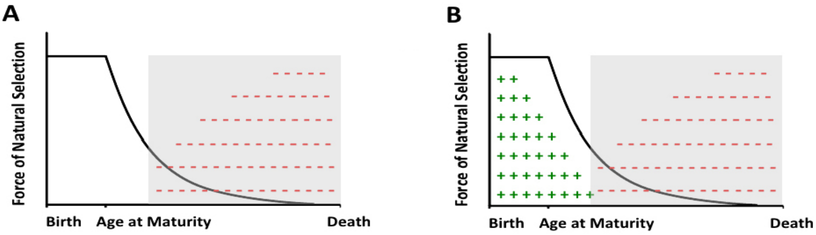
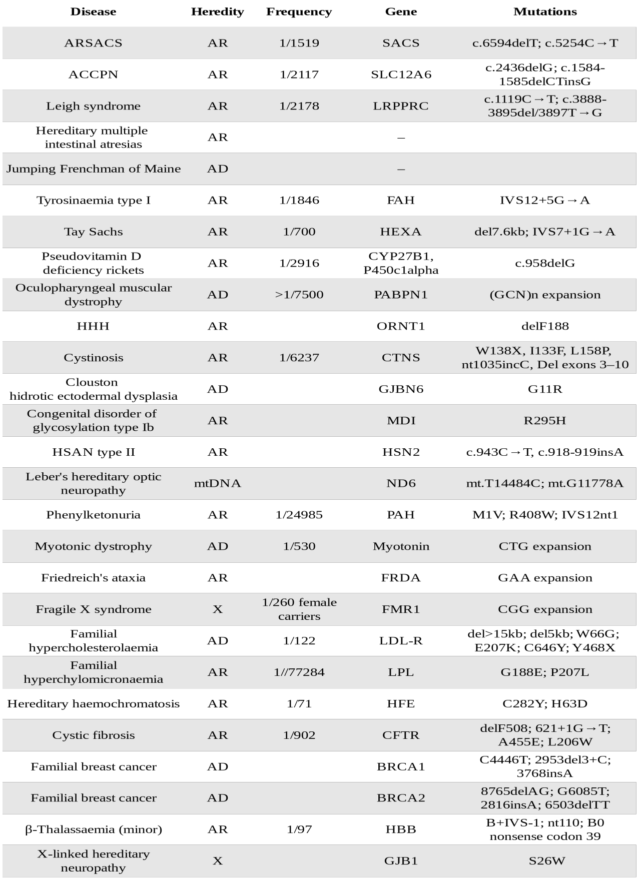


Figure 1.9. Schematic representation of mutation accumulation theory (A) and antagonistic pleiotropy theory (B). From Fabian and Flatt 2011.

Additional theories and hypotheses were then formulated to explain the persistence of specific diseases in human populations. For instance, the "hygiene hypothesis" states that the increasing incidence of autoimmune disorders in modern societies could be the result of the recent advances in hygiene and medical practices. The lack of exposure to infectious agents and parasites during childhood could lead to a defect in the establishment of a proper immune tolerance (Strachan 1989). Conversely, the "sodium retention" and the "thrifty genotype" hypotheses point to other phenomena that could be relevant for the dynamics of deleterious alleles. The two hypotheses suggest that an ancestral protective allele could be reverted to a deleterious one in the modern conditions. In particular, the "thrifty genotype" was advantageous in the past, facilitating the accumulation of acid fats to face starvation. In the modern conditions of western societies individuals carrying the protective allele would accumulate a reserve of energy for a famine that they will never face, resulting in an excess of acids fats having detrimental effects (Neel 1962). Similarly, in the “sodium retention” hypothesis alleles that in the past favored the accumulation of sodium in the original Western Africa populations are currently the major cause of the higher incidence of hypertension and cardiovascular diseases in African Americans. It is supposed that during the trip from Africa to the Americas in the slave trade, the major causes of death were salt-depletive diseases such as diarrhea, fevers and vomiting. Individuals with an enhanced genetic ability to conserve salt had a distinct survival advantage over others and were, therefore, more likely to transmit their genotype to the subsequent generations. In the present conditions, the previously advantageous genotype results in a higher incidence of hypertension and related disorders in their descendants (Wilson 1986). Even if these fascinating hypotheses have not been directly demonstrated, as a whole they prepared the ground for evolutionary medicine, whose objective is the understanding of why humans get sick and not only how. Besides ancient demographic history and the inter-breeding events with sister species described in section 1.1, also recent demographic events play an important role in defining the tendency of specific populations to be affected by certain diseases. The population bottleneck suffered by the ancestors of modern Ashkenazi Jews could explain the high prevalence of many diseases, such as Gaucher’s disease, dysautonomia, Tay-Sachs disease, cystic fibrosis and many others in that specific community

(Risch et al. 2003). Similarly, the current 6 millions French-Canadians living in Quebec originated from approximately 8,500 settlers who colonized the region between 1608 and 1759. The low effective population size of the founder population and the subsequent reduced efficiency of natural selection is at the basis of the high prevalence of many rare genetic disorders with Mendelian inheritance (Laberge et al. 2005).



Disease	Heredity	Frequency	Gene	Mutations
ARSACS	AR	1/1519	SACS	c.6594delT; c.5254C → T
ACCPN	AR	1/2117	SLC12A6	c.2436delG; c.1584-1585delCTinsG
Leigh syndrome	AR	1/2178	LRPPRC	c.1119C → T; c.3888-3895del/3897T → G
Hereditary multiple intestinal atresias	AR		–	
Jumping Frenchman of Maine	AD		–	
Tyrosinaemia type I	AR	1/1846	FAH	IVS12+5G → A
Tay Sachs	AR	1/700	HEXA	del7.6kb; IVS7+1G → A
Pseudovitamin D deficiency rickets	AR	1/2916	CYP27B1, P450c1alpha	c.958delG
Oculopharyngeal muscular dystrophy	AD	>1/7500	PABPN1	(GCN)n expansion
HHH	AR		ORNT1	delF188
Cystinosis	AR	1/6237	CTNS	W138X, I133F, L158P, nt1035incC, Del exons 3–10
Clouston hidrotic ectodermal dysplasia	AD		GJBN6	G11R
Congenital disorder of glycosylation type Ib	AR		MDI	R295H
HSAN type II	AR		HSN2	c.943C → T, c.918-919insA
Leber's hereditary optic neuropathy	mtDNA		ND6	mt.T14484C; mt.G11778A
Phenylketonuria	AR	1/24985	PAH	M1V; R408W; IVS12nt1
Myotonic dystrophy	AD	1/530	Myotonin	CTG expansion
Friedreich's ataxia	AR		FRDA	GAA expansion
Fragile X syndrome	X	1/260 female carriers	FMR1	CGG expansion
Familial hypercholesterolaemia	AD	1/122	LDL-R	del>15kb; del5kb; W66G; E207K; C646Y; Y468X
Familial hyperchylomicronaemia	AR	1//77284	LPL	G188E; P207L
Hereditary haemochromatosis	AR	1/71	HFE	C282Y; H63D
Cystic fibrosis	AR	1/902	CFTR	delF508; 621+1G → T; A455E; L206W
Familial breast cancer	AD		BRCA1	C4446T; 2953del3+C; 3768insA
Familial breast cancer	AD		BRCA2	8765delAG; G6085T; 2816insA; 6503delTT
β-Thalassaemia (minor)	AR	1/97	HBB	B+IVS-1; nt110; B0 nonsense codon 39
X-linked hereditary neuropathy	X		GJB1	S26W

Table 1.1. List of rare diseases more frequent or with distinctive features in French-Canadian population. ACCPN,

agenesis of corpus callosum and peripheral neuropathy; ARSACS, autosomal recessive spastic ataxia of Charlevoix–Saguenay; HHH, hyperornithinaemia–hyperammonaemia–homocitrullinuria; AR, autosomal recessive; AD, autosomal dominant. Adapted from Laberge et al. 2005.

Thus, alleles involved in both monogenic and polygenic diseases and therefore probably affecting fitness are not only subject to purifying selection as originally stated by the “mutation-selection balance” model. Environmental factors and demographic processes interplay with the genetic background and contribute in shaping the frequency of deleterious alleles. Therefore, an evolutionary framework is mandatory for the complete understanding of the genetic architecture of human diseases and to explain their persistence in current human populations. The “antagonistic pleiotropy”, the “hygiene”, the “thrifty genotype” and “sodium retention” hypotheses represent the effort to explain human pathologies through an evolutionary framework and suggest that human diseases should be regarded as the dark side of evolution, adaptation and technological progress. Quoting the famous sentence of Dobzhansky: “Nothing in biology makes sense except in the light of evolution” (Dobzhansky 1973); certainly human diseases are not an exception to this general concept and modern evolutionary medicine represent an attempt for their explanation. In spite of the achieved goals, a lot of theoretical and experimental work is still needed.

1.4 Classification of human diseases.

A disease is an abnormal condition of a part, organ, or system of an organism resulting from various causes, such as infection, inflammation, environmental factors, or genetic defect, and characterized by an identifiable group of signs, symptoms, or both. From a medical point of view, diseases can be classified by aetiology (the cause), pathogenesis (the mechanism), by symptoms and by the organ involved. Due to the unknown aetiology and pathogenesis, often classification is not trivial and frequently the diagnostic terms are used to classify the condition. A further level of complexity is given by the social interpretation of what is a pathological condition. For instance, obesity can represent a status symbol in societies facing persistent famine (Haslam and James 2005); in Hmong population epilepsy is considered a sign of spiritual gift (Fadiman 1998). In addition, what is definitely considered a disease in modern societies could not be in ancient communities. It has been hypothesized that mental conditions, such as schizophrenia and bipolar disorder, could be widely socially accepted in the past and that affected individuals could fulfill relevant social and religious roles and thus have higher probability to reproduce and pass the corresponding susceptibility variants to the offspring (Preti and Miotto 1997; Polimeni and Reiss 2002).

An alternative way to classify diseases is provided by the genetics underlying human pathologies. The classical dichotomous classification separates disorders in two main categories: Mendelian monogenic and complex polygenic diseases.

1.4.1 Mendelian diseases.

Mendelian disorders constitute the simplest category of genetic diseases and generally a deleterious variant in a single gene is the direct cause. The ~5,000 described monogenic pathologies are generally rare and collectively affect ~0.4% of the total population (Chong et al. 2015), ranging from 5 cases each 1,000 births for the familial combined hyperlipidemia (Gaddi et al. 2007) to the single worldwide case of ribose 5-phosphatate isomerase deficiency (Wamelink et al. 2010).

AD Genetic Disorder	Frequency Per 1000 births	AR Genetic Disorder	Frequency Per 1000 births
Familial combined hyperlipidemia	5.0	Cystic fibrosis	0.4
Familial hypercholsterolaemia	2.0	alpha-1-antitrypsin deficiency	0.2
Dominant otosclerosis	1.0	Phenylketonuria	0.1
Adult polycystic kidney disease	0.8	Congenital adrenal hyperplasia	0.1
Multiple exostoses	0.5	Spinal muscular atrophy	0.1
Huntington's disease	0.5	Sickle cell anaemia	0.1
Neurofibromatosis	0.4	Beta-Thalassaemia	0.05
Myotonic Dystrophy	0.2		
Congenital spherocytosis	0.2		
Polyposis coli	0.1		

Table 1.2. Frequency of the most common autosomal dominant (AD) and recessive (AR) Mendelian disorders in UK population. Data obtained from Genetic Alliance UK.

Usually monogenic diseases tend to be very severe, have an early onset and a strong impact on the lifestyle of affected individuals. These disorders run in families following classic Mendelian inheritance rules and the pedigree analysis of affected families reveals the exact mechanism of transmission, that is, whether the causal allele is dominant or recessive and whether it is located in the autosomal or sexual chromosomes. The accurate analysis of the inheritance transmission patterns and the availability of human genetic and recombination maps allowed the development of statistical methods for the discovery of genomic regions harboring the causal mutations. When for each individual of a pedigree the genotypes at several specific genetic markers are known, it is possible to calculate the likelihood (LOD score) for a given marker to be in linkage at a certain genetic distance from the causal allele (Botstein et al. 1980).

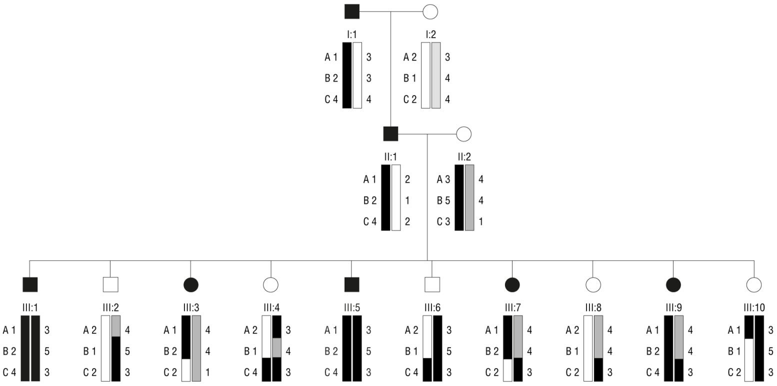

I:1
A 1 3
B 2 3
C 4 4
I:2
A 2 3
B 1 4
C 2 4
II:1
A 1 2
B 2 1
C 4 2
II:2
A 3 4
B 5 4
C 3 1
III:1
A 1 3
B 2 5
C 4 3
III:2
A 2 4
B 1 5
C 2 3
III:3
A 1 4
B 2 4
C 2 1
III:4
A 2 3
B 1 4
C 4 3
III:5
A 1 3
B 2 5
C 4 3
III:6
A 2 3
B 1 5
C 4 3
III:7
A 1 4
B 2 4
C 2 3
III:8
A 2 4
B 1 4
C 2 3
III:9
A 1 4
B 2 4
C 4 3
III:10
A 1 3
B 1 5
C 2 3

Figure 1.10. Example of three generation pedigree segregating an autosomal dominant trait. The linkage mapping method suggests that marker B2 is in linkage with causal variant. From Pulst 1999.

This methodology permitted the identification of many genomic regions of interest and the subsequent discovery of the corresponding causal variants, with relevant implications for the understanding of the biological processes underlying such Mendelian disorders. Most of the genetic variants that cause Mendelian diseases evolve under strong purifying selection due to the impairment on affected individuals and the consequent reduced fitness. Such causal alleles are thought to be introduced in our genomes by random mutations and maintained at low frequency in the population through the “mutation-selection balance” model (Reich and Lander 2001). But for any biological phenomenon, the exception is the only general rule. For some diseases with autosomal recessive transmission, it has been proved that the causal allele confers an advantage in specific environmental conditions. As described above, the causal variant for sickle cell anemia confers protection against malaria infection (Allison 1954); the mutated version of the *CFTR* gene causal of cystic fibrosis is thought to protect against dehydration caused by cholera infection (Gabriel et al. 1994) or tubercolosis (Poolman and Galvani 2007). Particular mutations on the triosephosphate isomerase gene cause an autosomal recessive enzymatic deficiency (Schneider et al. 1965), while other forms of the gene (null alleles) are lethal when found in homozygous state (Merkle and Pretsch 1989). Surprisingly, the frequency of heterozygous individuals for the null alleles is much higher than what is expected by chance (Schneider et al. 1996), leading to the suggestion that a major resistance to oxidative stress could explain such unexpected frequency (Ralser et al. 2006). Despite their simple transmission mechanism, the discovery of the genetic factors causal of Mendelian diseases is not always straightforward. Incomplete and age-dependent penetrance, phenotypic heterogeneity, incomplete dominance and codominance could lead to an erroneous interpretation of the transmission pattern and make vaguer the distinction between Mendelian and complex diseases.

1.4.2 Complex diseases.

The aetiology of complex polygenic diseases is the result of the interplay of several genetic and environmental factors interacting among them and resulting in a pathological phenotype only under certain conditions (Hunter 2005). Complex diseases affect a considerable proportion of the whole population and tend to manifest at late ages (Wright et al. 2003). In addition, a high level of phenotypic heterogeneity is typically observed and diagnosis is usually made considering both inclusion and exclusion criteria. Nevertheless, as monogenic disorders, complex diseases also tend to show familial clustering, even if they do not follow the classic rules of Mendelian inheritance (van Heyningen and Yeyati 2004). This non-regular familial clustering provides important information to disentangle the relevance of the genetic components contributing to such complex diseases. The analysis of phenotypic correlation within families at different relationship scales (siblings, cousins, second cousins, etc.) gives insight into the level of heritability of a given condition (Botstein and Risch 2003). Heritability is the proportion of phenotypic variance in any particular trait that is due to genetic factors. While the broader-sense heritability (H^2) captures the contribution of any genetic factor influencing the phenotype, including gene-gene interactions; the narrow-sense heritability (h^2) estimates the fraction of variation attributable to additive genetic factors (Visscher, Hill, and Wray 2008).

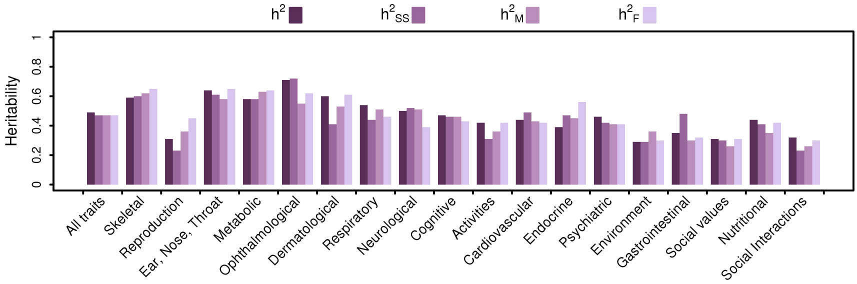


Figure 1.11. Meta-analysis estimates of heritability for 17,804 traits considering 2,748 publications regarding monozygotic and dizygotic twin studies. Details regarding traits categorization and global heritability estimates are in the original publication (Polderman et al. 2015). h^2, heritability on the whole set of twin pairs; h^2_{SS} heritability based only on same sex twin pairs; h^2_M heritability based only on males twin pairs; h^2_F heritability based only on females twin pairs.

The heritability of human diseases ranges widely, from ~1% for leukemia and stomach tumors (Czene, Lichtenstein, and Hemminki

2002) to 76% and 81% for chronic obstructive pulmonary disease and schizophrenia, respectively (Ingebrigtsen et al. 2010; Sullivan, Kendler, and Neale 2003). Due to the variable expressivity, the low penetrance of the increasing risk variants, the variable role of the environmental factors and the multifactorial nature of complex diseases, linkage analysis and positional cloning were not successful as for Mendelian disorders (Bush and Haines 2001). In contrast to strategies based on genome wide linkage analysis, that are able to detect causal variants without any *a priori* knowledge of the disease biology, approaches based on candidate genes require at least a minimal understanding of the biological processes underlying the disorders. Candidate genes methodologies represent a biological guess, but in contrast to linkage studies, a case-control design involving unrelated samples can be used to test genetic association (Tabor, Risch, and Myers 2002). Even if these candidate gene approaches provided insight for the understanding of complex diseases, many spurious associations due to inappropriate samplings or others factors influencing individual’s genetic composition were subsequently demonstrated (Hutchison et al. 2004; Ioannidis 2005; Hirschhorn et al. 2002). Only with the advent of high-throughput genotyping technologies it has been possible to join the advantage of genome wide analysis, as in linkage mapping methods, with the opportunity to recruit unrelated samples, as in the case-control setting studies (Kilpinen and Barrett 2016).

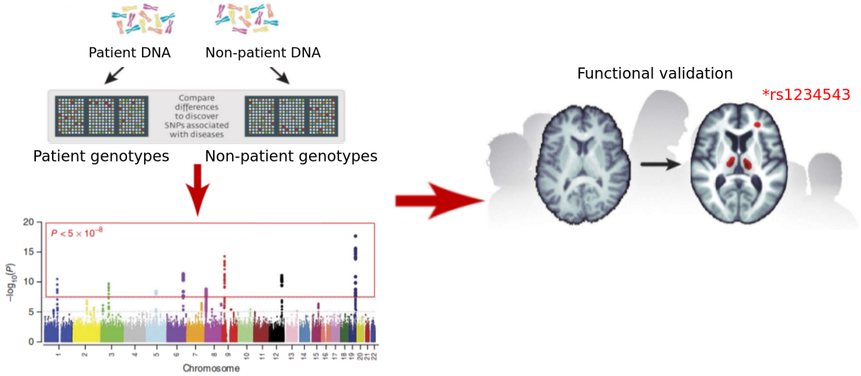

Patient DNA
Non-patient DNA
Compare differences to discover SNPs associated with diseases
Patient genotypes
Non-patient genotypes
Functional validation
*rs1234543
$P < 5 \times 10^{-8}$
20
15
10
5
0
$-\log_{10}(P)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Figure 1.12. Schematic representation of Genome Wide Association Studies (GWAS) work-flow. Adapted from Medland et al. 2014.

Around 5,000 significant associations between SNPs and human phenotypes have been discovered by Genome Wide Association Studies (GWAS) and are currently stored in the GWAS catalog database (Hindorff et al. 2009; Welter et al. 2014).

Figure 1.13. Diagram of all SNP-trait association with P-value ≤5.0×10^{-8} published in the GWAS catalog.

The lack of a precise definition of cases and controls for some traits, the multiple testing correction and the population stratification represent well-known limitations that can be corrected by a proper study design (Pearson and Manolio 2008). In addition to these preventable issues, GWA studies rely on the assumption that common genetic variation (Minor Allele Frequency >5%) plays a major role in explaining the heritable factors underlying complex diseases. The frequencies of increasing risk alleles depend on the evolutionary forces acting on the resulting phenotype. The Common Disease-Common Variants (CD-CV) hypothesis suggests that the numerous susceptibility variants, with small additive or multiplicative effects, are evolutionarily neutral and could increase in frequency by chance (Reich and Lander 2001; Pritchard and Cox 2002). Unluckily, GWAS hits explain only a fraction of the total phenotypic variance for most traits, leading to the so-called problem of the “missing heritability” (Manolio et al. 2009). For instance, using high quality genotypes of 139,576 SNPs for 4,414 samples from the Wellcome Trust Case Control Consortium (WTCCC), the explained heritability estimates by GWAS hits for Coronary Heart Disease (CAD) range from 20% to 32%, if using imputed SNPs from 1000 Genomes Project and/or correcting for LD (Gusev et al. 2013).

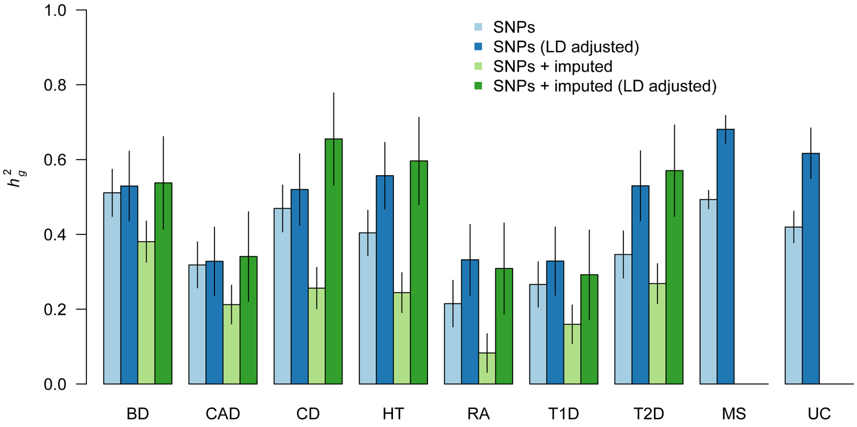


Figure 1.14. Heritability explained by GWAS hits for 8 different traits. Bipolar Disorder (BD), Coronary Artery Disease (CAD), Crohn's Disease (CD), Hypertension (HT), Rheumatoid Arthritis (RA), Type 1 Diabetes (T1D), Type 2 Diabetes (T2D), Multiple Sclerosis (MS), Ulcerative Colitis (UC). From Gusev et al. 2013.

In alternative, the Common Disease-Rare Variants (CD-RV) hypothesis suggests that most of the phenotypic variance of complex diseases should be due to rare variants (Minor Allele Frequency <1%) with moderate effects and recently originated in the population (Cirulli and Goldstein 2010). Due to their low frequency, rare variants are generally not included in commercial genotyping arrays and huge sample sizes are required to detect significant associations with a given trait. Nowadays, NGS technologies allow to capture the entire allelic frequency spectrum and permit to investigate the role of rare variants in human diseases.

Since the standard association tests used for common variants are clearly underpowered for rare variants, alternative statistical strategies emerged to detect associations. Rather than testing each single variant independently, the so-called “collapsing methods” consider jointly multiple variants within a given genomic region and compare their global burden between cases and controls (Asimit and Zeggini 2010; Bansal et al. 2010).

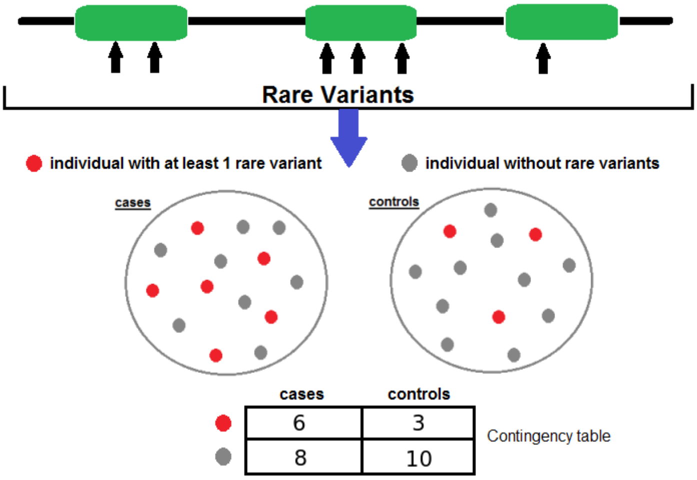

Rare Variants
individual with at least 1 rare variant
individual without rare variants
cases
controls
cases
controls
6
3
8
10
Contingency table

Figure 1.15. Schematic representation of “collapsing methods” statistical approach.

Test name	Description
ASUM	In the Adaptive Sum score test each variant site is evaluated for excess of minor alleles in controls and genotype codings are flipped. In a second stage a burden test is performed.
C-alpha	C-alpha tests for deviation of variance of minor allele counts in cases/controls.
CMC	CMC test collaps rare variants into a single indicator variable and performs a multivariate test to compare cases and controls.
KBAC	In the Kernel Based Adaptive Clustering method genotype frequencies are weighted by a hypergeometric density kernel function to test for differences between cases and controls.
RBT	In the Replication Based Test variant sites are scored based on -log transformation of probability of having more than observed variants in cases than in controls. The RBT statistic is defined as sum of the variant sites scores.
VT	In the Variable Thresholds method the burden test statistic of a group of variants will be maximized over subsets of variants defined by applying different minor allele frequency thresholds.
WSS	In the Weighted Sum method rare variant counts for each individual are accumulated and a weigthed term is applied to emphasize alleles with a low frequency in controls. The scores for all samples are ordered and the WS score is computed as the sum of ranks in cases.

Table 1.3. Description of different available “collapsing methods”.

The competing CD-CV and CD-RV hypotheses provide a valid theoretical framework for specific diseases (Gibson 2012), but emerging evidences suggest that both can be true even at a given locus. In particular, for Parkinson’s disease (PD) it has been demonstrated that both common and rare genetic variants have a role in its aetiology (Singleton and Hardy 2011). A rare missense mutation in the *SNCA* gene was discovered as the first genetic factor responsible for a Mendelian form of PD (Polymeropoulos et al. 1997), but several common variants affecting the expression of the same gene have also been reported as increasing risk factors for sporadic PD (Simon-Sanchez et al. 2009).

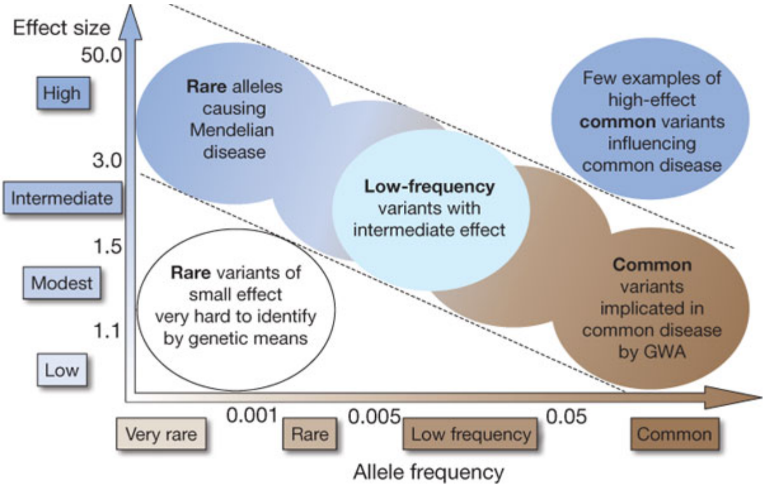

Effect size
50.0
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles causing Mendelian disease
Few examples of high-effect common variants influencing common disease
Low-frequency variants with intermediate effect
Rare variants of small effect very hard to identify by genetic means
Common variants implicated in common disease by GWA
0.001
0.005
0.05
Very rare
Rare
Low frequency
Common
Allele frequency

Figure 1.16. Power to detect variants depending on their frequency and effect on the phenotype. From Manolio et al. 2009.

1.5 Biological properties of genes related to human diseases.

The knowledge of the biological properties of genes related to human diseases is at the core of modern medical genetics. The understanding of their molecular functions and interactions is mandatory to gain a detailed comprehension of the pathological mechanisms and for the subsequent design of new therapeutic strategies. Association studies provided a valuable but limited contribution for the description of increasing risk variants for common diseases (Manolio et al. 2009). Additional insights come from the recent advances in sequencing technologies; whole genome, whole exome and target sequencing approaches have been successfully applied for the identification of rare causal variants using family design strategies (Ng et al. 2010; Proukakis et al. 2013). The characterization of genomic loci, discovered through different methodologies, resulted in the creation of large catalogs of variants related to human diseases. The Online Mendelian Inheritance in Man (OMIM) database (http://www.omim.org/) (Hamosh et al. 2005) and the Genome Wide Association Studies (GWAS) Catalog (https://www.ebi.ac.uk/gwas/) (Hindorff et al. 2009; Welter et al. 2014) are specifically designed to report mutations causing Mendelian hereditary disorders and susceptibility loci contributing to complex diseases, respectively.

Many authors have tried to characterize the biological and evolutionary properties of genes harboring causal variants or increasing risk factors for human disorders. At the beginning of the current century, the clamorous explosion of information technology strongly impacted genetics and biology, providing an extraordinary opportunity for large-scale analysis and comparisons through computational methods.

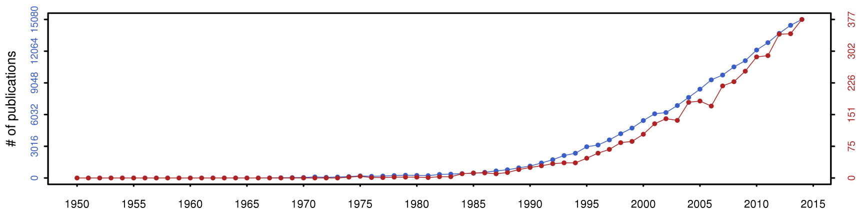


Figure 1.17. Number of articles indexed in PubMed by year when querying “Human disease genes” (blue) and “Human disease genes properties” (red).

In earlier studies, the causal variants of traits following Mendelian inheritance patterns were compared to genomic positions not involved in disease. For instance, Miller et al. compared the causal non-synonymous variants in 6 genes to the rest of non-synonymous variants in the same genes. Causal mutations were occurring in conserved regions and the resulting amino-acid changes were generally not observed along large phylogenetic trees. In addition, the disease causing amino-acid changes were biochemically more different compared to the rest of non-synonymous substitutions. Collectively, these observations indicated that causal alleles may not be tolerated in living species and may have a strong effect on the altered protein (Miller and Kumar 2001). Even if these preliminary studies were based in a very small number of genes and variants, the observed general statements were subsequently confirmed when using larger datasets (Subramanian and Kumar 2006; Kumar et al. 2011). Morbid genes are more conserved and evolutionary older than the rest of human genes, suggesting that genes harboring causal variants of Mendelian disorders are mainly target of negative selection (Kondrashov, Ogurtsov, and Kondrashov 2004; Lopez-Bigas and Ouzounis 2004). Effectively, these genes are enriched among those under strong purifying selection (Bustamante et al. 2005); specifically, causal variants of diseases following dominant transmission are located in more conserved and evolutionary older genomic regions than variants related to recessive ones (Blekhman et al. 2008; Furney, Albà, and López-Bigas 2006).

Conflicting results have been reported when comparing genes related to Mendelian disorders to genes increasing risk for complex diseases; some authors suggested higher evolutionary conservation in the former (Blekhman et al. 2008; Kondrashov, Ogurtsov, and Kondrashov 2004), while others reported opposite evidences (Smith and Eyre-Walker 2003; Thomas and Kejariwal 2004; Tu et al. 2006). Similarly, some authors reported higher connectivity in the protein-protein interaction network in the former (Barrenas et al. 2009), while others did not observe any difference, even if both groups were significantly different from genes not involved in human pathologies (Cai, Borenstein, and Petrov 2010).

Interestingly, genes implicated in the same disease tend to share similar network properties, cluster in the same region of the protein-protein interaction network, tend to interact among them, share a tendency to be co-expressed and similar functional categories (Barrenas et al. 2009; Goh et al. 2007). Moreover, most of the human diseases are functionally connected to others, leading to the idea of the existence of distinct and inter-connected disease-specific functional modules (Goh et al. 2007).

1.6 The human disease network.

At the molecular level, genes related to different patho-phenotypes are often functionally related, meaning that is not possible to consider diseases as independent entities. The systematic mapping of such dependencies between morbid conditions has culminated in the concept of the “diseasome”, a network whose nodes are diseases and whose links represent various molecular relationships between the disease-associated cellular components (Goh et al. 2007).

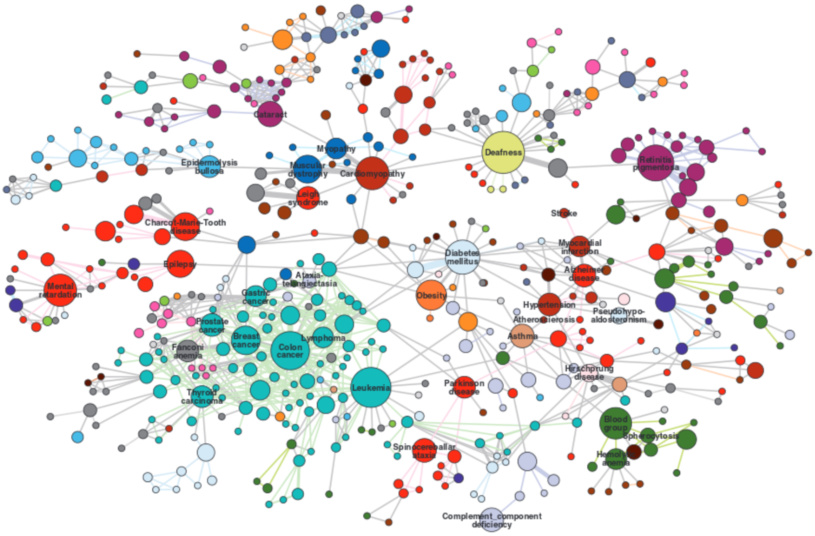

Cataract
Myopathy
Epidermolysis bullosa
Muscular dystrophy
Cardiomyopathy
Leigh syndrome
Deafness
Retinitis pigmentosa
Charcot-Marie-Tooth disease
Stroke
Epilepsy
Mental retardation
Ataxia-telangiectasia
Diabetes mellitus
Myocardial infarction
Alzheimer disease
Gastric cancer
Obesity
Hypertension
Prostate cancer
Atherosclerosis
Pseudohypo-aldosteronism
Breast cancer
Lymphoma
Asthma
Fanconi anemia
Colon cancer
Hirschprung disease
Parkinson disease
Leukemia
Thyroid carcinoma
Blood group
Spherocytosis
Spinocereballar ataxia
Hemolytic anemia
Complement_component deficiency

Figure 1.18. Graphical representation of the human diseasome. Nodes represent disorders and are connected to each other if they share at least one gene in which mutations are associated with both disorders. From Goh et al. 2007.

Uncovering such links not only helps to understand how different phenotypes are related at molecular level, but can also improve the comprehension of why certain groups of diseases arise together, yielding to new approaches for disease prevention, diagnosis and treatment (Barabasi, Gulbahce, and Loscalzo 2011).

The interplay between genes and diseases within the diseasome can take different forms that will be exhaustively described in the following sections.

1.6.1 Relations among human diseases.

The term pleiotropy refers to the functional effects of the same genetic variant on two or more unrelated phenotypes. A typical example of a pleiotropic gene with multiple molecular functions is the serum albumin (*ALB*), whose protein product is well known for binding fatty acids, steroids and toxic metabolites, but it is also involved in the oxidation of nitric oxide (Rafikova, Rafikov, and Nudler 2002). Pleiotropy may result in competing effects, some beneficial and others detrimental for the organism. As described above, in its famous theory of senescence, Williams proposed this antagonistic pleiotropy as the main molecular mechanism underlying the deteriorated health conditions at the latest stage of life (Williams 1957). The relations existing between rare recessive diseases and infections, as for instance those involving blood diseases and malaria (Allison 1954) or phenylketonuria and fungal infections (Woolf 1986), represent other examples of pleiotropic processes. Similarly, some genetic variants have a protective effect for strongly deleterious conditions, but at the same time increase the risk for other genetic diseases. Since it confers protection against cancers, the splice mutation E180 in the growth hormone receptor gene *GHR* is maintained at relatively high frequency in the general population, even if resulting in Laron syndrome when the variant is found in homozygous state (Guevara-Aguirre et al. 2011). Likewise, the high number of CAG repeats within the *HTT* gene are causal of Huntington's disease, while increase the tumor suppressor activity of *p53* (Carter and Nguyen 2011). Mainly, these examples consist on a rare disorder caused by the variant in homozygosity and a more common disease or infection, for which the variant reduces the risk in the heterozygous state. More complex interactions exist between pairs of common diseases, as for instance type 1 diabetes and celiac disease. Type 1 diabetes is caused by an autoimmune reaction against insulin producing β-cells and the disease affects approximately 0.4% of individuals of European ancestry. Celiac disease is also an autoimmune condition, that is the result of an inflammatory reaction in the small intestine and it affects ~0.1% of the European population. The variants located in *RGS1*, *CTLA4*, *SH2B3* and *PTPN2* genes show the same direction of association, increasing risk for both diseases. Conversely, specific alleles in *IL18RAP*, *TAGAP*, *INS*, *IL2RA*, *PTPN22*, *IL12A* and *LPP* genes have distinct effects in the two diseases (Smyth et al. 2008).

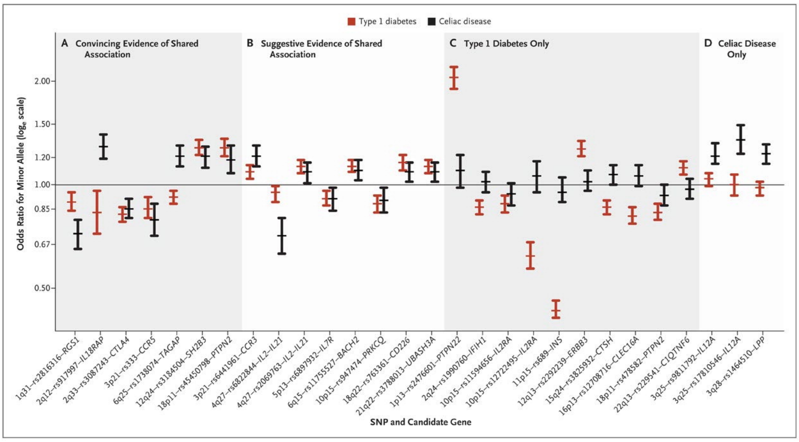


Figure 1.19. Odd-ratio of SNPs associated to either type 1 diabetes, celiac disease or both. From Smyth et al. 2008.

These observations suggest that different diseases may share common pathological processes, giving insight about why different diseases arise together, affecting the same individual. The comorbidity is the presence of one or more additional disorders co-occurring with a primary disease. For instance, according to the Centers for Disease Control and Prevention (CDC), ~60% of individuals affected by arthritis suffer also of heart disease, chronic pulmonary conditions or diabetes. In addition to existing relations between complex diseases, many co-occurrences of Mendelian and complex diseases have also been reported. For example, patients with beta-thalassemia, Huntington disease and Friederich’s ataxia often develop type 2 diabetes mellitus (De Sanctis et al. 1988; Podolsky, Leopold, and Sax 1972; Ristow 2004) and carriers of the genetic variants associated with Lujan-Fryns and Di George syndromes display an increased risk for schizophrenia (De Hert et al. 1996; Sinibaldi et al. 2004). The recent massive collection of clinical records have facilitated the detection of disease pairs significantly co-occurring. Park et al. (Park et al. 2012) analyzed the clinical records of ~13 millions patients stored in the Medicare database, using the relative risk (RR) as a quantitative measure of disease pairs to co-occur compared to random expectation. Similarly, Blair et al. (Blair et al. 2013) mined the electronic medical records, obtained from distinct regions of the United States and Denmark for over 110 million patients and tested for co-occurrences between 65 complex diseases and 95 Mendelian disorders.

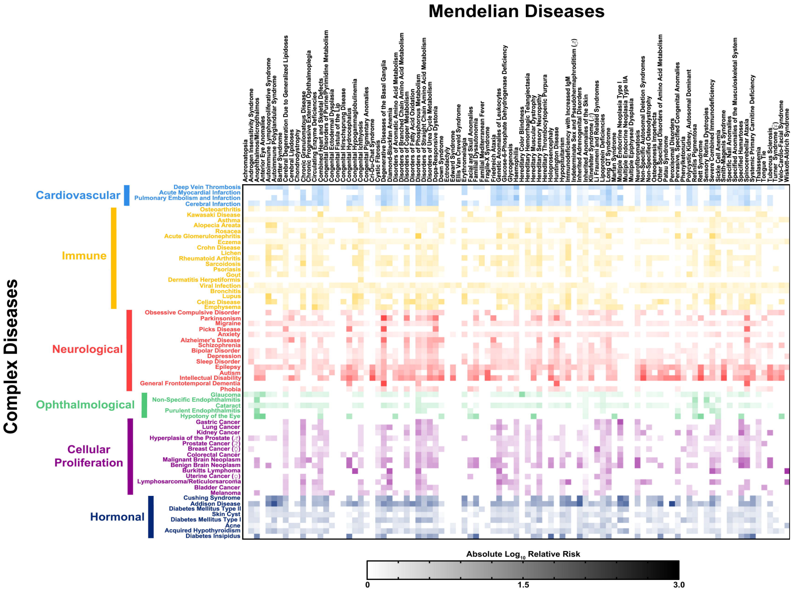


Figure 1.20. Co-occurrences between Mendelian and complex diseases obtained from clinical records. From Blair et al. 2013.

These efforts identified large lists of comorbidities, including relationships among complex diseases, Mendelian disorders and both complex and Mendelian patho-phenotypes. The interplay between human diseases is not a limited phenomenon, but it is widespread through the whole spectrum of diseases. Interestingly, both biological and clinical data unveil the connection between Mendelian monogenic and complex polygenic diseases, contributing to describe the full set of pathological processes underlying human diseases.

1.6.2 Blurring the boundaries between complex and Mendelian diseases.

The interconnected nature of human diseases does not refer only to the tendency of certain pathologies to share similar genetic background. The observed co-occurrence of Mendelian and complex diseases indicates the presence of extensive functional links between the two types of disease. In addition, the same genetic pathology might follow both complex and Mendelian transmission patterns. For instance, about 1-6% of the individuals suffering Alzheimer’s disease (AD) are affected by an early onset form of illness, following classical Mendelian inheritance (Campion et al. 1999). Regardless the contribution of association studies to depict the genomic loci contributing to disease risk (Bertram and Tanzi 2009; Kamboh et al. 2012; Beecham et al. 2014), most of the knowledge regarding AD pathogenesis has been produced by family studies, concerning the Mendelian forms of the disease. The three alleles ε-2, ε-3 and ε-4 of Apolipoprotein E (*APOE*) gene account for ~70% of familial AD cases (Campion et al. 1999). These mutations increase the production of the small protein Aβ42 (Selkoe 1999), which is the main component of the senile plaques, the hallmark of both familial and sporadic forms of AD. Even if *APOE* alleles explain only a small percentage of all AD cases (Campion et al. 1999), their description furnished an extraordinary contribution to formulate new hypotheses about the pathogenesis of the complex forms of the disease. Since then, genes involved in the deposition of abnormally processed amyloid products in the brain are regarded as candidate genes also for the sporadic form of AD.

Some diabetes forms also follow Mendelian inheritance patterns. Maturity onset diabetes of the young (MODY), which accounts for 2-5% of the diabetes cases in western societies, refers to diabetes forms following an autosomal dominant transmission pattern and caused by mutations in various genes. Even if the more common forms of diabetes are the result of complex interactions between several genetics and environmental factors, the pathological disruption of the insulin production is a shared mechanism with the monogenic forms. Mutations in the genes encoding for the enzyme GCK and the transcription factor HNF1A account for the vast majority of the Mendelian forms of diabetes, but also rare mutations in *HNF1α*, *IPF1*, *HNF-1β* and *NEUROD1* genes have been reported, providing additional clues for the complete understanding of the diverse molecular mechanisms interfering with the correct insulin production (Fajans, Bell, and Polonsky 2001).

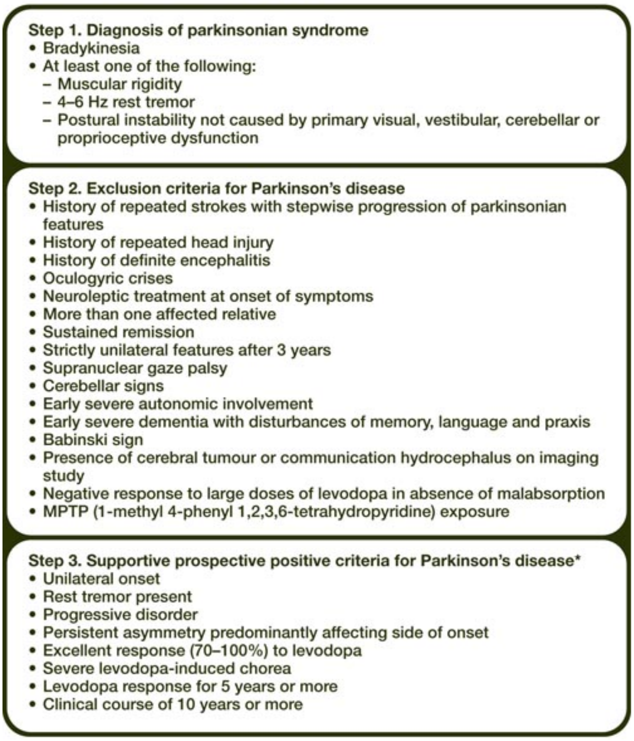
To conclude, the characterization of the genetic background of Mendelian forms of certain diseases has contributed to disclose the pathogenesis of the most common complex forms. Even if this “Mendelian” strategy did not identify all the increasing risk variants, it represented an efficient strategy for the formulation of novel hypotheses to explain complex diseases. Although medical genetics has largely benefited from the ideal dichotomous distinction in monogenic and complex diseases, most of the human genetic pathologies are located in a continuum between these two theoretical extremes.

1.7 Parkinson’s disease: a case study.

Parkinson’s disease (PD) is the second most common neurodegenerative disorder after Alzheimer’s disease and it is currently affecting ~6 million patients worldwide. Generally, symptoms appear after the sixth decade and comprise several motor dysfunctions, such as bradikynesia, resting tremor, rigidity but also affect cognitive capabilities (Poewe 2008). At the molecular level, PD is the result of the progressive degeneration of dopaminergic neurons in the substantia nigra of the brain, probably caused by the accumulation of protein aggregates, known as Lewy’s bodies and representing the hallmark of the disease (Forno 1996). Symptoms appear when more than half of the dopaminergic neurons have been lost and no treatment exists to slow down disease progression, while therapies only alleviate symptoms (Schapira 2005).

Rather than a single disorder, PD could be considered as an aggregate of symptoms that collectively affect the motor system and the cognitive functions. The phenotypic heterogeneity makes the diagnosis difficult, which comprises a large list of inclusive and exclusive criteria. Based on the guidelines provided by the UK National Institute for Health and Clinical Excellence (NICE), four different symptoms are considered cardinal in PD. The co-presence of bradykinesia with either resting tremor, or muscular rigidity or postural instability is necessary but not sufficient for the diagnosis. In addition, putative patients do not have to show any of the eleven exclusion criteria, comprising the occurrence of head injury or repeated strokes before the diagnosis, the use of antipsychotic or dopamine-depleting drugs, exposure to known neurotoxin and other neurological features, such as supranuclear gaze palsy and early severe dementia. Moreover, some supportive additional symptoms could further reinforce a proper diagnosis.

Even if through the whole 20th century many authors reported familial clustering of PD cases, only in the last 25 years the identification of causal variants permitted to describe the genetic components underlying the disease. For instance, one of the earliest linkage studies failed to detect the recently discovered α-synuclein multiplication occurring in a Swedish family (Planansky 1950).



Step 1. Diagnosis of parkinsonian syndrome
- Bradykinesia
- At least one of the following:
  - Muscular rigidity
  - 4–6 Hz rest tremor
  - Postural instability not caused by primary visual, vestibular, cerebellar or proprioceptive dysfunction

Step 2. Exclusion criteria for Parkinson's disease
- History of repeated strokes with stepwise progression of parkinsonian features
- History of repeated head injury
- History of definite encephalitis
- Oculogyric crises
- Neuroleptic treatment at onset of symptoms
- More than one affected relative
- Sustained remission
- Strictly unilateral features after 3 years
- Supranuclear gaze palsy
- Cerebellar signs
- Early severe autonomic involvement
- Early severe dementia with disturbances of memory, language and praxis
- Babinski sign
- Presence of cerebral tumour or communication hydrocephalus on imaging study
- Negative response to large doses of levodopa in absence of malabsorption
- MPTP (1-methyl 4-phenyl 1,2,3,6-tetrahydropyridine) exposure

Step 3. Supportive prospective positive criteria for Parkinson's disease*
- Unilateral onset
- Rest tremor present
- Progressive disorder
- Persistent asymmetry predominantly affecting side of onset
- Excellent response (70–100%) to levodopa
- Severe levodopa-induced chorea
- Levodopa response for 5 years or more
- Clinical course of 10 years or more

Table 1.4. Diagnostic criteria for PD suggested by UK Parkinson's disease Society Brain Bank.

Subsequently, several twin studies failed to confirm the genetic basis of PD, leading Duvoisin to conclude in 1987 that "the best available data do not support a role of heredity in the aetiology of PD" (Duvoisin 1987). Only few years later, analyzing the pedigree of the "Contursi kindred" originating from Southern Italy, Golbe and collaborators proved the contribution of genetic factors to PD (Golbe, Miller, and Duvoisin 1990; Golbe et al. 1993). Simultaneously, Wzolek et al. described dominant autosomal inheritance in three different families (Wszolek et al. 1993) and twin studies showed much higher concordance rates than previously thought (Burn et al. 1992), leading the very same Duvoisin to conclude that “these findings favor monogenic autosomal dominant inheritance and show reason to argue against a multifactorial aetiology” (Duvoisin and Johnson 1992). As I will largely describe later, evidently he was again wrong. Indeed, a growing number of genes and mutations related to the Mendelian forms of PD have been identified in rapid succession during the whole following decade.

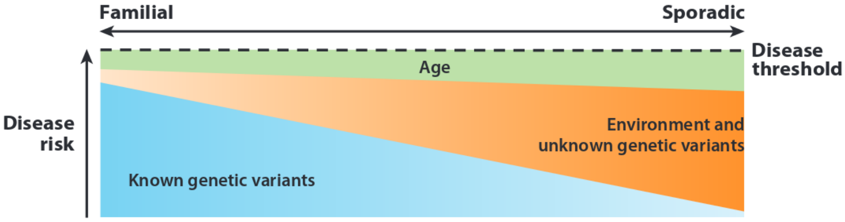

Familial
Sporadic
Disease threshold
Age
Disease risk
Environment and unknown genetic variants
Known genetic variants

Figure 1.21. Model illustrating interactions between genetic factors, environment and age for familial and sporadic forms of PD. From Shulman, De Jager, and Feany 2011.

1.7.1 Mendelian forms of Parkinson’s disease.

Parkinson’s disease (PD) can be divided in familial (Mendelian) and sporadic forms. A number of causal genes have been discovered for the Mendelian form, which constitutes 10-20% of the total cases. A missense mutation in the long arm of chromosome 4 is the first causal variant identified by linkage mapping to segregate with dominant autosomal familial form of PD (Polymeropoulos et al. 1997). The protein altering mutation Ala53Thr is located within the highly conserved *SNCA* gene whose longest transcript encodes for a protein composed of 140 amino-acids. Especially abundant in the presynaptic terminals of neurons, the *SNCA* protein product is the main component of the aggregates observed for several neurodegenerative pathologies (Giasson et al. 2000). Effectively, the *SNCA* protein is the major component of the Lewy's bodies, the cellular inclusions considered the pathological hallmark of both familial and sporadic forms of PD. The Ala53Thr mutation has been observed in several families throughout the world and haplotype data suggest that the variant originated from a single mutational event, occurring in a common ancestor (Markopoulou et al. 1999; Spira et al. 2001). Subsequent linkage analysis permitted the identification of two additional missense mutations in the *SNCA* gene, Ala30Pro (Krüger et al. 1998) and Glu46Lys (Zarranz et al. 2004), while the recent application of whole exome and target sequencing allowed the discovery of two more rare missense mutations in *SNCA* gene (Lesage et al. 2013; Proukakis et al. 2013). Mutations in the *LRRK2* gene are a much common cause of the dominant autosomal form of PD. Indeed, the missense mutation Gly2019Ser, firstly discovered in a Japanese family (Funayama et al. 2002), accounts for up to 7% of the affected individuals of European ancestry (Di Fonzo et al. 2005) and explains ~30% of PD familial cases from Ashkenazi Jews population (Ozelius et al.

2006). As for the *SNCA* gene, additional point mutations were later identified for the very large *LRRK2* gene (~144 Kb), including Arg1441Gly, Tyr1699Cys and Ile1122Val, among others (Paisán-Ruíz et al. 2004; Zimprich et al. 2004). Exome sequencing on affected families permitted the first identification and subsequent confirmation of the pathogenic role of the missense mutation Asp620Asn on the *VPS35* gene (Wider et al. 2008), which is involved in endosome-trans-Golgi trafficking and in the recycling of membrane associated proteins. Even if mutations in the *VPS35* gene only accounts for ~1% of the autosomal dominant PD cases (Ando et al. 2012), subsequent target sequencing of this gene in a large set of affected individuals and haplotype analysis indicated the presence of several independent mutational events, suggesting the presence of a mutational hotspot (Vilariño-Güell et al. 2011).

Causal variants for the autosomal recessive forms of PD have been identified using positional cloning in four causative genes, that is *PARK2*, *PINK1*, *DJ1* and *ATP13A2* (Matsumine et al. 1997; Valente et al. 2004; Bonifati et al. 2003; Ramirez et al. 2006). *PARK2* is involved in the proteasome-dependent degradation of proteins and in mitochondrial quality control by lysosome-dependent degradation. More than 170 different mutations have been identified throughout the gene and mutations in *PARK2* gene are the major cause of the familial recessive form of PD with an onset ≤40 years (Nuytemans et al. 2010). More than 50 homozygous and compound heterozygous mutations in *PINK1* gene have also been demonstrated to be related to the recessive form of PD (Nuytemans et al. 2010). It has been hypothesized that the mitochondrial serine/threonine kinase *PINK1* may phosphorylate mitochondrial proteins in response to cellular stress, protecting against mitochondrial dysfunction (Valente et al. 2004). Conversely, only few mutations have been reported for the pleiotropic gene *DJ1* (Valente et al. 2004), which encodes a chaperone with protease activity that acts as a transcriptional regulator, an antioxidant scavenger involved in tumorigenesis and in maintaining mitochondrial homeostasis (Ottolini et al. 2013). *ATP13A2* belongs to the P-type superfamily of ATPases that transport inorganic cations and other substrates across cell membranes (Schultheis et al. 2004). Loss of function mutations on *ATP13A2* gene have been related to a specific form of autosomal recessive PD, known as Kufor-Rakeb syndrome (Ramirez et al. 2006). Two different homozygous mutations (Arg741Gln, Arg747Trp) in two unrelated families have been identified through linkage analysis in the *PLA2G6* gene (Paisan-Ruiz et al. 2009), which encodes for an A2 phospholipase, a class of enzyme that catalyzes the release of fatty acids from phospholipids and which was previously related to other neurodegenerative pathologies (Morgan et al. 2006). Additional point mutations have been identified by linkage analysis on the *FBXO7* gene (Shojaee et al. 2008; Di Fonzo et al. 2009), econding for an F-box protein forming part of the E3 ubiquitin protein ligases (Winston et al. 1999) and directly interacting with *PARK2* protein product (Burchell et al. 2013). Several other loci, apparently following Mendelian inheritance, such as those located within

UCHL1, *EIF4G1*, *GIGYF2* and *HTRA2* genes, have been mapped in PD affected families, but their role in the disease is still somewhat controversial.

In addition to the described causal point mutations, many gene dosage alternations and structural rearrangements have been proved to play a major role on PD. By quantitative PCR amplification of *SNCA* exons, a whole gene triplication and duplication (Singleton 2003; Ibáñez et al. 2004; Chartier-Harlin et al. 2004) were reported to segregate with PD. More than 40 types of exon rearrangements have been observed in the *PARK2* gene, involving multi-exons and single exon deletions (Kitada et al. 1998; Deng et al. 2006; Lucking et al. 1998; Hedrich et al. 2001), but also duplications and triplications (Chaudhary et al. 2006; Sun et al. 2006; Simon-Sanchez et al. 2008; Lücking et al. 2001). Similarly, structural rearrangements in *DJ1* (Bonifati et al. 2003; Hedrich et al. 2004; Guo et al. 2010; Djarmati et al. 2004) and *PINK1* (Marongiu et al. 2007; Camargos et al. 2009; Li et al. 2005; Cazeneuve et al. 2009; Guo et al. 2010) genes have been detected to segregate with PD.

The described point mutations and structural variations together with other variants with relevant implications in the resulting proteins, such as frameshift indels, indicate that the monogenic forms of PD show high genetic heterogeneity, which result in a correspondent high phenotypic diversity. Genetic variants of any class are currently stored in large databases, such as OMIM (Hamosh et al. 2005), or in more specific catalogs precisely designated to collect variation related to Parkinson’s disease, as the

Parkinson’s Disease Mutation Database (PDmutDB) (Horaitis et al. 2007).

Collectively, the characterization of variants causal of Mendelian forms of PD represents an enormous source of clues for the understanding of the molecular mechanisms underlying the disease. Their knowledge permitted the identification of the metabolic pathways with a relevant role on PD pathogenesis. Among those, oxidative stress, mitochondrial respiration and protein degradation processes emerge as the most important. Based on these observations, three different hypotheses have been proposed to explain the death of dopaminergic neurons in PD affected individuals. The first suggests that oxidative stress, due to iron accumulation in the nervous system, leads to changes in calcium channel activity and altered proteolysis, resulting in α-synucleic aggregation and the subsequent formation of the detrimental cellular inclusions (Dexter and Jenner 2013). Since mutations in *SNCA*, *PARK2*, *PINK1* and *DJ1* have been associated to altered mitochondrial functions, an alternative hypothesis has been proposed to explain disease manifestation. Probably, the two proposed hypotheses are functionally overlapping, since abnormalities in mitochondrial structure and function could increase oxidative stress, which in turn increases mitochondrial dysfunctions (Dexter and Jenner 2013). Another hypothesis indicates the altered protein catabolism as the major mechanism leading to the formation of α-synuclein aggregates and neuronal death. The inhibition of the correct proteosomal functions could lead to increased oxidative and nitrative stress and thus alter mitochondrial functions (Dexter and Jenner 2013). Even if the precise sequence of events leading to neurons death is not fully understood, the study of the Mendelian forms of PD provided useful insights regarding the general pathogenesis and the altered molecular functions at the basis of the disease.

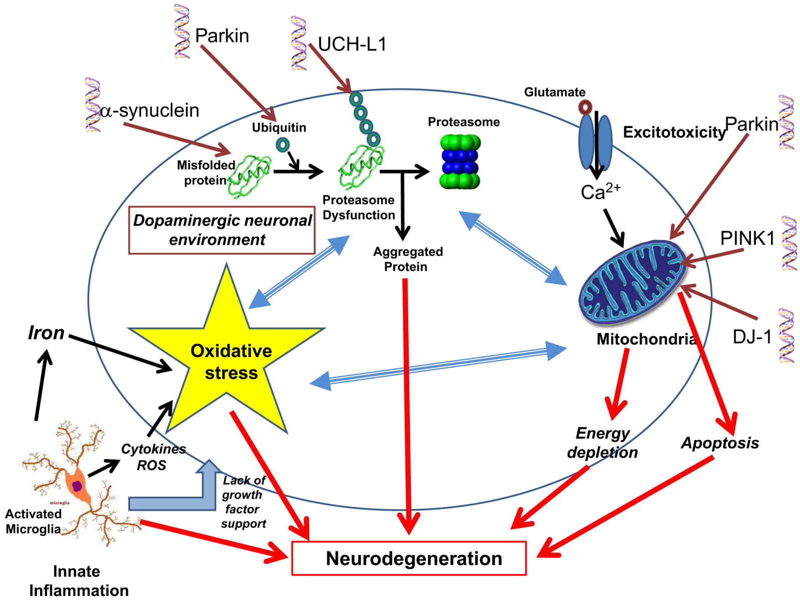

Parkin
UCH-L1
α-synuclein
Ubiquitin
Proteasome
Glutamate
Excitotoxicity
Parkin
Misfolded protein
Ca2+
Proteasome Dysfunction
Dopaminergic neuronal environment
PINK1
Aggregated Protein
Iron
Oxidative stress
Mitochondria
DJ-1
Energy depletion
Apoptosis
Cytokines ROS
Lack of growth factor support
Activated Microglia
Neurodegeneration
Innate Inflammation

Figure 1.22. Key molecular mechanisms widely accepted to contribute to the neurodegenerative process of PD. From Dexter and Jenner 2013.

1.7.2 Sporadic form of Parkinson's disease.

From a clinical point of view, the idiopathic form of PD is not distinguishable from some monogenic late onset forms. The high phenotypic heterogeneity characterizing PD is probably a consequence of the several factors at the basis of its pathogenesis. Sporadic PD is the result of the cumulative interaction between genetic background, lifestyle and various environmental factors. A simple way to describe this condition is that “genetics loads the gun and environment pulls the trigger”. Indeed, 40-60% of the phenotypic variance in PD can be explained by additive genetic effects (h^2, narrow-sense heritability), while the rest of the variance can be attributed to epistatic interactions, environmental factors and the interplay of genetics and environment (Hamza and Payami 2010). Among non-genetic risk factors, differences in geographical distribution have been observed for PD. For instance, PD prevalence is higher in the agricultural California central valley; living close to the fields, where persistent organic pollutants and pesticides have been extensively used (Costello et al. 2009; van der Mark et al. 2012). Several epidemiological, animal and *in vitro* studies evidenced that exposure to pesticides and herbicides, such as Rotenone and Paraquat among others, increases risk for PD (Goldman 2014) and a risk ratio of 1.6 was observed for ever being exposed to these chemical compounds (Noyce et al. 2012; Goldman 2014). Dysregulated iron homeostasis has also been implicated in PD aetiology for decades (Riederer et al. 1989), raising controversy regarding the higher prevalence of PD in individuals working in heavy industrial activities. However, even if agricultural-related and industrial jobs may involve a greater exposure to increasing risk chemical agents, the higher prevalence of PD in certain occupational categories needs to be fully proven. Other factors may actually reduce the risk to develop PD. Surprisingly, nicotine may act as a dopamine stimulant with a neuroprotective effect on tobacco smokers (Checkoway et al. 2002). The chemical compounds contained in the tobacco smoke may reduce the enzymatic activity of type B monoamine oxidase (MAO-B) (Fowler et al. 1996), that normally metabolizes dopamine and generates metabolites with a possible neurotoxic effect (Riederer et al. 1989). Furthermore, a significant association between coffee consumption and decreased risk for PD has been described, indicating that caffeine and its main metabolite paraxanthine have both a neuroprotective function (Tan et al. 2007).

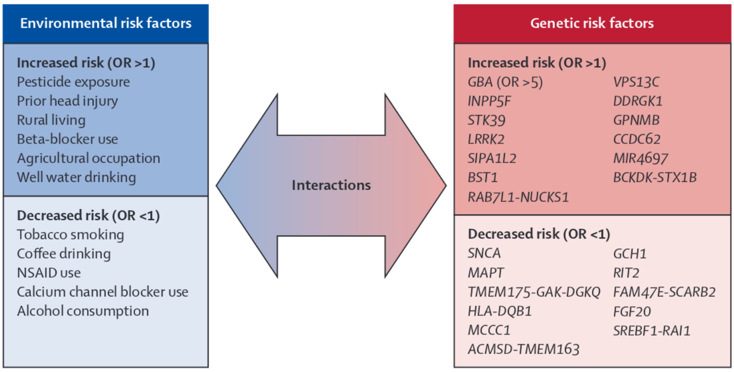

Environmental risk factors
Increased risk (OR >1)
Pesticide exposure
Prior head injury
Rural living
Beta-blocker use
Agricultural occupation
Well water drinking
Decreased risk (OR <1)
Tobacco smoking
Coffee drinking
NSAID use
Calcium channel blocker use
Alcohol consumption
Interactions
Genetic risk factors
Increased risk (OR >1)
GBA (OR >5)
VPS13C
INPP5F
DDRGK1
STK39
GPNMB
LRRK2
CCDC62
SIPA1L2
MIR4697
BST1
BCKDK-STX1B
RAB7L1-NUCKS1
Decreased risk (OR <1)
SNCA
GCH1
MAPT
RIT2
TMEM175-GAK-DGKQ
FAM47E-SCARB2
HLA-DQB1
FGF20
MCCC1
SREBF1-RAI1
ACMSD-TMEM163

Figure 1.23. Environmental and genetic risk factors for the development of PD. From Kalia and Lang 2016.

The first evidence that genetic factors contribute to sporadic PD was provided by a linkage study performed on 174 families with multiple individuals diagnosed with idiopathic PD. A genetic marker located near the tau microtubule-associated protein gene (*MAPT*) was observed to segregate with the disease (Scott et al. 2001). *MAPT* gene is expressed in the brain and it is a known increasing risk factor for Alzheimer's disease and other neurodegenerative disorders (Hutton et al. 1998; Baker et al. 1999). The long arm of chromosome 17, harboring the *MAPT* gene, contains a large inversion spanning 900 Kb which results in two main haplotypes: the more common non-inverted H1 haplotype and the inverted H2 haplotype. The significant association between the non-inverted H1 haplotype and PD (Pastor et al. 2000) is also reflected in the subsequent identification of several common variants in linkage disequilibrium with *MAPT* gene and increasing risk for PD. Under the assumption that common diseases are caused by common variants (CD-CV hypothesis), GWAS permitted the identification of many other genomic loci associated to PD. Initially, the reduced GWAS power, due to small sample sizes and low density genotyping arrays, resulted in the identification of increasing risk common variants mainly in linkage disequilibrium with genes previously related to the monogenic forms of PD. For example, four SNPs in the *SNCA* gene and three SNPs in the *MAPT* gene reached significance at genome wide level (Simon-Sanchez et al. 2009). Simultaneously, five common variants located in the *LRRK2* gene were described as increasing risk factors for PD (Satake et al. 2009). Interestingly, the variants located in the *MAPT* gene were not replicated in populations of Asian ancestry, whereas the *BST1* association hit (Satake et al. 2009) was not replicated in Europeans, indicating that PD could present inter-population genetic heterogeneity.

The increased power, due to the availability of high density panels of SNPs in genotyping arrays and the recruitment of larger sample sizes, allowed the identification of several loci increasing risk for PD (Hamza et al. 2010, Pankratz et al. 2008, Do et al. 2011). New potential metabolic routes and molecular pathways were thus proposed to have a role in disease pathogenesis. For instance, the association of the human leukocyte antigen (*HLA*) locus suggested a connection between the immune system and PD. The huge sampling recruitment operated by 23andMe (a personal genomics company) permitted a large association analysis, comprising 3,426 PD cases and ~30,000 controls. Besides confirming previously described associations, the 23andMe study suggested that the *GBA* gene could harbor a new possible increasing risk factor (Do et al. 2011). The *GBA* gene encodes for a lysosomal enzyme, known to harbor a variant that in homozygous state is causal of Gaucher's disease, a rare lipid storage disorder. Moreover, individuals affected by Gaucher's disease tend to be also affected by PD and the known missense causal variant N370S was already described to increase risk for idiopathic PD when found in heterozygous state (Goker-Alpan et al. 2004). Nevertheless, many of the identified increasing risk loci need further validation, as for example those in proximity of *SCARB2*, *SREBF1* and *RAI1* genes (Do et al. 2011), which have not been yet independently replicated.

After the first round of association studies, the efforts have been devoted to pool data of different GWAS to achieve greater power for the identification of risk factors with modest effect. In that sense, the purpose of the International Parkinson Disease Genomics Consortium (IPDGC) was to perform a meta-analysis, considering jointly the data provided by five different PD GWAS, accounting for a total of 5,333 PD cases and 12,019 controls in the discovery phase and 7,053 PD cases and 9,007 controls in the replication stage. In total, eleven loci reached genome wide significance, among which six (*MAPT, SNCA, HLA-DRB5, BST1*, *GAK* and *LRRK2*) were already described and five (in linkage disequilibrium with *ACMSD*, *STK39*, *MCCC1*, *LAMP3*, *SYT11*, *RAB25 CCDC62* and *HIP1R)* were identified thanks to the acquired meta-analysis power (IPDGC 2011a). Collectively, the eleven loci explain ~60% of the population attributable risk, which is the proportion of cases that would not occur in a population if these variants were not present (IPDGC 2011a). Subsequently, larger meta-analysis contributed to identify additional loci that in previous studies did not reach genome wide significance (IPDGC 2011b, Pankratz et al. 2012, Lill et al. 2012). The most recent and last effort to recover information through association studies was performed on 13,708 cases and 95,282 controls for the discovery phase and on 5,353 PD cases and 5,551 controls in the replication stage. The study, involving ~8 million SNPs, allowed the identification of six additional loci, in the proximity of *SIPA1L2*, *INPP5F*, *MIR4697*, *GCH1*, *VPS13C* and *DDRGK1* genes. These findings brought to a total of 28 independent risk variants in 24 different genomic loci (Nalls et al. 2014).

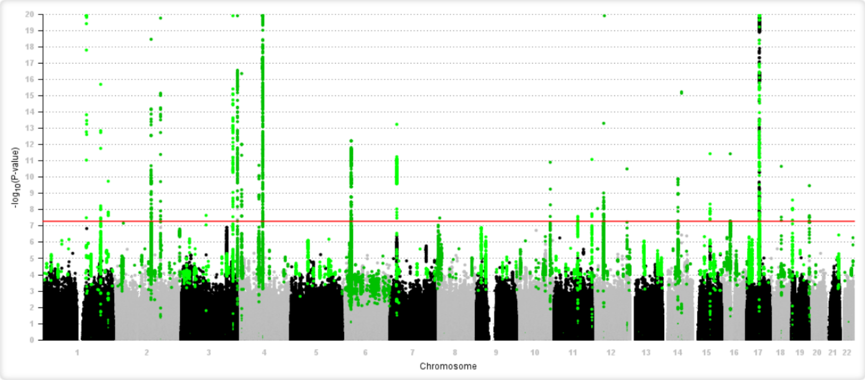


Figure 1.24. Manhattan plot of genome wide meta-analysis performed in PDGene database. From PDGene website.

Despite the large contribution provided by GWAS to disentangle PD genetic architecture, association studies also suffer possible spurious associations and do not have the power to identify all the increasing risk loci. For instance, a replication study involving 8,750 cases and 8,955 controls collected from the Genetic Epidemiology of PD (GEO-PD) consortium, suggested that the previously reported association involving *ACMSD* gene was a case of false positive, while the signal regarding *HLA-DRB5* was mainly due to population specific variation in allele frequencies (Sharma et al. 2012). Even if meta-analyses improved and confirmed the identification of many loci, none of the performed GWAS was able to detect in a single experiment all the reported increasing risk variants reaching genome wide significance. Larger samples sizes and the use of cohorts with more homogeneous phenotype are needed to further extend the knowledge of PD genetic factors through association studies. All the genetic signals detected in the last decade by GWAS are stored in large and general databases, such as the GWAS catalog (Hindorff et al. 2009; Welter et al. 2014), or in disease specific archives, specifically designed for the storage of the information regarding the variants linked to PD, such as the PDGene (Lill et al. 2012; Nalls et al. 2014) or the PDmutDB databases (Horaitis et al. 2007). Collectively, using a statistical method that applies a linear mixed model, it has been calculated that 27% of the total PD heritability may be explained by common variants included in commercial genotyping arrays, while the SNPs showing genome wide significance alone account for 3-5% of the phenotypic variance (Keller et al. 2012). These observations indicate that most of the risk loci still need to be identified, although present in the used arrays. This “missing heritability” could reside in common variants with smaller effects, but also in rare variants that are traditionally not captured in the genotyping arrays.

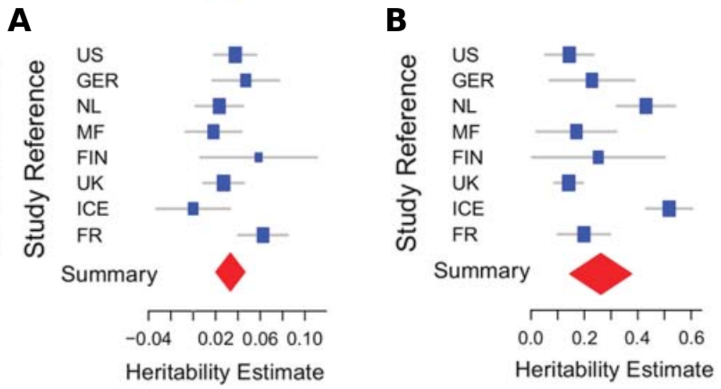


Figure 1.25. Explained heritability using significant SNPs (A) or all SNPs (B) of six GWAS data sets drawn from the International Parkinson's Disease Genomics Consortium (IPDGC). From Keller et al. 2012.

A further source of phenotypic variance could rely in structural variants and copy number variation, practically ignored by association studies but demonstrated to have a relevant role in the monogenic forms of PD. Moreover, it is of capital importance to describe the interactions among the already known increasing risk loci. Even if no significant interactions were detected between the alleles located in *SNCA* and *MAPT* genes, it has been hypothesized that patients carrying the corresponding susceptibility variants on both loci have higher risk than the simple combination of independent effects (Elbaz et al. 2011). Most GWAS signals fall in intergenic or gene desert regions, making difficult the identification of the causal variants driving disease risk. The sequencing of genomic regions around significant loci may provide an additional tool for capturing the genetic variation with a functional relevance in PD pathogenesis. In spite of such methodological limitations, GWAS permitted the detection of several reliable PD loci, shading light on the molecular mechanism underlying PD. Interestingly, association studies validated the role of genes previously involved in the familial forms of PD, indicating that similar pathways and metabolic routes are shared across disease forms. Finally, association studies allowed the identification of new potentially relevant functions that could contribute to a better understanding of the disease architecture and to the future design of improved treatments.

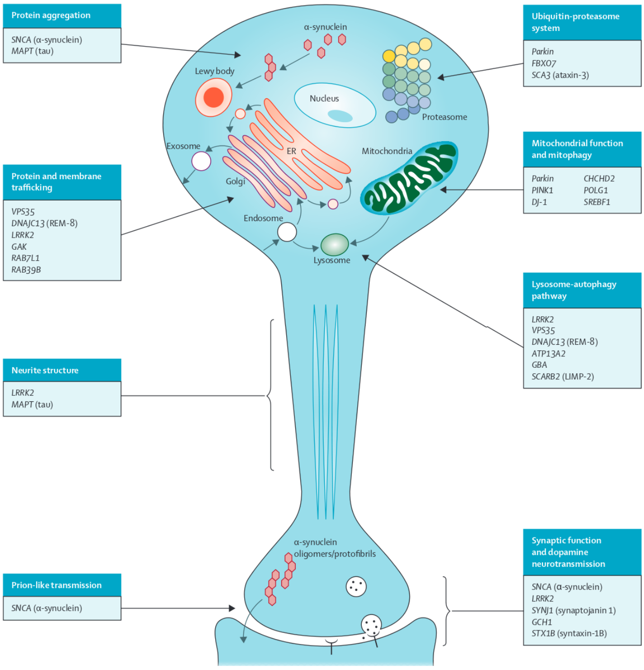

Protein aggregation
SNCA (α-synuclein)
MAPT (tau)
α-synuclein
Lewy body
Nucleus
Ubiquitin-proteasome system
Parkin
FBXO7
SCA3 (ataxin-3)
Proteasome
Exosome
ER
Mitochondria
Golgi
Protein and membrane trafficking
VPS35
DNAJC13 (REM-8)
LRRK2
GAK
RAB7L1
RAB39B
Mitochondrial function and mitophagy
Parkin
PINK1
DJ-1
CHCHD2
POLG1
SREBF1
Endosome
Lysosome
Lysosome-autophagy pathway
LRRK2
VPS35
DNAJC13 (REM-8)
ATP13A2
GBA
SCARB2 (LIMP-2)
Neurite structure
LRRK2
MAPT (tau)
α-synuclein oligomers/protofibrils
Prion-like transmission
SNCA (α-synuclein)
Synaptic function and dopamine neurotransmission
SNCA (α-synuclein)
LRRK2
SYNJ1 (synaptojanin 1)
GCH1
STX1B (syntaxin-1B)

Figure 1.26. Cellular processes involved in the pathogenesis of Parkinson’s disease. From Kalia and Lang 2016.

CHAPTER 2

Yes we can!
BARACK OBAMA

OBJECTIVES

Taking Parkinson’s disease (PD) as a model, the work presented in this thesis aims to contribute to the understanding of the genetic architecture of human polygenic diseases and to the description of the evolutionary and biological features of the genes linked to diseases. The three main objectives that will be addressed are:

1. To explore the role of rare variants in PD. Taking advantage of the improvement of sequencing technologies, in Chapter 3 we will describe variation in genomic loci previously described to have a role in sporadic PD. Since GWAS identified associated rather than causal variants, we chose to re-sequence at high coverage the protein coding and putative regulatory regions of some of the genes in linkage disequilibrium with the variants showing the strongest association with the disease. In addition, we also included genes that were known to harbor causal variants for both the dominant and the recessive forms of familial PD, in order to test the contribution of these genes to sporadic PD. The purpose is to describe and functionally characterize the full spectrum of genetic variation

within 38 PD candidate genes on 249 idiopathic PD cases and 145 unrelated controls of European ancestry. Leveraging annotation databases, we decided to investigate the role of both rare and common putative functionally relevant variants in PD. Given their small number of observations, low frequency variants cannot be detected by classic association tests and require the application of specific statistical methodologies to disentangle their role. Basically, these methods collapse in different ways all the variant information in a given gene or region and compare the global burden of rare variants between different sets of samples. Moreover, statistics from the field of molecular evolution, such as Tajima's D, can also be used to test for deviations in the allele frequency spectrum between cases and controls, providing an opportunity to test the Common Disease-Rare Variants (CD-RV) hypothesis to explain the PD missing heritability.

2. To analyze structural variants such as gene copy number in PD cases. For that, in Chapter 4 we will take advantage of new bioinformatics tools, based on the assumption that differences in depth of coverage among specific genomic regions and across multiple samples can be used to obtain a semi-quantitative estimation of copy number variation. Thus, when analyzing multiple samples in parallel in the same targeting experiment, Next Generation Sequencing offers the unique possibility to predict within the same experimental setting SNPs, small indels and larger structural variation, contributing to the understanding of the role of

different classes of genetic variants on the disease aetiology.

3. To study the biological and evolutionary properties of genes related to human disease. Given the observed functional interconnection between the familial and the sporadic forms of PD reported in Chapter 3, we decided to investigate the biological and evolutionary properties of different classes of genes related to the wide spectrum of human diseases. In Chapter 5, we will first try to assess whether genes related to all human pathologies represent a specific category of genes with particular evolutionary and functional characteristics within our genome. We then assess the selective pressures acting on different sequence types of human disease genes and explore whether the “mutation-selection balance” model is sufficient to account for their evolutionary history. Our final objective is to disentangle the properties of genes linked to different types of diseases, with a special interest in genes that have been found to be related to both monogenic and complex diseases. To that end, we aim to describe if and how the two types of diseases are interconnected and what is the role of Mendelian genes in the aetiology of complex diseases.

PART II

RESULTS

CHAPTER 3

NEXT GENERATION SEQUENCING ON CANDIDATE GENES RELATED TO PARKINSON'S DISEASE.

Nino Spataro, Francesc Calafell, Laura Cervera-Carles, Ferran Casals, Javier Pagonabarraga, Berta Pascual-Sedano, Antonia Campolongo, Jaime Kulisevsky, Alberto Lleó, Arcadi Navarro, Jordi Clarimon, Elena Bosch.

Published (Spataro et al. 2014).

Spataro N, Calafell F, Cervera-Carles L, Casals F, Pagonabarraga J, Pascual-Sedano B, et al. Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. Hum Mol Genet. 2015 Apr 1;24(7):2023–34. DOI: 10.1093/hmg/ddu616

CHAPTER 4

COPY NUMBER VARIATION ON PARKINSON'S DISEASE

Nino Spataro, Ana Roca-Umbert, Laura Cervera-Carles, Monica Valles, Roger Anglada, Javier Pagonabarraga, Berta Pascual-Sedano, Antonia Campolongo, Jaime Kulisevsky, Ferran Casals, Jordi Clarimon, Elena Bosch.

Under Review

Spataro N, Roca-Umbert A, Cervera-Carles L, Vallès M, Anglada R, Pagonabarraga J, et al. Detection of genomic rearrangements from targeted resequencing data in Parkinson's disease patients. Mov Disord. 2017 Jan;32(1):165–9. DOI: 10.1002/mds.26845

CHAPTER 5

PROPERTIES OF HUMAN DISEASE GENES

Nino Spataro, Juan Antonio Rodriguez, Arcadi Navarro, Elena Bosch.

Under Review

Spataro N, Rodríguez JA, Navarro A, Bosch E. Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. Hum Mol Genet. 2017 Jan 4;26(3):ddw405. DOI: 10.1093/hmg/ddw405

PART III

DISCUSSION

CHAPTER 6

> If you're not confused, you're not paying attention.
>
> TOM PETERS

DISCUSSION

As outlined in Chapter 2, the aim of this thesis is to contribute to the understanding of the genetic architecture of human complex diseases. In the first work, we investigated the role of SNPs and small indels detected in a set of 38 candidate genes for PD; whereas, in the second we predicted and subsequently validated the role of structural variations in sporadic PD cases. The observation that genes harboring causal variants for familial PD are collectively involved in the sporadic form of the disease, led us to disentangle the role of Mendelian genes in complex diseases and to investigate the evolutionary and functional properties of the whole set of genes related to human genetic disorders.

6.1 Rare variants in complex diseases.

Several re-sequencing studies have shown that, probably due to the recent explosive population growth, most of the human variation is rare (Casals and Bertranpetit 2012; Keinan and Clark 2012; Gao and Keinan 2014) and that rare variants are enriched in deleterious alleles (Zhu et al. 2011; Tennessen et al. 2012). These low frequency variants are not captured by the commercial SNP arrays used in association studies and at the same time do not have a sufficiently large effect to be detected in family studies. The rapid development of Next Generation Sequencing technologies and the application of new statistical methods, specifically designed for testing the association between complex traits and rare variants, have successfully demonstrated their contribution to colorectal cancer, plasma high-density lipoprotein levels, type 1 diabetes and blood pressure (Fearnhead et al. 2004; Cohen et al. 2004; Nejentsev et al. 2009; Ji et al. 2008). The exploration of variation in proximity of GWAS hits could contribute to unravel the role of rare variants into the disease. Interestingly, some of the GWAS hits for sporadic PD point to genes previously described to harbor rare causal variants for the familial forms of the disease. Through a targeting approach, we re-sequenced the protein coding and regulatory regions of 38 PD candidate genes, including genes related to the Mendelian and the sporadic forms of the disease. The analyzed dataset comprises 249 PD idiopathic cases and 145 unrelated controls of Spanish origin in which up to ~500 Kb of genomic material per sample was sequenced at high coverage. Our working hypothesis was that rare variants may play a prominent role also in the aetiology of complex diseases. To this end, we measured deviations in the allele frequency spectrum between cases and controls, using recently developed “collapsing methods” and classic evolutionary statistics, such as Tajima’s D. These tests can be applied even in moderate sample sizes to both individual genes separately or to groups of genes.

Our analysis confirmed the role of the huge inversion in chromosome 17 involving the *MAPT* gene as a high frequency risk factor in a population of European ancestry. After correcting for multiple testing, none of the common SNPs and small indels found by re-sequencing were significantly different in frequency between cases and controls, meaning that our study design was not able to detect the original association signals. Cases and controls displayed similar site-frequency spectrum on genes detected by GWAS to increase risk for PD, indicating that for these genes rare variants are equally distributed between the cohorts. By contrast, differences between cases and controls were consistently observed when focusing on rare putative functional variation occurring on genes previously linked to the Mendelian forms of PD. Indeed, code-altering variants in Mendelian genes with a Minor Allele Frequency (MAF) $\leq 1\%$, including non-synonymous SNPs, nonsense mutations and both frameshift and non-frameshift exonic indels, were more abundant in PD cases. This result was consistent with several collapsing methods as well with the Tajima’s D statistic.

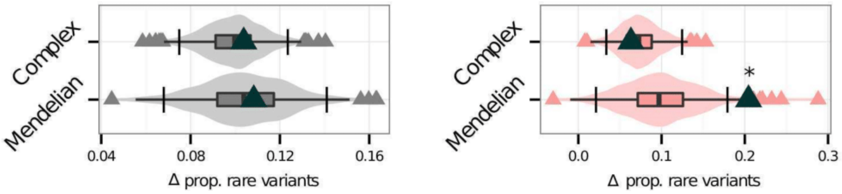


Figure 6.1. Difference in the proportion of rare variants between PD cases and controls. Violin plots represent the distribution of difference in proportion of rare variants when permuting 1,000 times cases and controls status. Triangles represent the observed case-control difference in the proportion of rare variants. Gray and light-red distributions refer to all rare variants and to code altering rare variants, respectively.

In our study, we also identified a considerable number of PD cases carrying variants known to be causal of the Mendelian forms of the disease. These rare causal variants do not explain alone the excess of low frequency variants observed for PD cases and when removing them from the analysis, rare variants still remained significantly more abundant in PD patients. Moreover, the excess of rare variants in Mendelian genes is present in both dominant and recessive genes separately. When testing each single Mendelian gene independently most of the genes showed a clear pattern toward an excess of rare variants in PD cases, even if not always significant.

Collectively, these results highlight the role that rare variants may have in the genetic architecture of this specific pathological condition. Even if our analysis is not able to identify which variant actually confers disease risk, our observations indicate that both the contrasting CD-CV and CD-RV hypotheses may be valid to explain the genetic factors underlying PD. Even more interestingly, our analysis highlights that rare variants may mediate the role of genes previously related to Mendelian PD in the sporadic form, suggesting the existence of a functional link between the two forms of the disease.

6.2 Copy number variation on Parkinson’s disease.

Most re-sequencing studies have been mainly focusing on variation of small size, including single point mutations and indels of a few base pairs. Different gene dosage alterations and structural rearrangements have been described to be related to familial PD and other neurological disorders. Mainly, these large genomic alterations were detected in a limited number of samples by using specific arrays designed to capture genome wide copy number variation or by using quantitative PCR on few candidate genes. Recently, different bioinformatics tools have been implemented to predict the copy number variation from depth of coverage data, relying on differences of coverage among genomic regions and across samples. Using our own re-sequencing data on candidate genes for PD, we confirmed by quantitative PCR all the 11 CNVs predicted by the XHMM software (Fromer et al. 2012) on 10 different PD cases.

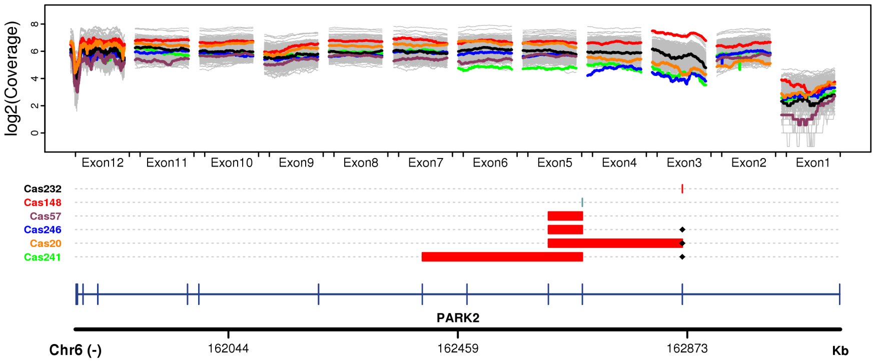


Figure 6.2. Schematic representation of coverage based predictions and validated CNVs on the *PARK2* gene. Colors in the upper panel match that of the individuals' labels in y-axis of the bottom panel. Gray background lines in the upper panel represent depth of coverage in the remaining samples of our dataset without predicted CNVs. At the bottom panel, red bars, light-blue bars and black diamonds represent deletions, duplications and frame-shift indels, respectively.

Among these CNVs, we observed 7 deletions and 4 duplications occurring in 3 different genes, previously described by other studies to be related to familial PD. Notably, we were able to detect structural variation in the *GBA* gene region, which shares 96% of identity with its neighboring pseudogene (*PGBA1*) and it is known to be a difficult region to re-sequence and investigate. The detection of these large variants, along with the description of known causal SNPs and frameshift indels allowed us to explain the genetic causal factors for 12 sporadic cases. Notably, exon dose alterations explained the disease phenotype of 6 of them, representing 2.4% of

PD cases in our study. Our analysis demonstrates the relevance of sequencing technologies for the discovery of all classes of variants and for the description of variation with a prominent role in disease aetiology, even if using moderate sample sizes and a limited portion of the human genome.

6.3 The role of Mendelian genes in complex diseases.

The functional link existing between Mendelian and sporadic PD outlined in Chapter 3 and 4, induced us to globally investigate the interplay between human rare monogenic and complex polygenic diseases. As a whole, genes related to human genetic pathological conditions represent a specific subset of the genome with particular evolutionary and biological features. Since diseases are diagnosed and observed in living humans, we expect those genes harboring causal or susceptibility variation for human diseases to be more functionally relevant than non disease gene, but not as much as the so-called essential genes (*i.e.* genes for which it is predictable that several functional mutations have lethal consequences in early stage of life). Protein-protein interaction network and expression profiles analyses confirmed our hypothesis and the intermediate functional relevance of human disease genes, that in turn, is also reflected in their intermediate protein coding evolutionary rates. Surprisingly, intra-species variation data indicate that the site frequency spectra of human disease genes are shifted toward intermediate frequency variants compared to the rest of the genome, suggesting that purifying selection is not the only evolutionary force acting on them, at least at their regulatory sequences. Moreover, human disease genes are enriched among those detected to show signatures of long lasting balancing selection, further indicating that adaptive evolutionary forces could contribute to shape the frequencies of causal alleles.

The biological features of human disease genes depends on the type of disease they are found to be associated with, that is monogenic Mendelian disorders or polygenic complex diseases. For instance, Mendelian genes emerged before and are more expressed than complex disease genes, while genes linked to both forms of pathologies show intermediate tendencies. Even if it is expected to observe genes related to Mendelian disorders under stronger purifying selection, we reported only slight differences compared to complex disease genes. Interestingly, genes related to disorders with Mendelian inheritance contribute higher than expected by chance to the risk for common polygenic diseases. More than 23% of genes harboring causal variants for monogenic disorders have been reported to increase risk for complex diseases. For 71% of the considered complex diseases at least 1 gene related to Mendelian disorders is observed among the increasing risk factors. The contribution of Mendelian genes in these complex traits varies widely, ranging from the 15 Mendelian genes reported for coronary heart disease, type 2 diabetes and prostate cancer, to the single Mendelian gene observed for 20 different complex diseases.

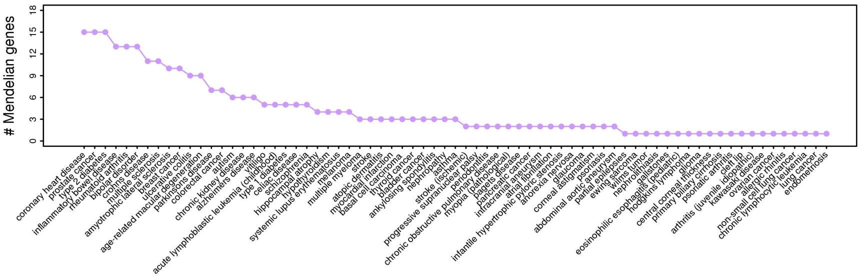


Figure 6.3. Number of genes related to Mendelian disorders located in proximity of increasing risk factors for complex diseases.

Even more interestingly, Mendelian genes tend to harbor variants with stronger effects on the phenotype, when compared to the rest of reported loci for a given complex disease. Together with the higher functional importance in the protein network, the tendency toward higher expression levels and the enrichment in relevant biological categories, these observations indicate that genes related to Mendelian disorders may play a major role in the aetiology of complex diseases. To conclude, we demonstrate that genes related to Mendelian disorders should be regarded as good potential candidates for complex diseases and our work highlights the importance of the extensive interplay existing among human genetic pathologies and between disease types.

6.4 Limitations.

In spite of the results achieved and the contribution given by this thesis to a better understanding of genetic architecture of PD, our study design suffers from different limitations. First of all, the budget restrictions did not allow a comprehensive genome wide investigation of the role of rare variants in idiopathic PD. Our analysis is mainly focused on genes previously observed to be related to familial PD and on genes located in proximity of the strongest association signals detected by GWAS. We further curated the list of candidate genes to study, selecting only those that were directly connected or in proximity of the rest of PD genes in the protein network. We only focused on exonic and putative regulatory sequences, further limiting the investigation of variation on the selected PD candidate genes. Even if the used sample size is sufficient to test in aggregate the association signal of rare variants in individual genes or groups of genes, it is not appropriate to detect significant association signals for a single rare variant. Thus, no clear candidate variants can be directly deduced to functionally validate *in vitro*. Theoretically, an association signal detected through GWA studies could result from multiple underlying low-frequency variants located within the same chromosome carrying the common risk allele (Dickson et al. 2010). Even if such "synthetic association" has not been systematically explored, different authors suggest that it does not explain most of GWAS results (Orozco, Barrett, and Zeggini 2010; Wray, Purcell, and

Visscher 2011). The difficulty of phasing singletons and other rare variants prevented us from testing whether rare variants enriched in PD cases are effectively located within chromosomes carrying particular common risk alleles.

Although we were able to detect structural variants from re-sequencing data, we ignore the false negative rate of the prediction method. Apparently, all the predicted CNVs were effectively present, but we do not know the fraction of real CNVs that the XHMM software is not able to identify.

The study presented in Chapter 5 is of course limited by the accuracy of the genetic information currently available for human diseases as well as by our incomplete knowledge regarding the true susceptibility/causal variants and their corresponding genes. Even if the underlying causal variants have been described for many Mendelian disorders, about half of all known Mendelian phenotypes still remain unsolved (Chong et al. 2015). For complex polygenic traits, the situation is even more obscure. Most of the associated variants identified by GWAS are not functionally characterized and most of them occur in intergenic regions, challenging the identification of the gene or functional element harboring the true causal variant. For each significant signal, a list of potential genes is reported in the GWAS catalog, which is based on the expertise of the authors on the biology of the disease. Inevitably, the lack of precise information regarding the genes harboring the true causal variants introduces false positives in the set of genes related to complex diseases. Conversely, the missing knowledge of all the genetic factors contributing to human complex diseases represents a source of false negatives. Putative human essential genes were inferred from essential genes detected by knock-out experiments in mice. Even if mice represent a useful model to study human biology, the list of human essential genes obtained probably also contains both false positives and negatives. Although all the used sets of genes (*i.e.* GWAS, hOMIM and essential genes) probably contain some erroneous assignations, we believe that the available information consent to have a reliable global description of the human genes properties. A further confounding factor could be due to the "publication bias" that could affect the protein-protein interaction network analysis and the classification of biological functions. Indeed, genes related to human diseases are generally more investigated than the rest of the genome, resulting in an overproduction of biological information, artificially inflating the number of known protein-protein interactions and gene functional annotation. Gene age analysis was based on data produced by multiple alignments to detect orthologous genes from very distant species in the phylogenetic tree. The limitations of alignment algorithms and the lack of the availability of genomic sequences for a wide range of species could lead to the wrong identification of the most distant species possessing an orthologous gene. Nevertheless, these limitations should equally affect all the considered sets of genes and should not bias our results toward a specific direction.

In spite of the considered limitations, we believe that in this thesis we use a good approximation for the identification of the genetic factors relevant for PD and for the description of the specific features of different sets of human genes, contributing to a better understanding of the genetic architecture of human diseases.

6.5 What's next?

The current cost-effective and standardized NGS technologies represent an incessant source of new biological information and continuously improve the knowledge of genetic factors contributing to diseases. Probably, in the near future, for each specific phenotype larger sample sizes will be sequenced, unveiling the role of both rare and common variants. Moreover, third generation sequencing technologies promise to generate single-molecule templates and long reads, enabling CNV discovery and direct haplotype phasing.

In spite of these predictable methodological and technical advances, additional theoretical frameworks are needed to include both epistasis and gene-environment interactions. Recently, it has been reported that 30% of the possible genotypes in the Green Fluorescent Protein (GFP) result in negative epistasis, while positive epistasis emerged only for 2% of the genotypes. These evidences indicate that frequently the joint effect of functional variants is stronger than their independent contribution (Sarkisyan et al. 2016). Similarly, gene-environment interactions affecting expression levels have been discovered in a recent twin study (Buil et al. 2015). Since most of the signals detected by GWAS are predicted to have a regulatory effect, these findings suggest that environmental factors might recover a fraction of the genetic and non-genetic variance associated to disease risk.

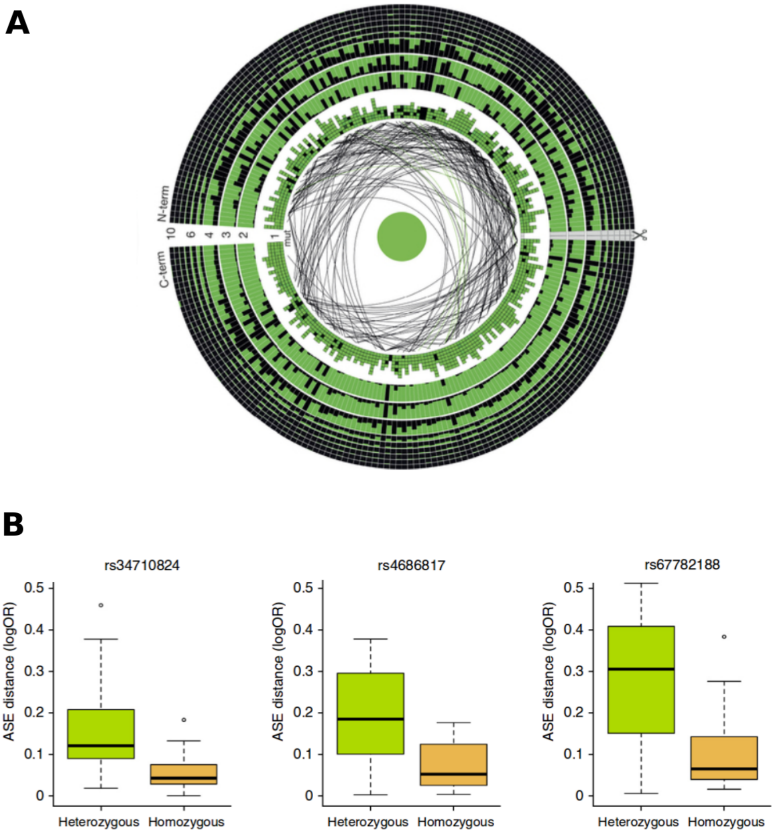


Figure 6.4. Examples of epistasis (A) and gene-environment interactions (B). In (A) the GFP protein sequence is arranged in circles; the inner one represents the wild-type protein, circles further away represent genotypes when from one to up to ten amino-acids are mutated simultaneously. At each circle, green and black bars represent the proportion of genotypes for a given position resulting in higher and lower fluorescent, respectively. Pairs of mutations with positive and negative epistasis are highlighted in green and black lines, respectively. From Sarkisyan et

al. 2016. Boxes in (B) display the difference in allele specific expression (ASE) between heterozygous (green) and homozygous (orange) monozygotic twin pairs at three SNPs of interest. Since monozygotic twins are genetically identical, any difference in ASE for two monozygotic siblings should be caused by environmental or epigenetic factors. From Buil et al. 2015.

Recently, the functional circuitry around a GWAS hit for obesity has been described in depth and the complete interpretation behind the elusive association has been provided (Claussnitzer et al. 2015). Translating a GWAS signal into a mechanistic process requires a combined approach that should include different related fields of modern biology, such as epigenomic annotation, gene expression profiles, genome editing and many more. Only integrating different sources of knowledge and using a comprehensive approach it will be possible to describe the molecular mechanism behind each GWAS hit. A detailed functional explanation is needed for all association signals if biology wants to reconcile with medicine and fulfill the promise of personalized medicine.

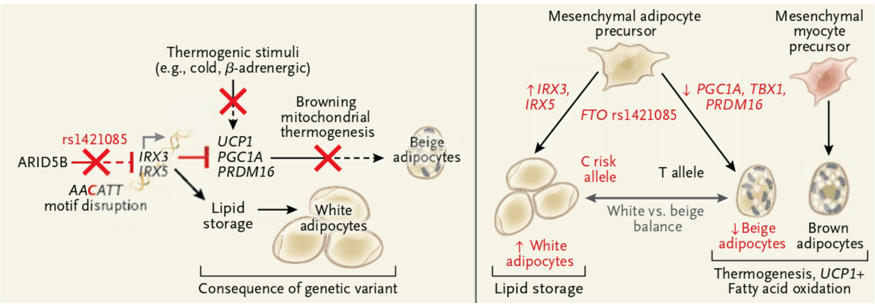

Thermogenic stimuli
(e.g., cold, β-adrenergic)
Browning
mitochondrial
thermogenesis
rs1421085
ARID5B
IRX3
IRX5
UCP1
PGC1A
PRDM16
Beige
adipocytes
AACATT
motif disruption
Lipid
storage
White
adipocytes
Consequence of genetic variant
Mesenchymal adipocyte
precursor
Mesenchymal
myocyte
precursor
↑IRX3,
IRX5
FTO rs1421085
↓ PGC1A, TBX1,
PRDM16
C risk
allele
T allele
White vs. beige
balance
↑ White
adipocytes
↓Beige
adipocytes
Brown
adipocytes
Lipid storage
Thermogenesis, UCP1+
Fatty acid oxidation

Figure 6.5. Description of functional circuitry for variant rs1421085 located in *FTO* locus. From Claussnitzer et al. 2015.

In the near future I would like to contribute to the functional characterization of GWAS signals related to PD and try to unveil how different genetic variants interact to generate particular phenotypes of interest. In collaboration with the Consiglio's group at the Center of Regenerative Medicine in Barcelona (CMRB), we will try to assess the genetic background at the basis of the incomplete penetrance of the Gly2019Ser mutation in the *LRRK2* gene, which accounts for up to 7% of the affected individuals of European ancestry (Di Fonzo et al. 2005). To this end, we are analyzing the whole exome of individuals carrying the mutation, but which are asymptomatic for PD. We will try to compare their genetic profiles with controls that do not carry the mutation and with symptomatic carriers. Putative candidate variants will be validated *in vitro* using neuronal cells obtained from reprogrammed induced pluripotent stem cells (iPSC). The comparison of specific endophenotypes among these different samples will shed light on the protective mechanisms behind the reduced penetrance of Gly2019Ser mutation, which in turn will contribute to a further understanding of pathogenic processes behind PD (Sánchez-Danés et al. 2012).

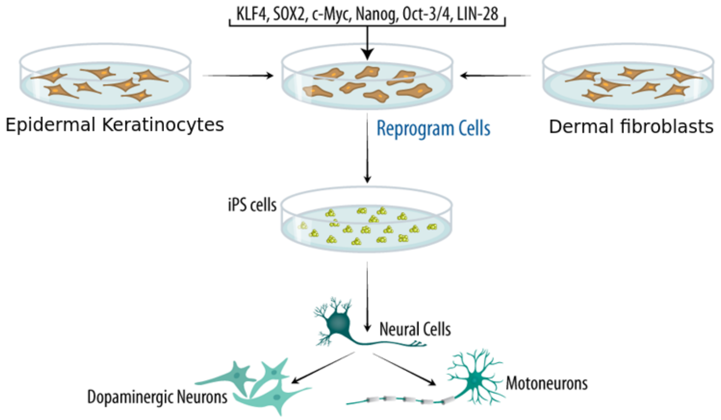

KLF4, SOX2, c-Myc, Nanog, Oct-3/4, LIN-28
Epidermal Keratinocytes
Reprogram Cells
Dermal fibroblasts
iPS cells
Neural Cells
Dopaminergic Neurons
Motoneurons

Figure 6.6. Schematic representation of the protocol to obtain reprogrammed neural cells from adult cell types. Adapted from Amabile and Meissner 2009.

In addition, I would like to continue exploring the relationships between Mendelian disorders and complex diseases. Currently, association studies represent an outdated technology and the whole field is moving towards sequencing experiments. Due to the large number of tests performed in GWAS, many true associated loci may not pass the multiple testing correction, limiting the current knowledge of the factors contributing to disease risk. Authors claiming a new association need to be sure that the identified variants are effectively associated. For this reason very stringent multiple testing thresholds have been usually applied to ensure the identification of true positives but at the cost of increasing the number of false negatives. Weaker association signals may be detected by increasing the sample size, which in turn implies an increase in experimental costs and poses new challenges regarding samples recruitment, stratification and homogeneity of phenotypes. Alternatively, I propose that weak signals could be still detected from existing GWAS data if we reanalyze them according to candidate gene like approaches. In Chapter 5, I demonstrated the role of genes related to Mendelian disorders in complex polygenic diseases. These genes are extensively located in proximity of susceptibility factors for complex diseases; moreover clinical records suggest that many Mendelian disorders co-occur with specific complex diseases beyond random expectation. Co-occurring traits are probably sharing similar disrupted metabolic routes and aetiology mechanisms. The broad knowledge concerning the exact pathological processes behind Mendelian disorders could shed light on the co-occurring complex diseases. Given that many of the causal genetic factors for Mendelian disorders are known, we could use this information to test whether these factors have also a role in the co-occurring polygenic diseases.

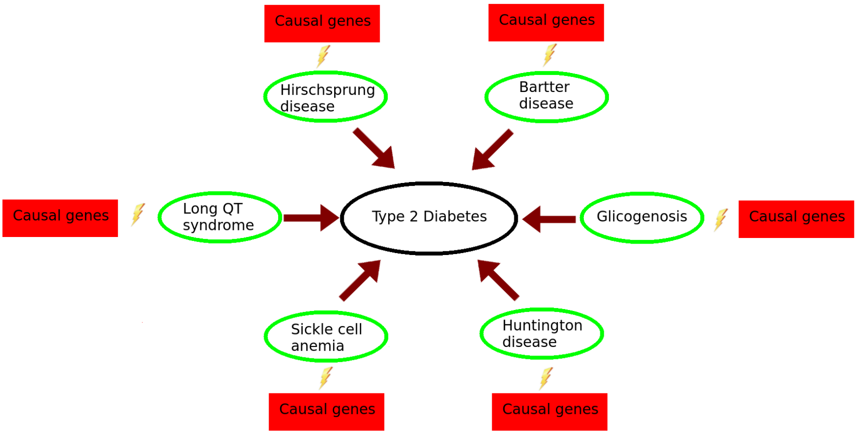

Causal genes
Causal genes
Hirschsprung disease
Bartter disease
Causal genes
Long QT syndrome
Type 2 Diabetes
Glicogenosis
Causal genes
Sickle cell anemia
Huntington disease
Causal genes
Causal genes

Figure 6.7. Example of candidate Mendelian genes for a complex trait. Red blocks indicate the genes harboring causal variants for the Mendelian traits in the green circles, which in turn co-occur with type 2 diabetes.

Similarly to the 3 “Rs” at the basis of the green economy, I am convinced that it is still possible to “reuse” previous GWAS, “reducing” the searching space to candidate gene sets and “recycling” clinic records regarding comorbidity to generate new genetic information about complex diseases.

REUSE
REDUCE
RECYCLE

Figure 6.8. The “3Rs” of the green economy.

CHAPTER 7

Yes we could...

An extremely synthetic personal evaluation of Obama’s presidency.

CONCLUSIONS

In this thesis, I showed that Next-generation DNA sequencing contributes to the discovery of variation linked to diseases and helps to understand their genetic architecture. Currently, many different classes of genetic variants can be rapidly investigated in a single cost effective experiment and in a large number of samples. Moreover, the many available biological databases furnish a unique possibility to depict the properties of genes harboring variants with a relevant role in genetic diseases, allowing to generate new knowledge employing preexisting information. “We are on the leading edge of a true revolution in medicine”, enthusiastically stated Francis Collins in his book “*The Language of Life*” (Collins 2009). Undoubtedly, the whole biology and medical genetics strongly benefited from the progress represented by the “*-omics*” sciences. Regardless of the results achieved from the sequencing of the first human genome, many of the promises about the understanding of genetic diseases still remain to be fulfilled. The exalted revolution has still not come completely true and probably many years are still needed before it takes its final form. In spite of the several ethical, legal and social concerns raised by the massive accumulation of genetic data, in the near future medicine will shift more and more toward sequencing and information technology based analysis, leading to many little discoveries that collectively will show the way to the “holy grail” of personalized medicine.

PART IV

APPENDIX

LIST OF PUBLICATIONS

List of articles published in Peer-Reviewed journals:

1. Engelken J, Espadas G, Mancuso FM, Bonet N, Scherr AL, Jímenez-Álvarez V, Codina-Solà M, Medina-Stacey D, **Spataro N**, Stoneking M, Calafell F, Sabidó E, Bosch E.
   Signatures of Evolutionary Adaptation in Quantitative Trait Loci Influencing Trace Element Homeostasis in Liver. Mol Biol Evol. 2016 Mar.
   https://www.ncbi.nlm.nih.gov/pubmed/26582562
2. **Spataro N**, Calafell F, Cervera-Carles L, Casals F, Pagonabarraga J, Pascual-Sedano B, Campolongo A, Kulisevsky J, Lleó A, Navarro A, Clarimón J, Bosch E.
   Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. Hum Mol Genet. 2015 Apr.
   https://www.ncbi.nlm.nih.gov/pubmed/25504046
3. Farfán M, **Spataro N**, Sanglas A, Albarral V, Lorén JG, Bosch E, Fusté MC.
   Draft Genome Sequence of the Aeromonas diversa Type Strain. Genome Announc. 2013 Jun.
   https://www.ncbi.nlm.nih.gov/pubmed/23792745
4. **Spataro N**, Farfán M, Albarral V, Sanglas A, Lorén JG, Fusté MC, Bosch E.
   Draft Genome Sequence of Aeromonas molluscorum Strain

848TT, Isolated from Bivalve Molluscs. Genome Announc. 2013 Jun.
https://www.ncbi.nlm.nih.gov/pubmed/23788549

List of submitted papers:

1. **Spataro N**, Rodriguez J, Navarro A, Bosch E.
   Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology.
2. Rodriguez J, Marigorta U, Hughes D, **Spataro N**, Bosch E, Navarro A.
   Evidence that the antagonistic pleiotropy and mutation accumulation theories influence human senescence and disease.
3. **Spataro N**, Roca-Umbert A, Cervera-Carles L, Vallès M, Anglada R, Pagonabarraga J, Pascual-Sedano B, Campolongo A, Kulisevsky J, Casals F, Clarimón J, Bosch E.
   Detection of genomic rearrangements from targeted re-sequencing data in Parkinson's disease patients.

SUPPLEMENTARY MATERIALS.

Next generation sequencing on candidate genes related to Parkinson's disease.

The Supplementary Material for the article presented in Chapter 3 (Spataro et al. 2014) is available at: http://hmg.oxfordjournals.org/content/24/7/2023/suppl/DC1 and below. Supplementary Tables can be found in Electronic Appendix 1.

Spataro N, Calafell F, Cervera-Carles L, Casals F, Pagonabarraga J, Pascual-Sedano B, et al. Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. Supplementary material. Hum Mol Genet. 2015 Apr 1;24(7):2023–34. DOI: 10.1093/hmg/ddu616

SUPPLEMENTARY MATERIALS.

Copy Number Variation on Parkinson’s disease.

The Supplementary Material for the article presented in Chapter 4 is available below.

Spataro N, Roca-Umbert A, Cervera-Carles L, Vallès M, Anglada R, Pagonabarraga J, et al. Detection of genomic rearrangements from targeted resequencing data in Parkinson’s disease patients. Supplementary material 2. Mov Disord. 2017 Jan;32(1):165–9. DOI: 10.1002/mds.26845

SUPPLEMENTARY MATERIALS.

Properties of human disease genes.

The Supplementary Material for the article presented in Chapter 5 is available below. Supplementary Tables can be found in Electronic Appendix 2.

Spataro N, Rodríguez JA, Navarro A, Bosch E. Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. Supplementary material. Hum Mol Genet. 2017 Jan 4;26(3):ddw405. DOI: 10.1093/hmg/ddw405

BIBLIOGRAPHY

Abi-Rached, L, Jobin, M J, Kulkarni, S, McWhinnie, A, Dalva, K, Gragert, L, Babrzadeh, F et al. 2011. “The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans.” *Science* 334 (6052): 89–94.

Adams, J. 1816. “A Treatise on the Supposed Hereditary Properties of Diseases.” *London: Callow*, April.

Allison, A C. 1954. “Protection Afforded by Sickle-Cell Trait Against Subtertian Malarial Infection.” *British Medical Journal* 1 (4857): 290–94.

Amabile, G, and Meissner, A. 2009. “Induced Pluripotent Stem Cells: Current Progress and Potential for Regenerative Medicine.” *Trends in Molecular Medicine* 15 (2): 59–68.

Ando, M, Funayama, M, Li, Y, Kashihara, K, Murakami, Y, Ishizu, N, Toyoda, C et al. 2012. “VPS35 Mutation in Japanese Patients with Typical Parkinson’s Disease.” *Movement Disorders* 27 (11): 1413–17.

Asimit, J, and Zeggini, E. 2010. “Rare Variant Association Analysis Methods for Complex Traits.” *Annual Review of Genetics* 44 (1). Annual Reviews: 293–308.

Baker, M, Litvan, I, Houlden, H, Adamson, J, Dickson, D, Perez-Tur, J, Hardy, J, Lynch, T, Bigio, E, and Hutton, M. 1999. “Association of an Extended Haplotype in the Tau Gene with Progressive Supranuclear Palsy.” *Human Molecular Genetics* 8 (4): 711–15.

Bansal, V, Libiger, O, Torkamani, A, and Schork, N J. 2010. “Statistical Analysis Strategies for Association Studies Involving Rare Variants.” *Nat Rev Genet* 11 (11): 773–85.

Barabasi, A, Gulbahce, N, and Loscalzo, J. 2011. “Network Medicine: A Network-Based Approach to Human Disease.”

Nat Rev Genet 12 (1): 56–68.

Barrenas, F, Chavali, S, Holme, P, Mobini, R, and Benson, M. 2009. “Network Properties of Complex Human Disease Genes Identified through Genome-Wide Association Studies.” *PLoS ONE* 4 (11): 2–7.

Beecham, G W, Hamilton, K, Naj, A C, Martin, E R, Huentelman, M, Myers, A J, Corneveaux, J J, et al. 2014. “Genome-Wide Association Meta-Analysis of Neuropathologic Features of Alzheimer’s Disease and Related Dementias.”. *PLoS Genetics* 10 (9): e1004606.

Bertram, L, and Tanzi, R E. 2009. “Genome-Wide Association Studies in Alzheimer’s Disease.” *Human Molecular Genetics* 18 (R2): R137–45.

Blair, D R, Lyttle, C S, Mortensen, J M, Bearden, C F, Jensen, A B, Khiabanian, H, Melamed, R, et al. 2013. “A Nondegenerate Code of Deleterious Variants in Mendelian Loci Contributes to Complex Disease Risk.” *Cell* 155 (1): 70–80.

Blekhman, R, Man, O, Herrmann, L, Boyko, A R, Indap, A, Kosiol, C, Bustamante, C D, Teshima, K M, and Przeworski, M. 2008. “Natural Selection on Genes That Underlie Human Disease Susceptibility.” *Current Biology* 18 (12): 883–89.

Bonifati, V, Rizzu, P, van Baren, M J, Schaap, O, Breedveld, G J, Krieger, E, Dekker, M C J, et al. 2003. “Mutations in the DJ-1 Gene Associated with Autosomal Recessive Early-Onset Parkinsonism.” *Science* 299 (5604): 256–59.

Botstein, D, White, R L, Skolnick, M and Davis, R W. 1980. “Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms.” *American Journal of Human Genetics* 32 (3): 314–31.

Botstein, D, and Neil, R. 2003. “Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease.” *Nature Genetics* 33 Suppl (march): 228–37. .

Boveri, T. 1904. *Ergebnisse uber die Konstitution der* *chromatischen Substanz des Zellkerns*. Jena: G. Fischer.

Buil, A, Brown, A A, Lappalainen, T, Vinuela, A, Davies, M N, Zheng, H, Richards, J B, et al. 2015. “Gene-Gene and Gene-Environment Interactions Detected by Transcriptome Sequence Analysis in Twins.” *Nat Genet* 47 (1): 88–91.

Burchell, V S, Nelson, D E, Sanchez-Martinez, A, Delgado-Camprubi, M, Ivatt, R M, Pogson, J H, Randle, S J, et al. 2013. “The Parkinson’s Disease-Linked Proteins Fbxo7 and Parkin Interact to Mediate Mitophagy.” *Nat Neurosci* 16 (9): 1257–65.

Burn, D J, Mark, MH, Playford, E D, Maraganore, D M, Zimmerman, T R, Duvoisin, R C, Harding, A E, Marsden, C D, and Brooks, D J. 1992. “Parkinson’s Disease in Twins Studied with 18–dopa and Positron Emission Tomography.” *Neurology* 42 (10): 1894–1900.

Bush, W S, and Haines, J. 2001. “Overview of Linkage Analysis in Complex Traits.” In *Current Protocols in Human Genetics*.

Bustamante, C D, Fledel-Alon, A, Williamson, S, Nielsen, R, Hubisz, M T, Glanowski, S, Tanenbaum, D M, et al. 2005. “Natural Selection on Protein-Coding Genes in the Human Genome.” *Nature* 437 (7062): 1153–57.

Cai, J J, Borenstein, E, and Petrov, D. 2010. “Broker Genes in Human Disease.” *Genome Biology and Evolution* 2 (January): 815–25.

Callaway, E. 2011. “Ancient DNA Reveals Secrets of Human History.” *Nature* 476: 136–37.

Camargos, S T, Oliveira Dornas, L, Momeni, P, Lees, A, Hardy, J, Singleton, A and Cardoso, F. 2009. “Familial Parkinsonism and Early Onset Parkinson’s Disease in a Brazilian Movement Disorders Clinic: Phenotypic Characterization and Frequency of SNCA, PRKN, PINK1, and LRRK2 Mutations.” *Movement Disorders* 24 (5): 662–66.

Campion, D, Dumanchin, C, Hannequin, D, Dubois, B, Belliard, S, Puel, M, Thomas-Anterion, C et al. 1999. “Early-Onset Autosomal Dominant Alzheimer Disease: Prevalence, Genetic

Heterogeneity, and Mutation Spectrum.” *American Journal of Human Genetics* 65 (3): 664–70.

Carter, A J R, and Nguyen, A Q. 2011. “Antagonistic Pleiotropy as a Widespread Mechanism for the Maintenance of Polymorphic Disease Alleles.” *BMC Medical Genetics* 12 (December): 160.

Casals, F, and Bertranpetit, J. 2012. “Human Genetic Variation, Shared and Private.” *Science* 337 (6090): 39–40.

Cazeneuve, C, Sân, C, Ibrahim, S A, Mukhtar, M M, Kheir, M M, LeGuern, E, Brice, A and Salih, M A. 2009. “A New Complex Homozygous Large Rearrangement of the PINK1 Gene in a Sudanese Family with Early Onset Parkinson’s Disease.” *Neurogenetics* 10 (3): 265–70.

Chartier-Harlin, M C, Kachergus, J, Roumier, C, Mouroux, V, Douay, X, Lincoln, S, Levecque, C et al. 2004. “α-Synuclein Locus Duplication as a Cause of Familial Parkinson’s Disease.” *Lancet* 364 (9440): 1167–69.

Chaudhary, S, Behari, M, Dihana, M, Swaminath, P V, Govindappa, S T, Jayaram, S, Goyal, V et al. 2006. “Parkin Mutations in Familial and Sporadic Parkinson’s Disease among Indians.” *Parkinsonism & Related Disorders* 12 (4): 239–45.

Checkoway, H, Powers, K, Smith-Weller, T, Franklin, G M, Longstreth, W T and Swanson, P D. 2002. “Parkinson’s Disease Risks Associated with Cigarette Smoking, Alcohol Consumption, and Caffeine Intake.” *American Journal of Epidemiology* 155 (8): 732–38.

Chong, J X, Buckingham, K J, Jhangiani, S N, Boehm, C, Sobreira, N, Smith, J D, Harrell, T M et al. 2015. “The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities.” *American Journal of Human Genetics* 97 (2): 199–215.

Cirulli, E T, and Goldstein, D B. 2010. “Uncovering the Roles of Rare Variants in Common Disease through Whole-Genome Sequencing.” *Nat Rev Genet* 11 (6): 415–25.

Clarke, L, Zheng-Bradley, X, Smith, R, Kulesha, E, Xiao, C, Toneva, I, Vaughan, B et al. 2012. “The 1000 Genomes Project: Data Management and Community Access.” *Nat Meth* 9 (5): 459–62.

Claussnitzer, M, Dankel, S N, Kim, K H, Quon, G, Meuleman, W, Haugen, C, Glunk, V et al. 2015. “FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.” *New England Journal of Medicine* 373 (10): 895–907.

Cohen, J C, Kiss, R S, Pertsemlidis, A, Marcel, Y L, McPherson, R and Hobbs, H H. 2004. “Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol.” *Science* 305 (5685): 869–72.

Collins, F S. 2009. *The Language of Life DNA and the Revolution in Personalized Medicine*. Pymble, NSW ; New York, NY: HarperCollins

Costello, S, Cockburn, M, Bronstein, J, Zhang, X and Ritz, B. 2009. “Parkinson’s Disease and Residential Exposure to Maneb and Paraquat From Agricultural Applications in the Central Valley of California.” *American Journal of Epidemiology* 169 (8): 919–26.

Creighton, H B and McClintock, B. 1931. “A Correlation of Cytological and Genetical Crossing-Over in Zea Mays.” *Proceedings of the National Academy of Sciences of the United States of America* 17 (8): 492–97.

Czene, K, Lichtenstein, P and Hemminki, K. 2002. “Environmental and Heritable Causes of Cancer among 9.6 Million Individuals in the Swedish Family-Cancer Database.” *International Journal of Cancer* 99 (2): 260–66.

De Hert, M, Steemans, D, Theys, P, Fryns, J P and Peuskens, J. 1996. “Lujan-Fryns Syndrome in the Differential Diagnosis of Schizophrenia.” *American Journal of Medical Genetics - Seminars in Medical Genetics* 67 (2): 212–14.

De Sanctis, V, Zurlo, M G, Senesi, E, Boffa, C, Cavallo, L and Di Gregorio, F. 1988. “Insulin Dependent Diabetes in Thalassaemia.” *Archives of Disease in Childhood* 63 (1): 58–62.

de Visser, J A, Cooper, T F and Elena, S F. 2011. “The Causes of Epistasis.” *Proceedings of the Royal Society of London B: Biological Sciences* 278 (1725): 3617–24.

Deng, H, Le, W, Hunter, CB and et al. 2006. “Heterogeneous Phenotype in a Family with Compound Heterozygous Parkin Gene Mutations.” *Archives of Neurology* 63 (2): 273–77.

Deschamps, M, Laval, G, Fagny, M, Itan, Y, Abel, L, Casanova, J L, Patin, E and Quintana-Murci, L. 2016. “Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes.” *The American Journal of Human Genetics* 98 (1): 5–21.

Dexter, D T and Jenner, P. 2013. “Parkinson Disease: From Pathology to Molecular Disease Mechanisms.” *Free Radical Biology and Medicine* 62 (September): 132–44.

Di Fonzo, A, Rohé, C F, Ferreira, J, Chien, H F, Vacca, L, Stocchi, F, Guedes, L et al. 2005. “A Frequent LRRK2 Gene Mutation Associated with Autosomal Dominant Parkinson’s Disease.” *Lancet* 365 (9457): 412–15.

Di Fonzo, A, Dekker, M C J, Montagna, P, Baruzzi, A, Yonova, E H, Correia Guedes, L, Szczerbinska, A et al. 2009. “FBXO7 Mutations Cause Autosomal Recessive, Early-Onset Parkinsonian-Pyramidal Syndrome”. *Neurology* 72 (3): 240-45

Dickson, S P, Wang, K, Krantz, I, Hakonarson, H and Goldstein, D B. 2010. “Rare Variants Create Synthetic Genome-Wide Associations.” *PLoS Biol* 8 (1): e1000294.

Djarmati, A, Hedrich, K, Svetel, M, Schäfer, N, Juric, V, Vukosavic, S, Hering, R et al. 2004. “Detection of Parkin (PARK2) and DJ1 (PARK7) Mutations in Early-Onset Parkinson Disease: Parkin Mutation Frequency Depends on Ethnic Origin of Patients.” *Human Mutation* 23 (5): 525.

Do, C B, Tung, J Y, Dorfman, E, Kiefer, A K, Drabant, E M, Francke, U, Mountain, J L et al. 2011. “Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson’s Disease.” *PLoS Genet* 7 (6): e1002141.

Dobzhansky, T. 1973. "Nothing in Biology Makes Sense except in the Light of Evolution." *The American Biology Teacher* 35 (3): 125–29.

Drake, J W. 1970. *The Molecular Basis of Mutation*. *San Francisco: Holden-Day*.

Durbin, R M, Altshuler, D L, Abecasis, G R, Bentley, D R, Chakravarti, A, Clark, A G, Collins, F S et al. 2010. "A Map of Human Genome Variation from Population-Scale Sequencing." *Nature* 467 (7319): 1061–73.

Duvoisin, R C. 1987. "Genetics of Parkinson's Disease." *Advances in Neurology* 45: 307–12.

Duvoisin, R C and Johnson, W G. 1992. "Hereditary Lewy-Body Parkinsonism and Evidence for a Genetic Etiology of Parkinson's Disease." *Brain Pathology* 2 (4): 309–20.

Elbaz, A, Ross, O A, Ioannidis, J P A, Soto-Ortolaza, A I, Moisan, F, Aasly, J, Annesi, G et al. 2011. "Independent and Joint Effects of the MAPT and SNCA Genes in Parkinson Disease." *Annals of Neurology* 69 (5):

Fabian, D and Flatt, T. 2011. "The Evolution of Aging." *Nature Education Knowledge* 3 (10): 9–10.

Fadiman, Anne. 1998. *The Spirit Catches You and You Fall down : A Hmong Child, Her American Doctors, and the Collision of Two Cultures*. New York: Farrar, Straus and Giroux.

Fajans, S S, Bell, G I and Polonsky, K S. 2001. "Molecular Mechanisms and Clinical Pathophysiology of Maturity-Onset Diabetes of the Young." *New England Journal of Medicine* 345 (13): 971–80.

Fearnhead, N S, Wilding, J L, Winney, B, Tonks, S, Bartlett, S, Bicknell, D C, Tomlinson, I P M, McC. Mortensen, N J and Bodmer, W F. 2004. "Multiple Rare Variants in Different Genes Account for Multifactorial Inherited Susceptibility to Colorectal Adenomas." *Proceedings of the National Academy of Sciences of the United States of America* 101 (45): 15992–97.

Fernandez-Canon, J M, Granadino, B, Beltran-Valero De Bernabe, D, Renedo, M, Fernandez-Ruiz, E, Penalva, M A and Rodriguez De Cordoba, S. 1996. “The Molecular Basis of Alkaptonuria.” *Nat Genet* 14 (1): 19–24.

Fisher, R A. 1930. *The Genetical Theory of Natural Selection*. *Clarendon Press, Oxford* .

Forno, L S. 1996. “Neuropathology of Parkinson’s Disease.” *Journal of Neuropathology & Experimental Neurology* 55 (3): 259–72.

Fowler, J S, Volkow, N D, Wang, G J, Pappas, N, Logan, J, MacGregor, R, Alexoff, D et al. 1996. “Inhibition of Monoamine Oxidase B in the Brains of Smokers.” *Nature* 379 (6567): 733–36.

Fromer, M, Moran, J L, Chambert, K, Banks, E, Bergen, S E, Ruderfer, D M, Handsaker, R E et al. 2012. “Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth.” *American Journal of Human Genetics* 91 (4): 597–607.

Funayama, M, Hasegawa, K, Kowa, H, Saito, M, Tsuji, S and Obata, F. 2002. “A New Locus for Parkinson’s Disease (PARK8) Maps to Chromosome 12p11.2-q13.1.” *Annals of Neurology* 51 (3): 296–301.

Furney, S J, Albà, M M and López-Bigas, N. 2006. “Differences in the Evolutionary History of Disease Genes Affected by Dominant or Recessive Mutations.” *BMC Genomics* 7 (1): 1–11.

Gabriel, S E, Brigman, K N, Koller, B H, Boucher, R C and Stutts, M J. 1994. “Cystic Fibrosis Heterozygote Resistance to Cholera Toxin in the Cystic Fibrosis Mouse Model.” *Science* 266 (5182): 107–9.

Gaddi, A, Cicero, A F G, Odoo, F O, Poli A, Paoletti, A R and On behalf of the Atherosclerosis and Metabolic Diseases Study Group. 2007. “Practical Guidelines for Familial Combined Hyperlipidemia Diagnosis: An up-Date.” *Vascular Health and Risk Management* 3 (6): 877–86.

Gao, F and Keinan, A. 2014. “High Burden of Private Mutations due to Explosive Human Population Growth and Purifying Selection.” *BMC Genomics* 15 (4): 1–7.

Garrod, A E. 1902. “The incidence of Alkaptonuria: a study in chemical individuality.” *The Lancet* 160 (4137): 1616–20.

Giasson, B I, Duda, J E, Murray, I V J, Chen, Q, Souza, J M, Hurtig, H I, Ischiropoulos, H, Trojanowski, J Q and Lee, V M Y. 2000. “Oxidative Damage Linked to Neurodegeneration by Selective α-Synuclein Nitration in Synucleinopathy Lesions.” *Science* 290 (5493): 985–89.

Gibson, G. 2012. “Rare and Common Variants: Twenty Arguments.” *Nat Rev Genet* 13 (2): 135–45.

Glass, Bentley. 1947. “Maupertuis and the Beginnings of Genetics.” *The Quarterly Review of Biology* 22 (3): 196–210.

Goh, K, Cusick, M E, Valle, D, Childs, B, Vidal, M and László Barabási, A. 2007. “The Human Disease Network.” *Proceedings of the National Academy of Sciences* 104 (21): 8685–90.

Goker-Alpan, O, Schiffmann, R, LaMarca, M, Nussbaum, R, McInerney-Leo, A and Sidransky, E. 2004. “Parkinsonism among Gaucher Disease Carriers.” *Journal of Medical Genetics* 41 (12): 937–40.

Golbe, L I, Lazzarini, A M, Schwarz, K O, Mark, M H, Dickson, D W and Duvoisin, R C. 1993. “Autosomal Dominant Parkinsonism with Benign Course and Typical Lewy-Body Pathology.” *Neurology* 43 (11): 2222–27.

Golbe, L I, Miller, D C and Duvoisin, R C. 1990. “Autosomal Dominant Lewy-Body Parkinson’s Disease.” *Advances in Neurology* 53: 287–92.

Goldman, S M. 2014. “Environmental Toxins and Parkinson’s Disease.” *Annual Review of Pharmacology and Toxicology* 54 (1): 141–64.

Green, R E, Krause, J, Briggs, A W, Maricic, T, Stenzel, U, Kircher, M, Patterson, N et al. 2010. “A Draft Sequence of the Neandertal Genome.” *Science* 328 (5979): 710–22.

Guevara-Aguirre, J, Balasubramanian, P, Guevara-Aguirre, M, Wei, M, Madia, F, Cheng, C W, Hwang, D et al. 2011. “Growth Hormone Receptor Deficiency Is Associated With a Major Reduction in Pro-Aging Signaling, Cancer and Diabetes in Humans.” *Science Translational Medicine* 3 (70): 70ra13–70ra13.

Guo, J, Zhang, X , Nie, L, Zhang, H, Liao, B, Li, J, Wang, L, Yan, X and Tang, B. 2010. “Mutation Analysis of Parkin, PINK1 and DJ-1 Genes in Chinese Patients with Sporadic Early Onset Parkinsonism.” *Journal of Neurology* 257 (7): 1170–75.

Gusev, A, Bhatia, G, Zaitlen, N, Vilhjalmsson, B J, Diogo, D, Stahl, E A, Gregersen, P K et al. 2013. “Quantifying Missing Heritability at Known GWAS Loci.” *PLoS Genet* 9 (12): e1003993.

Haldane, J B S. 1934. “A Mathematical Theory of Natural and Artificial Selection Part X. Some Theorems on Artificial Selection.” *Genetics* 19 (5): 412–29.

Haldane, J B S. 1941. “New Paths in Genetics.” London: Allen & Unwin.

Haldane, J b S. 1949a. “Disease and Evolution.” *Ricerca Science Supplement* 19: 3–10.

Haldane, J B S. 1949b. “The Rate of Mutation of Human Genes.” *Hereditas* 35 (S1): 267–73.

Hamosh, A, Scott, A F, Amberger, J S, Bocchini, C A and McKusick, V A. 2005. “Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders.” *Nucleic Acids Research* 33 (suppl 1): D514–17.

Hamza, T H and Payami, H. 2010. “The Heritability of Risk and Age at Onset of Parkinson’s Disease after Accounting for Known Genetic Risk Factors.” *Journal of Human Genetics* 55 (4): 241–43.

Hamza, T H, Zabetian, C P, Tenesa, A, Laederach, A, Montimurro, J, Yearout, D, Kay, D M et al. 2010. “Common Genetic Variation in the HLA Region Is Associated with Late-Onset Sporadic Parkinson’s Disease.” *Nat Genet* 42 (9): 781–85.

Hardy, G H. 1908. “Mendelian proportions in a mixed population.” *Science* 28 (706): 49–50.

Haslam, D W and James, W P T. 2005. “Obesity.” *The Lancet* 366 (9492): 1197–1209.

Hedrich, K, Eskelson, C, Wilmot, B, Marder, K, Harris, J, Garrels, J, Meija-santana, H et al. 2004. “Distribution, Type, and Origin of” 19 (10): 1146–57.

Hedrich, K, Kann, M, Lanthaler, A J, Dalski, A, Eskelson, C, Landt, O, Schwinger, E et al. 2001. “The Importance of Gene Dosage Studies: Mutational Analysis of the” 10 (16): 1649–56.

Hindorff, L A, Sethupathy, P, Junkins, H A, Ramos, E M, Mehta, J P, Collins, F S and Manolio, T A. 2009. “Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (23). National Academy of Sciences: 9362–67.

Hirschhorn, J N, Lohmueller, K, Byrne, E and Hirschhorn, K. 2002. “A Comprehensive Review of Genetic Association Studies.” *Genet Med* 4 (2): 45–61. 2.

Horaitis, O, Talbot, C C, Phommarinh, M, Phillips, K M and Cotton, R G H. 2007. “A Database of Locus-Specific Databases.” *Nat Genet* 39 (4): 425.

Hunt, D M, Dulai, K S, Bowmaker, J K and Mollon, J D. 1995. “The Chemistry of John Dalton’s Color Blindness.” *Science* 267 (5200): 984–88.

Hunter, D J. 2005. “Gene-Environment Interactions in Human Diseases.” *Nat Rev Genet* 6 (4): 287–98.

Hutchison, K E, Stallings, M, McGeary, J and Bryan, A. 2004. “Population Stratification in the Candidate Gene Study: Fatal

Threat or Red Herring?” *Psychological Bulletin*. 345:80309-0345.

Hutton, M, Lendon, C L, Rizzu, P, Baker, M, Froelich, S, Houlden, H, Pickering-Brown, S et al. 1998. “Association of Missense and 5[prime]-Splice-Site Mutations in Tau with the Inherited Dementia FTDP-17.” *Nature* 393 (6686): 702–5.

Ibáñez, P, Bonnet, A M, Débarges, B, Lohmann, E, Tison, F, Agid, Y, Dürr, A, Brice, A and Pollak, P. 2004. “Causal Relation between α-Synuclein Locus Duplication as a Cause of Familial Parkinson’s Disease.” *The Lancet* 364 (9440): 1169–71.

Ingebrigtsen, T, Thomsen, S F, Vestbo, J, van der Sluis, S, Kyvik, K O, Silverman, E K, Svartengren, M and Backer, V. 2010. “Genetic Influences on Chronic Obstructive Pulmonary Disease – A Twin Study.” *Respiratory Medicine* 104 (12): 1890–95.

International HapMap Consortium. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437 (7063): 1299–1320.

International Parkinson Disease Genomics Consortium. 2011a. “Imputation of Sequence Variants for Identification of Genetic Risks for Parkinson’s Disease: A Meta-Analysis of Genome-Wide Association Studies.” *The Lancet* 377 (9766): 641–49. d

International Parkinson’s Disease Genomics Consortium. 2011b. “A Two-Stage Meta-Analysis Identifies Several New Loci for Parkinson’s Disease.” *PLoS Genet* 7 (6): e1002142.

Ioannidis, J P A. 2005. “Why Most Published Research Findings Are False.” *PLoS Med* 2 (8): e124.

Jameson, J L and Kopp, P A. 2014. “Principles of Human Genetics.” In *Harrison’s Principles of Internal Medicine*. McGraw-Hill.

Ji, W, Foo, J N, O’Roak, B J, Zhao, H, Larson, M G, Simon, D B, Newton-Cheh, C, State, M W, Levy, D and Lifton, R P. 2008. “Rare Independent Mutations in Renal Salt Handling Genes Contribute to Blood Pressure Variation.” *Nat Genet* 40 (5): 592–99.

Kalia, L V and Lang, A E. 2016. "Parkinson's Disease." *The Lancet* 386 (9996): 896–912.

Kamboh, M I, Demirci, F Y, Wang, X, Minster, R L, Carrasquillo, M M, Pankratz, V S, Younkin, S G et al. 2012. "Genome-Wide Association Study of Alzheimer's Disease." *Transl Psychiatry* 2 (May): e117.

Keinan, A and Clark, A G. 2012. "Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants." *Science* 336 (6082): 740–43.

Keller, M F, Saad, M, Bras, J, Bettella, F, Nicolaou, N, Simón-Sánchez, J, Mittag, F et al. 2012. "Using Genome-Wide Complex Trait Analysis to Quantify 'missing Heritability' in Parkinson's Disease." *Human Molecular Genetics* 21 (22): 4996–5009.

Kilpinen, H and Barrett, J C. 2016. "How next-Generation Sequencing Is Transforming Complex Disease Genetics." *Trends in Genetics* 29 (1): 23–30.

Kitada, T, Hattori, N, Matsumine, H and Mizuno, Y. 1998. "Mutations in the Parkin Gene Cause Autosomal Recessive Juvenile Parkinsonism" 169 (1993): 166–69.

Koch, R. 1890. "Ueber Bakteriologische Forschung: Vortrag in Der 1." *Internationalen Medicinischen Congresses Am 4. August 1890. Berlin: Verlag von August Hirschwald.,* September.

Kondrashov, F A, Ogurtsov, A Y and Kondrashov, A S. 2004. "Bioinformatical Assay of Human Gene Morbidity." *Nucleic Acids Research* 32 (5): 1731–37.

Krüger, R, Kuhn, W, Müller, T, Woitalla, D, Graeber, M, Kösel, S, Przuntek, H, Epplen, J T, Schöls, L and Riess, O. 1998. "Ala30Pro Mutation in the Gene Encoding Alpha-Synuclein in Parkinson's Disease." *Nature Genetics* 18 (2): 106–8.

Kumar, S, Dudley, J T, Filipski, A and Liu, L. 2011. "Phylomedicine: An Evolutionary Telescope to Explore and Diagnose the Universe of Disease Mutations." *Trends in Genetics* 27 (9): 377–86.

La Du, B N, Zannoni, V G, Laster, L and Seegmiller, J E. 1958. “The nature of the defect in tyrosine metabolism in Alkaptonuria.” *Journal of Biological Chemistry* 230 (1): 251–60.

Laberge, A M, Michaud, J, Richter, A, Lemyre, E, Lambert, M, Brais, B and Mitchell, G A. 2005. “Population History and Its Impact on Medical Genetics in Quebec.” *Clinical Genetics* 68 (4): 287–301.

Lalueza-Fox, C and Gilbert, T P. 2011. “Paleogenomics of Archaic Hominins.” *Current Biology* 21 (24): R1002–9.

Lander, E S, Linton, L M, Birren, B, Nusbaum, C, Zody, M C, Baldwin, J, Devon, K et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–

Lappalainen, I, Almeida-King, J, Kumanduri, V, Senf, A, Dylan Spalding, J, Ur-Rehman, S, Saunders, G et al. 2015. “The European Genome-Phenome Archive of Human Data Consented for Biomedical Research.” *Nat Genet* 47 (7): 692–95.

Leinonen, R, Sugawara, H, Shumway, M and on behalf of the International Nucleotide Sequence Database Collaboration. 2011. “The Sequence Read Archive.” *Nucleic Acids Research* 39 (Database issue): D19–21.

Lesage, S, Anheim, M, Letournel, F, Bousset, L, Honoré, A, Rozas, N, Pieri, L et al. 2013. “G51D α-Synuclein Mutation Causes a Novel Parkinsonian–pyramidal Syndrome.” *Annals of Neurology* 73 (4): 459–71.

Li, Y, Tomiyama, H, Sato, K, Hatano, Y, Yoshino, H, Atsumi, M, Kitaguchi, M et al. 2005. “Clinicogenetic Study of PINK1 Mutations in Autosomal Recessive Early-Onset Parkinsonism.” *Neurology* 64 (11): 1955–57.

Lill, C M, Roehr, J T, McQueen, M B, Kavvoura, F K, Bagade, S, Schjeide, B M, Schjeide, L M et al. 2012. “Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson’s Disease Genetics: The PDGene Database.” *PLoS Genet* 8 (3): e1002548.

Lopez-Bigas, N and Ouzounis, C A. 2004. “Genome-Wide Identification of Genes Likely to Be Involved in Human Genetic Disease.” *Nucleic Acids Res* 32.

Lucking, C B, Abbas, N, Durr, A, Bonifati, V, Bonnet, A M, Broucker, T, Michele, G, Wood, N W, Agid, Y and Brice, A. 1998. “Homozygous Deletions in Parkin Gene in European and North African Families with Autosomal Recessive Juvenile Parkinsonism. The European Consortium on Genetic Susceptibility in Parkinson’s Disease and the French Parkinson’s Disease Genetics Study Group.” *Lancet* 352 (9137): 1355–56.

Lücking, C B, Bonifati, V, Periquet, M, Vanacore, N Brice, A and Meco, G. 2001. “Pseudo-Dominant Inheritance and Exon 2 Triplication in a Family with Parkin Gene Mutations.” *Neurology* 57 (5): 924–27.

MacArthur, D G, Balasubramanian, S, Frankish, A, Huang, N, Morris, J, Walter, K, Jostins, L et al. 2012. “A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes.” *Science* 335 (6070): 823–28.

Manolio, Teri a, Francis S Collins, Nancy J Cox, David B Goldstein, Lucia a Hindorff, David J Hunter, Mark I McCarthy, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461 (7265): 747–53.

Markopoulou, K, Wszolek, Z K, Pfeiffer, R F and Chase, B A. 1999. “Reduced Expression of the G209A α-Synuclein Allele in Familial Parkinsonism.” *Annals of Neurology* 46 (3): 374–81.

Marongiu, R, Brancati, F, Antonini, A, Ialongo, T, Ceccarini, C, Scarciolla, O, Capalbo, A et al. 2007. “Whole Gene Deletion and Splicing Mutations Expand the PINK1 Genotypic Spectrum.” *Human Mutation* 28 (1): 98.

Martin, E. 1809. “Hereditary Blindness.” *Medical and Physical Recorder Baltimore* 1: 273–79.

Matsumine, H, Saito, M, Shimoda-Matsubayashi, S, Tanaka, H, Ishikawa, A, Nakagawa-Hattori, Y, Yokochi, M et al. 1997.

“Localization of a Gene for an Autosomal Recessive Form of Juvenile Parkinsonism to Chromosome 6q25.2-27.” *American Journal of Human Genetics* 60 (3): 588–96.

Medawar, P B. 1952. “An Unsolved Problem of Biology.” *Evolution in Health and Disease*.

Medland, S E, Jahanshad, N, Neale, B M and Thompson, P M. 2014. “Whole-Genome Analyses of Whole-Brain Data: Working within an Expanded Search Space.” *Nat Neurosci* 17 (6): 791–800.

Mendel, G. 1951. “Versuche Über Pflanzen-Hybriden.” *Journal of Heredity* 42 (1): 3.

Merkle, S and Pretsch, W. 1989. “Characterization of Triosephosphate Isomerase Mutants with Reduced Enzyme Activity in Mus Musculus.” *Genetics* 123 (4): 837–44.

Miller, M P and Kumar, S. 2001. “Understanding Human Disease Mutations through the Use of Interspecific Genetic Variation.” *Human Molecular Genetics* 10 (21): 2319–28.

Moore, J H. 2003. “The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases.” In *Human Heredity*, 56:73–82.

Morgan, N V, Westaway, S K, Morton, J E V, Gregory, A, Gissen, P, Sonek, S, Cangul, H et al. 2006. “PLA2G6, Encoding a Phospholipase A2, Is Mutated in Neurodegenerative Disorders with High Brain Iron.” *Nat Genet* 38 (7): 752–54.

Morgan, T H. 1910. “Sex limited inheritance in *Drosophila*.” *Science* 32 (812): 120–22.

Morgan, T H. 1911. “Chromosomes and associative inheritance.” *Science* 34 (880): 636–38.

Nalls, M A, Pankratz, N, Lill, C M, Do, C B, Hernandez, D G, Saad, M, DeStefano, A L et al. 2014. “Large-Scale Meta-Analysis of Genome-Wide Association Data Identifies Six New Risk Loci for Parkinson’s Disease.” *Nature Genetics* 056 (9): 1–7.

Neel, J V. 1962. “Diabetes Mellitus: A ‘Thrifty’ Genotype Rendered Detrimental by ‘Progress’?” *American Journal of Human Genetics* 14 (4): 353–62.

Nejentsev, S, Walker, N, Riches, D, Egholm, M and Todd, J A. 2009. “Rare Variants of IFIH1, a Gene Implicated in Antiviral Responses, Protect Against Type 1 Diabetes.” *Science* 324 (5925): 387–89.

Ng, S B, Buckingham, K J, Lee, C, Bigham, A W, Tabor, H K, Dent, K M, Huff, C D et al. 2010. “Exome Sequencing Identifies the Cause of a Mendelian Disorder.” *Nat Genet* 42 (1): 30–35.

Noyce, A J, Bestwick, J P, Silveira-Moriyama, L, Hawkes, C H, Giovannoni, G, Lees, A J and Schrag, A. 2012. “Meta-Analysis of Early Nonmotor Features and Risk Factors for Parkinson Disease.” *Annals of Neurology* 72 (6): 893–901.

Nuytemans, K, Theuns, J, Cruts, M and Van Broeckhoven, C. 2010. “Genetic Etiology of Parkinson Disease Associated with Mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 Genes: A Mutation Update.” *Human Mutation* 31 (7): 763–80.

Orozco, G, Barrett, J C and Zeggini, E. 2010. “Synthetic Associations in the Context of Genome-Wide Association Scan Signals.” *Human Molecular Genetics* 19 (R2): R137–44.

Ottolini, D, Calì, T, Negro, A and Brini, M. 2013. “The Parkinson Disease-Related Protein DJ-1 Counteracts Mitochondrial Impairment Induced by the Tumour Suppressor Protein p53 by Enhancing Endoplasmic Reticulum–mitochondria Tethering.” *Human Molecular Genetics* 22 (11): 2152–68.

Ozelius, L J, Senthil, G, Saunders-Pullman, R, Ohmann, E, Deligtisch, A, Tagliati, M, Hunt, A L et al. 2006. “LRRK2 G2019S as a Cause of Parkinson’s Disease in Ashkenazi Jews.” *New England Journal of Medicine* 354 (4): 424

Paisan-Ruiz, C, Bhatia, K P, Li, A, Hernandez, D, Davis, M, Wood, N W, Hardy, J, Houlden, H, Singleton, A and Schneider, S A. 2009. “Characterization of PLA2G6 as a Locus for Dystonia-Parkinsonism.” *Annals of Neurology* 65 (1): 19–23.

Paisán-Ruíz, C, Jain, S, Evans, E W, Gilks, W P, Simón, J, van der Brug, M, López de Munain, A et al. 2004. “Cloning of the Gene Containing Mutations That Cause PARK8-Linked Parkinson’s Disease.” *Neuron* 44 (4): 595–600.

Pankratz, N, Beecham, G W, Destefano, A L, Dawson, T M, Doheny, K F, Factor, S A, Hamza, T H et al. 2012. “Meta-Analysis of Parkinson’s Disease: Identification of a Novel Locus, RIT2.” *Annals of Neurology* 71 (3): 370–84.

Pankratz, N, Wilk, J B, Latourelle, J C, DeStefano, A L, Halter, C, Pugh, E W, Doheny, K F et al. 2008. “Genomewide Association Study for Susceptibility Genes Contributing to Familial Parkinson Disease.” *Human Genetics* 124 (6): 593–605.

Park, S, Yang, J S, Kim, J, Shin, Y E, Hwang, J, Park, J, Jang, S K and Kim, S. 2012. “Evolutionary History of Human Disease Genes Reveals Phenotypic Connections and Comorbidity among Genetic Diseases.” *Scientific Reports* 2 (October): 757.

Parkes, E A. 2013. “Mode of Communication of Cholera. By John Snow, MD: Second Edition – London, 1855, Pp 162. .” *International Journal of Epidemiology* 42 (6): 1543–52.

Parrish, I. 1845. “Biographical Memoir of John C. Otto, M.D., Late Vice President of the College of Physicians Read before the College by Appointment, March 4th, 1845.” Philadelphia: W.F. Geddes, printer.

Pasteur, L. 1881. “On the germ theory.” *Science* os-2 (63): 420–22.

Pastor, P, Ezquerra, M, Muñoz, E, Martí, M J, Blesa, R, Tolosa, E and Oliva, R. 2000. “Significant Association between the Tau Gene A0/A0 Genotype and Parkinson’s Disease.” *Annals of Neurology* 47 (2): 242–45.

Pauling, L, Itano, H A, Singer, S J and Wells, I C. 1949. “Sickle Cell Anemia, a Molecular Disease.” *Science* 110 (2865): 543–48.

Pearson, T A, and Manolio, T A. 2008. “How to Interpret a Genome-Wide Association Study.” *JAMA* 299 (11): 1335–44.

Planansky, K. 1950. “Paralysis Agitans: A Clinical and Genetic Study.” *American Journal of Human Genetics* 2 (2): 200–201.

Podolsky, S, Leopold, N A and Sax, D S. 1972. “Increased frequency of diabetes mellitus in patients with Huntington’s chorea.” *The Lancet* 299 (7765): 1356–59.

Poewe, W. 2008. “Non-Motor Symptoms in Parkinson’s Disease.” *European Journal of Neurology* 15 (April): 14–20.

Polderman, T J C, Benyamin, B, de Leeuw, C A, Sullivan, P F, van Bochoven, A, Visscher, P M and Posthuma, D. 2015. “Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies.” *Nat Genet* 47 (7): 702–9.

Polimeni, J and Reiss, J P. 2002. “How Shamanism and Group Selection May Reveal the Origins of Schizophrenia.” *Medical Hypotheses* 58 (3): 244–48.

Polymeropoulos, M H, Lavedan, C, Leroy, E, Ide, S E, Dehejia, A, Dutra, A, Pike, B et al. 1997. “Mutation in the α-Synuclein Gene Identified in Families with Parkinson’s Disease.” *Science* 276 (5321): 2045–47.

Poolman, E M and Galvani, A P. 2007. “Evaluating Candidate Agents of Selective Pressure for Cystic Fibrosis.” *Journal of the Royal Society Interface* 4 (12): 91–98.

Preti, A and Miotto, P. 1997. “Creativity, Evolution and Mental Illnesses.” *Journal of Memetics - Evolutionary Models of Information Transmission* 1.

Pritchard, J K and Cox, N J. 2002. “The Allelic Architecture of Human Disease Genes: Common Disease-Common Variant...or Not?” *Human Molecular Genetics* 11 (20): 2417–23.

Proukakis, C, Dudzik, C G, Brier, T, MacKay, D S, Cooper, J M, Millhauser, G L, Houlden, H and Schapira, A H. 2013. “A Novel α-Synuclein Missense Mutation in Parkinson Disease.” *Neurology* 80 (11: 1062–64.

Prufer, K, Racimo, F, Patterson, N, Jay, F, Sankararaman, S, Sawyer, S, Heinze, A et al. 2014. “The Complete Genome Sequence of a Neanderthal from the Altai Mountains.” *Nature* 505 (7481): 43–49.

Pulst, S M. 1999. “Genetic Linkage Analysis.” *Archives of Neurology* 56 (6): 667–72.

Rafikova, O, Rafikov, R and Nudler, E. 2002. “Catalysis of S-Nitrosothiols Formation by Serum Albumin: The Mechanism and Implication in Vascular Control.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (9): 5913–18.

Ralser, M, Heeren, G, Breitenbach, M, Lehrach, H and Krobitsch, S. 2006. “Triose Phosphate Isomerase Deficiency Is Caused by Altered Dimerization?Not Catalytic Inactivity?of the Mutant Enzymes.” *PLoS ONE* 1 (1): e30.

Ramirez, A, Heimbach, A, Grundemann, J, Stiller, B, Hampshire, D, Cid, L P, Goebel, I et al. 2006. “Hereditary Parkinsonism with Dementia Is Caused by Mutations in ATP13A2, Encoding a Lysosomal Type 5 P-Type ATPase.” *Nat Genet* 38 (10): 1184–91.

Reich, D E and Lander, E S. 2001. “On the Allelic Spectrum of Human Disease.” *Trends in Genetics* 17 (9): 502–10.

Reich, D E, Green, R E, Kircher, M, Krause, J, Patterson, N, Durand, E Y, Viola, B et al. 2010. “Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia.” *Nature* 468 (7327): 1053–60.

Riederer, P, Konradi, C, Hebenstreit, G and Youdim, M B H. 1989. “Neurochemical Perspectives to the Function of Monoamine Oxidase.” *Acta Neurologica Scandinavica* 80 (November): 41–45.

Riederer, P, Sofic, E, Rausch, W D, Schmidt, B, Reynolds, G P, Jellinger, K and Youdim, M B H. 1989. “Transition Metals, Ferritin, Glutathione, and Ascorbic Acid in Parkinsonian Brains.” *Journal of Neurochemistry* 52 (2): 515–20.

Risch, N, Tang, H, Katzenstein, H and Ekstein, J. 2003.

“Geographic Distribution of Disease Mutations in the Ashkenazi Jewish Population Supports Genetic Drift over Selection.” *The American Journal of Human Genetics* 72 (4): 812–22.

Ristow, M. 2004. “Neurodegenetive Disorders Associated with Diabetes Mellitus.” *Journal of Molecular Medicine* 82 (8): 510–29.

Sánchez-Danés, A, Richaud-Patin, Y, Carballo-Carbajal, I, Jiménez-Delgado, S, Caig, C, Mora, S, Di Guglielmo, C et al. 2012. “Disease-specific Phenotypes in Dopamine Neurons from Human iPS-based Models of Genetic and Sporadic Parkinson’s Disease.” *EMBO Molecular Medicine* 4 (5): 380–95.

Sankararaman, S, Mallick, S, Dannemann, M, Prufer, K, Kelso, J, Paabo, S, Patterson, N and Reich, D E. 2014. “The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans.” *Nature* 507 (7492): 354–57.

Sarkisyan, K S, Bolotin, D A, Meer, M V, Usmanova, D R, Mishin, A S, Sharonov, G V, Ivankov, D M et al. 2016. “Local Fitness Landscape of the Green Fluorescent Protein.” *Nature* 533 (7603): 397–401.

Satake, W, Nakabayashi, Y, Mizuta, I, Hirota, Y, Ito, C, Kubo, M Kawaguchi, T et al. 2009. “Genome-Wide Association Study Identifies Common Variants at Four Loci as Genetic Risk Factors for Parkinson’s Disease.” *Nat Genet* 41 (12): 1303–7.

Schapira, A H V. 2005. “Present and Future Drug Treatment for Parkinson’s Disease.” *Journal of Neurology, Neurosurgery & Psychiatry* 76 (11): 1472–78.

Schneider, A S, Valentine, W N, Hattori, M and Heins, H L. 1965. “Hereditary Hemolytic Anemia with Triosephosphate Isomerase Deficiency.” *New England Journal of Medicine* 272 (5): 229–35.

Schneider, A, Westwood, B, Yim, C, Cohen-Solal, M, Rosa, R, Labotka, R, Eber, S, Wolf, R, Lammi, A and Beutler, E. 1996. “The 1591C Mutation in Triosephosphate Isomerase (TPI) Deficiency. Tightly Linked Polymorphisms and a Common

Haplotype in All Known Families.” *Blood Cells, Molecules, and Diseases* 22 (2): 115–25.

Schultheis, P J, Hagen, T T, O’Toole, K K, Tachibana, A, Burke, C R, McGill, D L, Okunade, G W and Shull, G E. 2004. “Characterization of the P5 Subfamily of P-Type Transport ATPases in Mice.” *Biochemical and Biophysical Research Communications* 323 (3): 731–38.

Scott, W K, Nance, M A, Watts, R L et al. 2001. “Complete Genomic Screen in Parkinson Disease: Evidence for Multiple Genes.” *JAMA* 286 (18): 2239–44.

Selkoe, D J. 1999. “Translating Cell Biology into Therapeutic Advances in Alzheimer’s Disease.” *Nature* 399 (Supplementary): A23–31.

Sharma, M, Ioannidis, J P A, Aasly, J O, Annesi, G, Brice, A, Van Broeckhoven, C, Bertram, L, et al. 2012. “Large-Scale Replication and Heterogeneity in Parkinson Disease Genetic Loci.” *Neurology* 79 (7).: 659–67.

Shojaee, S, Sina, F, Banihosseini, S S, Kazemi, M H, Kalhor, R, Shahidi, G A, Fakhrai-Rad, H, Ronaghi, M and Elahi, E. 2008. “Genome-Wide Linkage Analysis of a Parkinsonian-Pyramidal Syndrome Pedigree by 500 K SNP Arrays.” *The American Journal of Human Genetics* 82 (6): 1375–84.

Shulman, J M, De Jager, P L and Feany, M B. 2011. “Parkinson’s Disease: Genetics and Pathogenesis.” *Annual Review of Pathology* 6 (January): 193–222.

Simon-Sanchez, J, Scholz, S, del Mar Matarin, M, Fung, H C, Hernandez, D, Gibbs, J R, Britton, A, Hardy, J, and Singleton, A. 2008. “Genomewide SNP Assay Reveals Mutations Underlying Parkinson Disease.” *Human Mutation* 29 (2): 315–22.

Simon-Sanchez, J, Schulte, C, Bras, J M, Sharma, M, Gibbs, J R, Berg, D, Paisan-Ruiz, C et al. 2009. “Genome-Wide Association Study Reveals Genetic Risk Underlying Parkinson’s Disease.” *Nat Genet* 41 (12).: 1308–12.

Simonti, C N, Vernot, B, Bastarache, L, Bottinger, E, Carrell, D S, Chisholm, R L, Crosslin, D R et al. 2016. “The Phenotypic Legacy of Admixture between Modern Humans and Neandertals.” *Science* 351 (6274): 737–41.

Singleton, A. 2003. “alpha-Synuclein Locus Triplication Causes Parkinson’s Disease.” *Science* 302 (5646): 841–841.

Singleton, A and Hardy, J. 2011. “A Generalizable Hypothesis for the Genetic Architecture of Disease: Pleomorphic Risk Loci.” *Human Molecular Genetics* 20 (R2): R158–62.

Sinibaldi, L, De Luca, A, Bellacchio, E, Conti, E, Pasini, A, Paloscia, C, Spalletta, G, Caltagirone, C, Pizzuti, A and Dallapiccola, B. 2004. “Mutations of the Nogo-66 Receptor (RTN4R) Gene in Schizophrenia.” *Human Mutation* 24 (6): 534–35.

Smith, N G C and Eyre-Walker, A. 2003. “Human Disease Genes: Patterns and Predictions.” *Gene* 318 (October): 169–75.

Smyth, D J, Plagnol, V, Walker, N M, Cooper, J D, Downes, K, Yang, J H M, Howson, J M M et al. 2008. “Shared and Distinct Genetic Variants in Type 1 Diabetes and Celiac Disease.” *The New England Journal of Medicine* 359 (26): 2767–77.

Spataro, N, Calafell, F, Cervera-Carles, L, Casals, F, Pagonabarraga, J, Pascual-Sedano, B, Campolongo, A et al. 2014. “Mendelian Genes for Parkinson’s Disease Contribute to the Sporadic Forms of the Disease.” *Human Molecular Genetics* 24 (7): 2023–34.

Spira, P J, Sharpe, D M, Halliday, G, Cavanagh, J and Nicholson, G A. 2001. “Clinical and Pathological Features of a Parkinsonian Syndrome in a Family with an Ala53Thr α-Synuclein Mutation.” *Annals of Neurology* 49 (3): 313–19.

Stoneking, M and Krause, J. 2011. “Learning about Human Population History from Ancient and Modern Genomes.” *Nat Rev Genet* 12 (9): 603–14.

Strachan, D P. 1989. “Hay Fever, Hygiene, and Household Size.” *BMJ : British Medical Journal* 299 (6710): 1259–60.

Subramanian, S and Kumar, S. 2006. “Evolutionary Anatomies of Positions and Types of Disease-Associated and Neutral Amino Acid Mutations in the Human Genome.” *BMC Genomics* 7 (1): 1–9.

Sullivan, P F, Kendler, K S and Neale, M C. 2003. “Schizophrenia as a Complex Trait: Evidence from a Meta-Analysis of Twin Studies.” *Archives of General Psychiatry* 60 (12): 1187–92.

Sun, M, Latourelle J C, Wooten, G et al. 2006. “Influence of Heterozygosity for Parkin Mutation on Onset Age in Familial Parkinson Disease: The Genepd Study.” *Archives of Neurology* 63 (6): 826–32.

Sutton, W S. 1903. “The chromosomes in heredity.” *The Biological Bulletin* 4 (5): 231–50.

Tabor, H K, Risch, N J and Myers, R M. 2002. “Candidate-Gene Approaches for Studying Complex Genetic Traits: Practical Considerations.” *Nat Rev Genet* 3 (5): 391–97.

Tan, E K, Chua, E, Fook-Chong, S M, Teo, Y Y, Yuen, Y, Tan, L and Zhao, Y. 2007. “Association between Caffeine Intake and Risk of Parkinson’s Disease among Fast and Slow Metabolizers.” *Pharmacogenetics and Genomics* 17 (11).

Tattersall, I. 2009. “Human Origins: Out of Africa.” *Proceedings of the National Academy of Sciences* 106 (38): 16018–21.

Tennessen, J A, Bigham, A W, O’Connor, T D, Fu, W, Kenny, E E, Gravel, S, McGee, S et al. 2012. “Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes.” *Science* 337 (6090): 64–69.

The 1000 Genomes Project Consortium. 2012. “An Integrated Map of Genetic Variation from 1,092 Human Genomes.” *Nature* 491 (7422): 56–65.

The 1000 Genomes Project Consortium. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526 (7571): 68–74.

Thomas, P D and Kejariwal, A. 2004. “Coding Single-Nucleotide Polymorphisms Associated with Complex vs. Mendelian

Disease: Evolutionary Evidence for Differences in Molecular Effects.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (43): 15398–403.

Tu, Z, Wang, L, Xu, M, Zhou, X, Chen, T and Sun, F. 2006. “Further Understanding Human Disease Genes by Comparing with Housekeeping Genes and Other Genes.” *BMC Genomics* 7 (1): 1–13.

Valente, E M, Abou-Sleiman, P M, Caputo, V, Muqit, M M K, Harvey, K, Gispert, S, Ali, Z et al. 2004. “Hereditary Early-Onset Parkinson’s Disease Caused by Mutations in PINK1.” *Science* 304 (5674): 1158–60.

van der Mark, M, Brouwer, M, Kromhout, H, Nijssen, P, Huss, A and Vermeulen, R. 2012. “Is Pesticide Use Related to Parkinson Disease? Some Clues to Heterogeneity in Study Results.” *Environmental Health Perspectives* 120 (3): 340–47.

van Heyningen, V and Yeyati, P L. 2004. “Mechanisms of Non-Mendelian Inheritance in Genetic Disease.” *Human Molecular Genetics* 13 (REV. ISS. 2): 225–33.

Venter, J C, Adams, M D, Myers, E W, Li, P W, Mural, R J, Sutton, G G, Smith, H O et al. 2001. “The Sequence of the Human Genome.” *Science* 291 (5507): 1304–51.

Vilariño-Güell, C, Wider, Ross, O A, Dachsel, J C, Kachergus, J M, Lincoln, S J, Soto-Ortolaza, A I et al. 2011. “VPS35 Mutations in Parkinson Disease.” *The American Journal of Human Genetics* 89 (1): 162–67.

Visscher, Peter M, Matthew A Brown, Mark I McCarthy, and Jian Yang. 2012. “Five Years of GWAS Discovery.” *American Journal of Human Genetics* 90 (1): 7–24.

Visscher, P M, Hill, W G and Wray, N R. 2008. “Heritability in the Genomics Era: Concepts and Misconceptions.” *Nat Rev Genet* 9 (4): 255–66.

de Vries, H. 1901. “Die Mutationstheorie. Versuche Und Beobachtungen Über Die Entstehung von Arten Im Pflanzenreich.” Leipzig: Veit & comp.

Wamelink, M M C, Grüning, N M, Jansen, E E W, Bluemlein, K, Lehrach, H, Jakobs, C and Ralser, M. 2010. "The Difference between Rare and Exceptionally Rare: Molecular Characterization of Ribose 5-Phosphate Isomerase Deficiency." *Journal of Molecular Medicine* 88 (9): 931–39.

Weinberg, W. 1908. "Über Den Nachweis Der Vererbung Beim Menschen ." *Jahreshefte Des Vereins Für Vaterländische Naturkunde in Württemberg* 64 : 369–82.

Welter, D, MacArthur, J, Morales, J, Burdett, T, Hall, P, Junkins, H, Klemm, A et al. 2014. "The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations." *Nucleic Acids Research* 42 (Database issue): D1001–6.

Wider, C, Skipper, L, Solida, A, Brown, L, Farrer, M, Dickson, D, Wszolek, Z K and Vingerhoets, F J G. 2008. "Autosomal Dominant Dopa-Responsive Parkinsonism in a Multigenerational Swiss Family." *Parkinsonism & Related Disorders* 14 (6): 465–70.

Williams, G C. 1957. "Pleiotropy, Natural Selection, and the Evolution of Senescence." *Evolution* 11 (4): 398–411.

Wilson, T W. 1986. "History of salt supplies in West Africa and blood pressures today." *The Lancet* 327 (8484): 784–86.

Winston, J T, Koepp, D M, Zhu, C, Elledge, S J and Harper, J W. 1999. "A Family of Mammalian F-Box Proteins." *Current Biology* 9 (20): 1180–S3.

Wolfe, N D, Dunavan, C P and Diamond, J. 2007. "Origins of Major Human Infectious Diseases." *Nature* 447 (7142): 279–83.

Woolf, L I. 1986. "The Heterozygote Advantage in Phenylketonuria." *American Journal of Human Genetics* 38 (5): 773–75.

Wray, N R, Purcell, S M and Visscher, P M. 2011. "Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results." *PLoS Biol* 9 (1): e1000579.

Wright, A, Charlesworth, B, Rudan, I, Carothers, A and Campbell, H. 2003. “A Polygenic Basis for Late-Onset Disease.” *Trends in Genetics : TIG* 19 (2): 97–106.

Wright, S. 1932. “The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution.” *Proceedings of the Sixth International Congress of Genetics* 1:356-66.

Wszolek, Z K, Cordes, M, Calne, D B, Munter, M D, Cordes, I and Pfeifer, R F. 1993. “Hereditary Parkinson’s Disease: Report on Three Families with Autosomal-Dominant Inheritance Pattern .” *Nervenarzt* 64 (5): 331–35.

Zarranz, J J, Alegre, J, Gómez-Esteban, J C, Lezcano, E, Ros, R, Ampuero, I, Vidal, L et al. 2004. “The New Mutation, E46K, of α-Synuclein Causes Parkinson and Lewy Body Dementia.” *Annals of Neurology* 55 (2): 164–73.

Zhu, Q, Ge, D, Maia, J M, Zhu, M, Petrovski, S, Dickson, S P, Heinzen, E L, Shianna, K V and Goldstein, D B. 2011. “A Genome-Wide Comparison of the Functional Properties of Rare and Common Genetic Variants in Humans.” *The American Journal of Human Genetics* 88 (4): 458–68.

Zimprich, A, Biskup, S, Leitner, P, Lichtner, P, Farrer, M, Lincoln, S, Kachergus, J et al. 2004. “Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology.” *Neuron* 44 (4): 601–7.