Chapter 1

Introduction

1.1 The Human Genome

1.1.1 What is a genome?

The genome is the entire genetic information of an organism. The human genome is composed of nuclear and mitochondrial DNA. It has been estimated that the nuclear genome contains approximately protein-coding 30000 genes. Therefore, less than 2% of the human genome encodes for proteins, the reminder is composed of repetitive DNA sequences, which may play a role in the modulation of gene expression [1].

1.1.2 Human genetic variations

Even though any two genomes are roughly 99.9% identical, the analysis of human DNA sequences has revealed the existence of millions of variants among the 3.2 billion bases of the human genome [2, 3]. These sequence variations are important

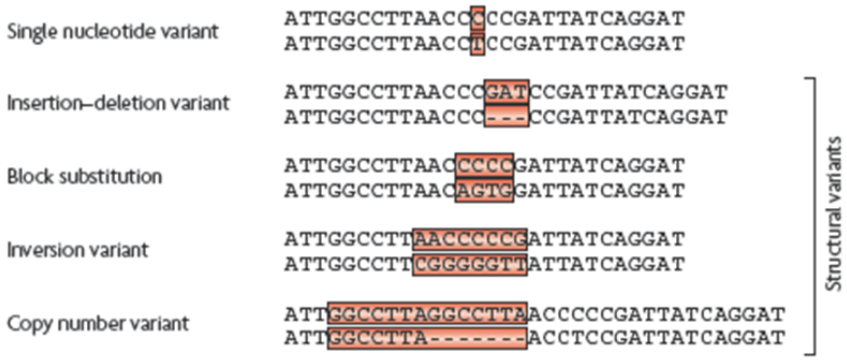

Figure 1.1: **Types of human genetic variants.**Adapted from Frazer et al. (2009) [5].

since they are responsible for the heritable phenotypic variation among individuals, including disease susceptibility and quantitative trait variability [2]. The 1000 Genomes Project has led to the identification of 15 million variants with a frequency higher than 1% in each one of different population groups from Europe, Asia, Africa and the Americas. It has been estimated that these represent over 95% of the human genome variants accessible to the current technologies [4].

Variants in the human genome can be classified into two different types: single nucleotide variants and structural variants (Figure 1.1) [5].

Single nucleotide polymorphisms (SNPs) Single nucleotide polymorphisms are the most common type of variants in the human genome. A single nucleotide polymorphism (SNP) is defined as a sequence variation in which a DNA base is substituted by another with a frequency greater than 1%. In order to create a public resource of single nucleotide polymorphisms (SNPs) in the human genome, in

1999 the SNP Consortium (TSC) was constituted. The initial aim was to identify 300000 SNPs in two years. At the end of 2001, 1.4 million SNPs were released into the public domain, exceeding the initial expectations [6]. A large collaborative effort to identify and characterize SNPs has, subsequently, been undertaken by the International HapMap Project and the 1000 Genomes Project [7–9]. Currently around 15 million SNPs have been identified in the human genome [10].

Structural variants Structural variants include insertions or deletions of one or more bases, block substitutions, inversions and copy number variants [11]. Insertions and deletions involve, respectively, the addition or loss to one or more nucleotide in the DNA sequence. More than 1 million of short indels have been identified during the pilot of the 1000 Genome Project [10]. Block substitutions consist a series of adjacent nucleotides which differs between two genomes [11]. Inversions are rearrangements in which the order of the DNA sequence is reversed in a specific chromosomal region. Copy number variants are segments of DNA which are present at a different copy number compared to the reference genome [12]. As part of the 1000 Genomes project, a map of copy number variants (CNVs) based on whole genome DNA sequencing of 185 human samples has been created. It included 22025 deletions and 6000 other structural variants (SVs), including insertions and tandem duplications [13]. The identification and characterization structural variants (SVs) of the human genome is important, since in some genomic regions SVs can influence gene dosage and therefore they could contribute to disease susceptibility [12].

1.1.3 The Human Genome Project

The Human Genome Project was organized mainly to map and sequence the human genome. It was proposed by US Department of Energy (DOE) and the National Institutes of Health, initially as a 15-year project, at the end of the 1980s. It started formally in 1990 and was completed in 2003, two years earlier than planned, due to the rapid improvement of required technologies. During the early stages, the Wellcome Trust (UK) gave its contribution to the project and subsequently organizations in France, Japan, China and other countries joined. The project goals were not only to map and sequence the human genome, but also to develop new technologies to study the human genome, to improve tools to store and analyze the genomic data and to address ethical, legal and social issue (ELSI). In addition to the Human Genome Project, genome projects for other organisms, such as *Escherichia coli* and *Mus musculus* (mouse) started. Sequencing the genomes of other organisms has been proved to be important for the functional characterization of the human orthologous genes [14]. In 1998 the physical map of 30181 human genes was released [15]. On the 25^{th} of June 2000, President Bill Clinton announced the completion of the first draft of the entire human genome. In February 2001 the initial working draft of the human genome sequence was published [16, 17]. In April 2003, the completion of the Human Genome Project was announced and in 2004 the finished human genome sequence was published [3].

1.1.4 The ENCODE Project

Following the release of the finished human genome sequence, the scientific and medical communities recognized the need to gain a better understanding of the

functional elements encoded in the human genome sequence. Therefore, in September 2003 the National Human Genome Research Institute (NHGRI) launched the Encyclopedia of DNA Elements (ENCODE) project, aimed at identifying and characterize all these functional elements. A wide variety of experimental and computational methods were employed to annotate genes and regulatory regions of the human genome (Figure 1.2). Gene annotation was done mainly through manual curation. The annotation process involved the collection of all evidences of transcripts supported by experimental data from public databases and experimental validation by RT-PCR for novel transcribed loci. Pseudogenes were detected by similarity to other protein-coding genes. The resulting catalog of ENCODE annotated genes was termed GENCODE. Regulatory elements were identified through DNase hypersensitivity assays, measurement of DNA methylation, and mapping of histone modifications and transcription factor (TF) binding sites by chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-Seq). In order to use the obtained data to interpret the role of human genetic variations associated with diseases, the ENCODE consortium is working to integrate their data with those from other large-scale studies, such as the 1000 Genomes Project [18].

1.1.5 The HapMap Project

The International HapMap Project started in 2002 with the aim of characterizing the common DNA sequence variants and creating a resource to be used to identify genes underlying diseases and influencing the response to drugs [7]. The Project was a collaboration between scientists in different countries, such as Japan, the United Kingdom, Canada, China and the United State (www.genome.gov). A to-

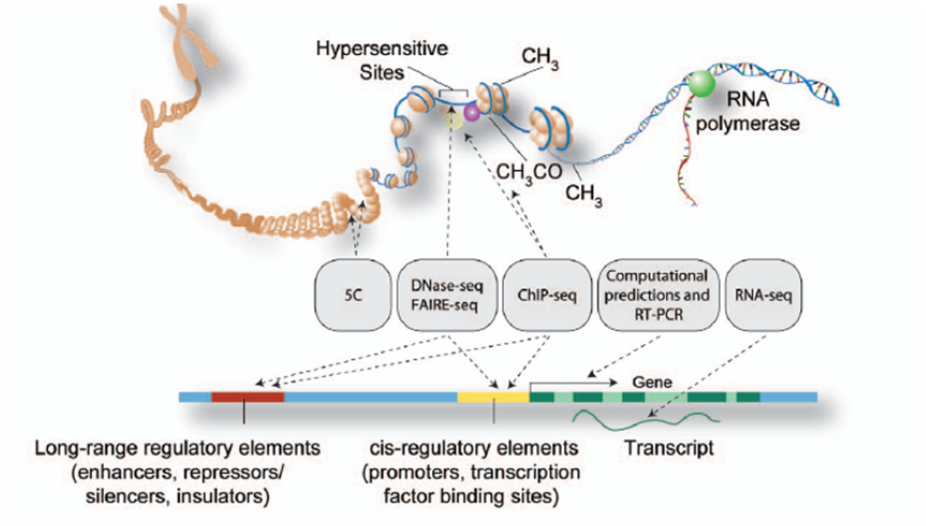

Figure 1.2: **Organization of the ENCODE project**
Representation of the methods used to detect functional elements (gray boxes). Adapted from The ENCODE Project Consortium et al. (2011) [18]

tal of 270 DNA samples from four different populations were collected: 30 trios of northern and western European ancestry living in US Utah (CEU); 30 trios of Yoruba people living in Ibadan, Nigeria (YRI); 45 unrelated Japanese living in Tokyo, Japan (JPT); and 45 unrelated Han Chinese living in Beijing, China (CHB) [7]. In Phase I of the project, the goal was to genotype at least one common SNP every 5 kilobases (kb) across the genome in each of one of the collected DNA samples [8]. More than 1.3 million SNPs were genotyped, of which 1 million passed quality control filters and were polymorphic in the studied samples [8]. The description of these data, including a fine-scale genetic map spanning the human genome, was released in 2005 [8].

In Phase II of the project, an additional 2 million SNPs were genotyped in the same sample collection. The resulting SNPs data were estimated to include approximately 25–35% of all the 9–10 million common SNPs (minor allele frequency

(MAF) greater than 0.05) in the human genome [9]. This phase of the provided increased resolution of the fine-scale genetic map spanning the human genome, improving the power to detect recombination hotspots [9].

In phase III of the project, in order to generate a data set of common and rare both SNPs and copy number polymorphisms (CNPs), 1.6 million common SNPs were genotyped in a larger collection of samples, including 1184 individuals from 11 populations and ten 100-kilobase regions were sequenced in 692 of these individuals. In this phase of the study, population-specific differences among low-frequency variants was estimated and the improvement in imputation accuracy using the larger reference panel was shown [19]. The International HapMap Project has been cardinal for the realization of well-powered, large-scale, genome-wide association studies [9].

1.1.6 The 1000 Genomes Project

The 1000 Genomes Project was set in 2008 with the aim of identifying most of the polymorphic DNA variants of the human genome (allele frequency of 1% or higher) [10]. Being too expensive to deeply sequence the whole genome and given that specific genomic regions contain a limited number of haplotypes , the overall plan was to to sequence at low-coverage (4x) about 2500 samples and to combine the data to impute in each sample the variants not directly detected by the low-coverage sequencing (http://www.1000genomes.org).

To determine whether the strategy of the overall plan was adequate, three pilot projects were undertaken: low-coverage (2-4x) whole-genome sequencing of 179 samples from four populations; high-coverage (20-60x) sequencing of two trios; and

target sequencing of 8140 exons in 697 individuals from seven populations. The results of the analysis of the pilot data were published in October 2010 [10]. Approximately 15 million SNPs, 1 million short insertions and deletions, and 20000 additional structural variants (SVs) were detected and their location, allele frequency and surrounding haplotype structure was defined [10]. These data provided provided a better understanding of the human genetic variation and have already been proven useful to identify SNPs associated with complex diseases [4].

1.2 Investigating the role of genetic in complex diseases

1.2.1 Recurrence risk ratio and heritability

Evidence of the role played by genetic factors in conferring susceptibility to diseases is the familial aggregation. Familial aggregation is usually assessed calculating the recurrence risk ratio. The recurrence risk ratio in relatives is the ratio between the prevalence of the disease in relatives and the prevalence of the disease in the population. For several complex diseases, on average a recurrence risk ratio in relatives of 2 has been found. Interpreting the meaning of the recurrence risk ratio in relatives, it has to be considered that it is a measure of all the factors, genetic and environmental, which contribute to familial aggregation [20].

An estimation of the importance of genetic factors in determining the susceptibility to a disease is provide by the heritability. The heritability (h^2) is defined as the proportion of the total phenotypic variance which is explained by the genetic variance. The heritability can be estimated calculating concordance rates

Table 1.1: **Estimates of heritability.**

Disorder	Prevalence(%)	Heritability(%)	
Hypertension	29	50	[23]
Obesity	22	77	[24]
Migraine	15	57	[25]
Asthma	12	75	[26]
Coronary artery disease	7	32	[27]
Schizophrenia	4	85	[28]

in monozygotic and dizygotic twins [21]. In table 1.1 estimates of heritability for some common disorders are shown.

Published recurrence risk ratio and heritability could be overestimates owing to the increased shared environment in monozygotic twins compared to dizygotic twins [22].

1.2.2 Identification of causative genes

Several strategies have been used to identify susceptibility genes for complex disorders.

Linkage analysis Linkage analysis aims to identify a disease locus analyzing co-segregation of the trait of interest with genetic markers using a family based approach. Linkage studies have been most successful in mapping genomic loci containing causal variants for mendelian diseases [29, 30]. However, they have had limited success in identifying risk loci for complex diseases. The likely reason is that the complex diseases are probably due not only to rare genetic variants, but

also to common genetic variants for the detection of which linkage studies are underpowered [31–33]. It has been estimated that loci with a moderate effect could not be detected even with 3000 sib pairs [34]. Moreover, linkage analysis is more difficult for complex diseases compared to monogenic disorders, since the underlying genetic component is likely to be polygenic and some complex disorders are probably genetically heterogeneous with different mechanism involved in different subtypes [21]

Association studies Genetic association studies are performed comparing the frequency of one or more genetic variants in affected subjects with the frequency in a matched control group, in order to establish whether there is an association between any of the tested variants and the disease [35]. When an association between a genetic variant and the disease is found, there are three possible scenarios underlying the finding: 1) the variant has a causal role in the disease (direct association), 2) the variant has not a causal role in itself but it is linked to the causal variant (indirect association), or 3) existence of confounding factors, such as population stratification or admixture, generating false association [35].

Risch (2000) has estimated that case–control association studies provide adequate power to detect variants with minor allele frequency higher than 10% and a genetic relative risks (GRR) as low as 1.5 (Figure 1.3) [34].

The association studies initially performed were candidate gene association studies, in which genes to be tested were chosen on the basis of previous positional and functional information and a limited number of genetic variants were tested for association with the disease [36]. The availability of catalogues of common sequence variants (i.e. HapMap) and the advances in genotyping technologies made

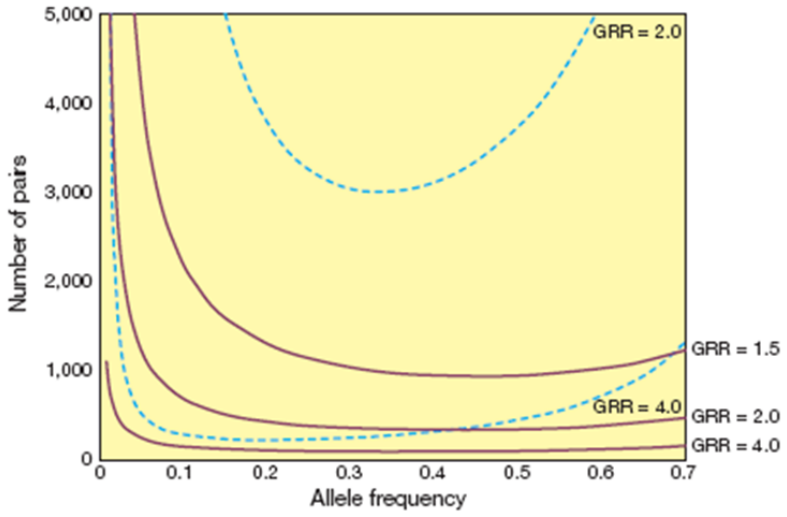

Figure 1.3: **Comparison of linkage with association analysis for detecting genetic effects**
Linkage analysis (dashed lines) is based on affected sib pairs (ASPs) considering a completely linked and informative marker. Association analysis (continuous lines) is based on casecontrol pairs. A multiplicative model is assumed, where the genotype relative risk (GRR) of the high-risk homozygote is the square of the value of GRR for the heterozygote, which is given in the figure. For variants with high relative risks (GRR ≥ 4) and intermediate minor allele frequencies (MAF = 0.050.50) linkage analysis can identify the disease locus. However, for variants with modest relative risks (GRR ≥ 2), linkage analysis requires unrealistically large samples. By contrast, adequate power for the detection of variants with modest GRR can be provided by casecontrol association studies [34]. Adapted from Rich (2000) [34].

possible to perform genome-wide association studies (GWAS), in which variants are tested for association with the disease, without any previous assumption [37]. GWAS have proven to be an effective approach to identify common genetic variants which confer disease susceptibility.

The first published GWAS demonstrated the association between age-related macular degeneration (AMD) and a common variant in the complement factor H gene (CFH). Even if the sample size was small (96 cases and 50 controls), it was possible to identify the association between AMD and CFH since the variant had a large effect on the disease risk [38].

In 2006–2007 several GWAS were published. These led to the identification of common genetic variants associated with several common diseases, such as obesity [39], coronary heart disease [40], type 1 and type 2 diabetes [36, 40–43].

Since then, hundreds of GWAS have been completed and have led to the identification of more than 1000 variants associated with complex diseases and traits (Figure 1.4 and 1.5) [36, 37, 40, 44–50].

The results of the GWAS completed so far have improved our understanding of the genetic structure of complex diseases. Since most of the associated common genetic variants are outside coding regions, it has been suggested that variation in gene expression regulation has an important role in complex diseases [36]. In contrast with monogenic disorders, it has been found that usually the effect size of each associated variant detected through GWAS was modest and explained only a small fraction of the phenotypic variance in the population [36, 51–53].

It has been suggested that rare variants, copy number variations, gene-gene interactions and epigenetic mechanisms may be the source of the 'missing heritability' [54]. Moreover, it is possible that part of the unexplained heritability is

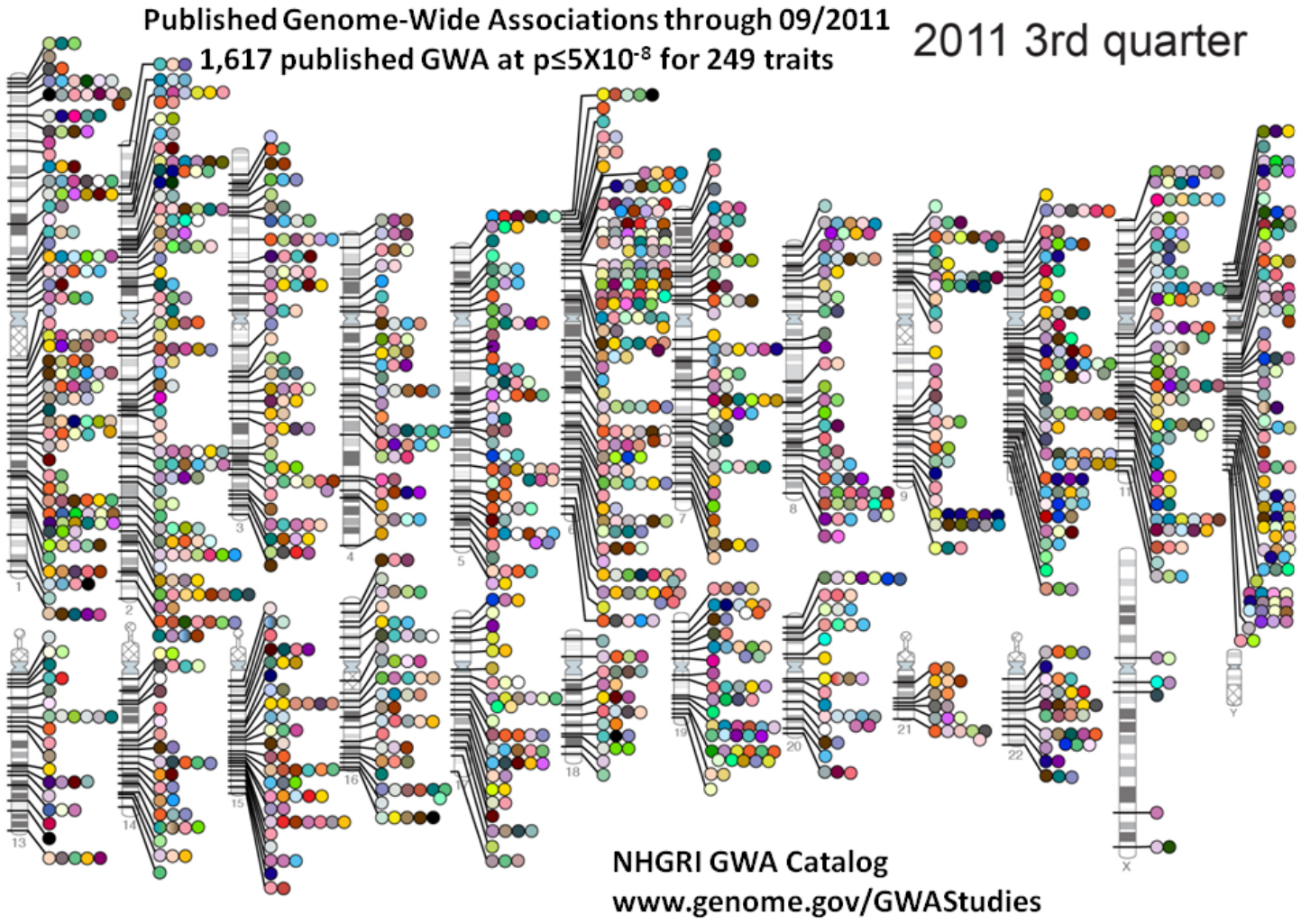

Figure 1.4: **Published GWAS.** Adapted from http://www.genome.gov/gwastudies/.

Figure 1.5: **Published GWAS.** Legend of Figure 1.4

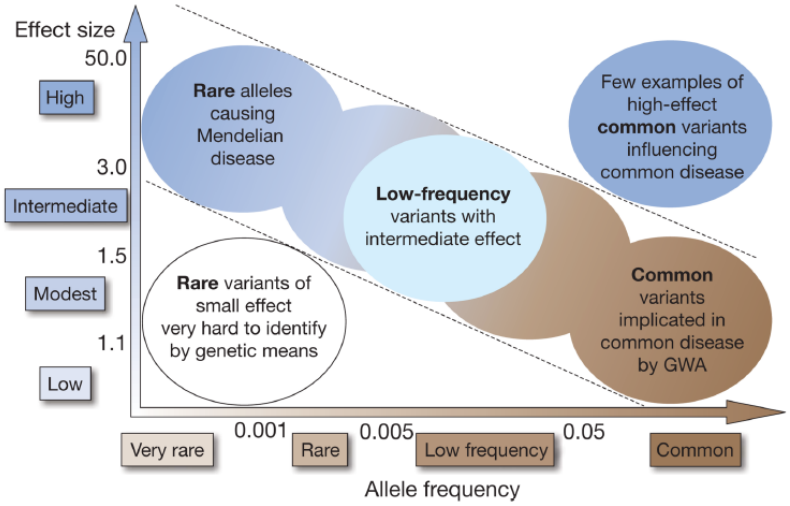

Figure 1.6: **Genetic variants frequency and diseases susceptibility**
GWAS performed so far have mainly identified variants of frequency higher than 5% and modest effect size. Higher penetrance and lower frequency variants could be probably identified in the future thanks to the advances in sequencing technologies. Adapted from Manolio et al. (2009) [54].

due to additional common variants of small effect, which have not been detected yet and they might be discovered increasing the sample size (Figure 1.6) [36].

Exome and whole genome sequencing Nowadays, the identification of rare variants which influence disease susceptibility has become possible thanks to technological advances in sequencing and bioinformatics approaches [55]. Over the last decade, the shift away from Sanger sequencing, considered as first–generation technology, to new massively parallel methods referred to as next-generation sequencing (NGS) has allowed reduced costs of sequencing [56, 57]. In order to identify rare variants associated with complex diseases, the most comprehensive study design will eventually involve sequencing the whole genome in a large number of recruited individuals [51]. However, whole-genome sequencing is still too expensive

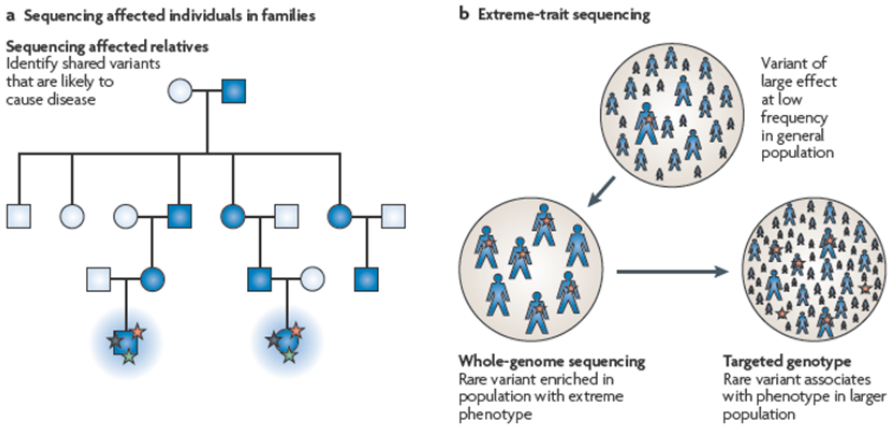

Figure 1.7: **Strategies for identifying rare variants**
In (a) shaded individuals are affected, and sequencing individuals at one or both ends of a trait distribution. In (b) the size of the individual represents the severity of the phenotype. Adapted from Cirulli et al. (2010) [51]

to be applicable to large sample sizes. Two study designs are currently used to discover rare variants underlying diseases: family-based sequencing and extreme-trait sequencing (Figure 1.7) [51]. The first design (family-based sequencing) consists in sequencing cases from families that have multiple affected individuals [51]. Initially the most distantly affected relatives could be sequenced and then rare shared variants could be followed-up checking their segregation in the family and their presence in independent cases [51]. The second design (extreme-trait sequencing) consists in sequencing individuals who are at the extreme ends of a phenotype distribution. Identified variants could be followed-up checking the co-segragation in the available families and genotyping them in larger sets of samples. This approach could be of value in identifying rare variants, which are not private and have modest to high effect sizes [51].

Another strategy which could be pursued, until the cost of whole-genome sequencing drops, is to sequence specific regions, such as the exome [51,56]. However, it has to be considered that whole-exome sequencing has its limitations, since structural variants are often not detectable and non coding regulatory regions, which may have a role in conferring diseases susceptibility, are not targeted.

1.3 Migraine

1.3.1 Clinical features

Migraine is a common neurological disease characterized by unilateral, pulsating, recurrent headaches of moderate-severe intensity, aggravated by physical activity, lasting for $4 - 72$ hours. Headache attacks can be associated with nausea, vomiting, phonophobia and photophobia [58]. The unilateral and pulsating features of migraine are the most characteristic ones compared to other types of headache. When the pain is severe, patients tend to look for a dark and quiet place, where they can lie down and try to sleep. Some patients associate the onset of the migraine attack to particular food (cheese, chocolate and fatty food), alcoholic drinks (red wine) and high caffeine intake or caffeine withdrawal. In some women, the attacks tend to occur in the premenstrual period [59].

A migraine attack can have four phases: the prodromal phase, the aura phase, the headache phase and the postdromal phase (Figure 1.8) [58]. The prodromal phase may precede the headache onset by several hours or days. Patients may experience irritability, euphoria, fatigue, depression, drowsiness, craving for particular food, dizziness, constipation, diarrhea, polyuria or other visceral manifesta-

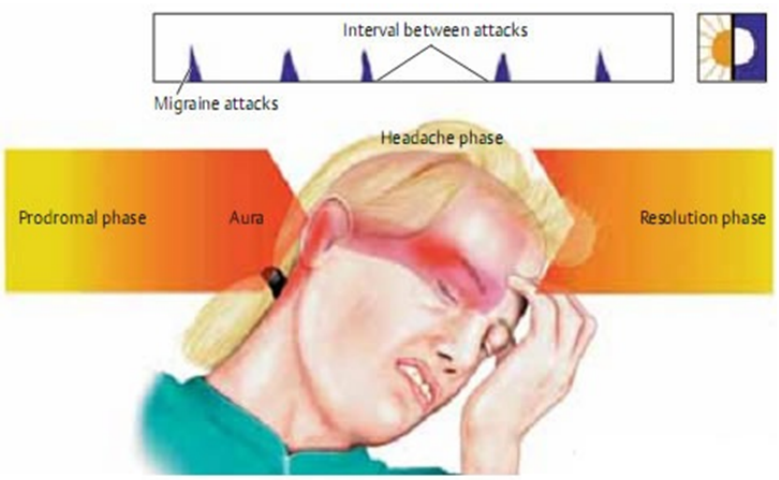

Figure 1.8: **Migraine attack phases.** Adapted from Rohkamm et al (2004)

tions [60]. The aura phase is experienced by one third of migraine patients [61,62]. The aura consists of transient neurological symptoms, which usually precede the headache [63]. Visual symptoms are the most common form of aura (99%), followed by sensory, aphasic and motor symptoms. The aura usually develops over 5–20 minutes and lasts less than one hour. The onset of the migraine headache is usually gradual. The headache typically it is unilateral and pulsating. However, it may be bilateral at the onset or start on one side and spread to the other side. During this phase patients may experience nausea, vomiting, photophobia, phonophobia, osmophobia, sensitivity to sounds and lights [58]. The postdromal phase may last for several days after the headache has resolved. Patients report increased sensitivity in the area in which the migraine headache was located, fatigue, mood changes and impaired concentration [59].

Between migraine attacks, the migrainous patients are asymptomatic [59].

1.3.2 Classification and diagnosis

Classification and diagnosis of migraine are based on diagnostic criteria established by the International Headache Society (IHS) (Table 1.2 and 1.3) [58].

In the IHS classification two main types of migraine are described: migraine without aura (MO) and migraine with aura (MA). MA is observed in one third of migraine patients [61,62]. Individuals may have attacks of only migraine with aura (pure MA) or they may have a combination of attacks both with and without aura in variable proportions (MA/MO). Since both types of migraine share identical headache symptoms and frequently co-occur in the same individual, it is thought that they are expressions of the same disease, with pure MO and pure MA at the ends of a wide clinical spectrum [63–65].

1.3.3 Epidemiology

Migraine is a common condition. The life-time prevalence of migraine is about 15% in the developed countries. Migraine is most prevalent in Europe and least prevalent in Africa (Figure 1.9) [66]. The prevalence is higher among women, peaks in the middle of life and declines thereafter (Figure 1.10) [67]. Migraine is a disease of public health interest since not only its costs are huge, €27 billion for whole Europe in 2004, but it is also associated with high disability and reduced quality of life [68]. Migraine patients suffer a median of 12 migraine attacks per year, 25% have at least two attacks per month and 10% have weekly attacks [63, 69]. Given its frequency and the severity of the symptoms, migraine can be associated with reduced work performance, impaired family and social life, leading to substantial personal suffering [70, 71].

Table 1.2: **Classification of migraine subtypes according to the International Headache Society.**

1.	**MIGRAINE**
1.1.	**Migraine without aura (MO)**
1.2.	**Migraine with aura (MA)**
	1.2.1. Typical aura with migraine headache
	1.2.2. Typical aura with non-migraine headache
	1.2.3. Typical aura without headache
	1.2.4. Familial hemiplegic migraine (FHM)
	1.2.5. Sporadic hemiplegic migraine (SHM)
	1.2.6. Basilar-type migraine
1.3.	**Childhood periodic syndromes that are commonly precursors of migraine**
1.4.	**Retinal migraine**
1.5.	**Complications of migraine**
1.6.	**Probable migraine**

Table 1.3: **International Headache Society diagnostic criteria for migraine without and with aura.**

1.1.	**Migraine without aura (MO)**
A	At least 5 attacks fulfilling criteria B-D.
B	Headache attacks lasting 4-72 hours (untreated or unsuccessfully treated).
C	Headache has at least two of the following characteristics: unilateral location pulsating quality moderate or severe pain intensity aggravation by or causing avoidance of routine physical activity (eg, walking or climbing stairs)
D	During headache at least one of the following: nausea and/or vomiting photophobia and phonophobia
E	Not attributed to another disorder.
1.2.	**Migraine with aura (MA)**
A	At least 2 attacks fulfilling criterion B
B	Migraine aura fulfilling criteria B-C for one of the subforms 1.2.1-1.2.6
C	Not attributed to another disorder
1.2.1	**Typical aura with migraine headache**
A	At least 2 attacks fulfilling criteria B-D
B	Aura consisting of at least 1 of the following, but no motor weakness: 1. Fully reversible visual symptoms including positive and/or negative features 2. Fully reversible sensory symptoms including positive and/or negative features 3. Fully reversible dysphasic speech disturbance
C	At least two of the following: 1. Homonymous visual symptoms and/or unilateral sensory symptoms 2. At least one aura symptom develops gradually over 5 minutes and/or different aura symptoms occur in succession over $\geq$ 5 minutes 3. Each symptom lasts $\geq$ 5 and $\leq$ 60 minutes
D	Headache fulfilling criteria B-D for Migraine without aura begins during the aura or follows aura within 60 minutes
E	Not attributed to another disorder

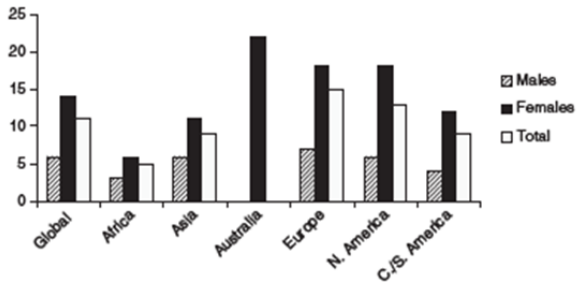

Figure 1.9: **Prevalence of migraine in adults of different countries.**
Adapted from Stovner at al. 2007 [66]

Acute and prophylactic treatments of migraine are not optimal [63]. Triptans, used as acute treatment, lead to pain relief in only 33-57% of cases [72] . The three most commonly prescribed prophylactic agents (propranolol, amitriptyline and verapamil) prevent migraine in only 50% of the patients [73].

1.3.4 Comorbid disorders

Migraine has been found to be comorbid with other neurological disorders, psychiatric diseases and cardiovascular disorders. The association between migraine and epilepsy has been shown in several studies. Both migraine and epilepsy are episodic neurological disorders, they share pathophysiological mechanisms and some clinical features. It has been shown that both disorders may be due to mutations in the same genes (*CACNA1A*, *ATP1A2*, and *SCN1A*) and that migraine can precede or follow an epileptic seizure [74, 75]

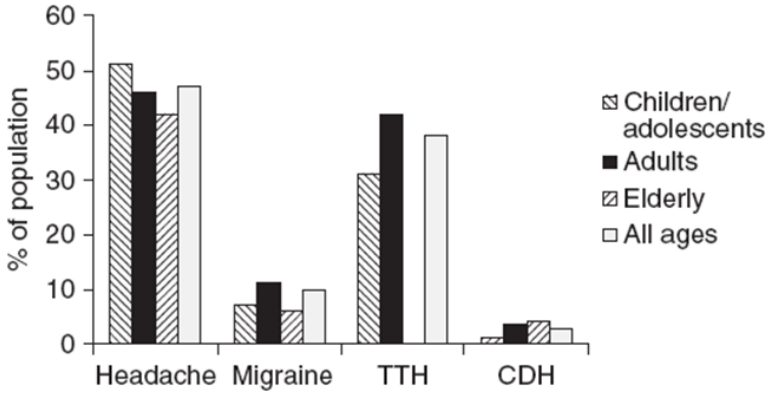

Figure 1.10: **Prevalence of different headaches in different age categories.** TTH, Tension-type headache; CDH, chronic daily headache. Adapted from Stovner at al. 2007 [66]

Migraine has been found to be associated with ischemic stroke in several studies. This association is bidirectional, since it has been shown that ischemic stroke can develop as a complication of an attack of migraine with aura and that migraine with aura is a risk factor for the development of ischemic stroke [76].

Both migraine with and without aura have been found to be associated with cardiovascular diseases (CVD) and with risk factors for CVD. In a recent population study conducted on 6102 migraineurs and 5243 controls, Bigal et al. (2010) found that migraine was associated with an increased risk of myocardial infarction, stroke, and claudicatio. Moreover, migraineurs were found to have, more frequently than controls, diabetes, hypertension and high cholesterol [77].

Several studies have shown an increased risk of affective and anxiety disorders in patients with migraine compared to the general population. The presence of a

comorbid psychiatric disorders may affect migraine evolution. It has been shown that the presence of a psychiatric disorder is a risk factor for the transformation of migraine into a chronic form [78]. Identifying the comorbid disorders is important since it can improve the management of the patient.

1.3.5 Pathogenesis

The pathogenic mechanisms underlying migraine are poorly understood. Migraine is currently considered to be a neurovascular disorder [63]. There is evidence that cortical spreading depression (CSD) is the pathophysiological substrate of the aura symptoms [79]. CSD is a process characterized by a slowly propagating wave of neuronal depolarization, followed by a suppression of neuronal activity [80, 81]. This neuronal event is accompanied by changes in cortical blood oxygenation, which progress slowly and contiguously over the cerebral cortex. An initial increase in blood oxygenation, possibly reflecting vasodilatation, is followed by a reduction in cortical blood oxygenation, possibly reflecting vasoconstriction [82].

Although CSD was initially considered to be present only in MA, CSD-like changes have been shown to occur during MO [83]. Therefore, it has been suggested that a silent aura occurs in MO and that the two main forms of migraine (MA and MO) share a common pathogenesis [84]. The link between CSD and migraine has been also provided by the observation that CSD activates the trigeminovascular system and evokes a series of events consistent with the development of the headache [85, 86] . Bolay et al. (2002) have shown that metabolites released in the extracellular space, during CSD, activate perivascular trigeminal afferents [86]. The activated trigeminal nerves transmit impulses to trigeminal

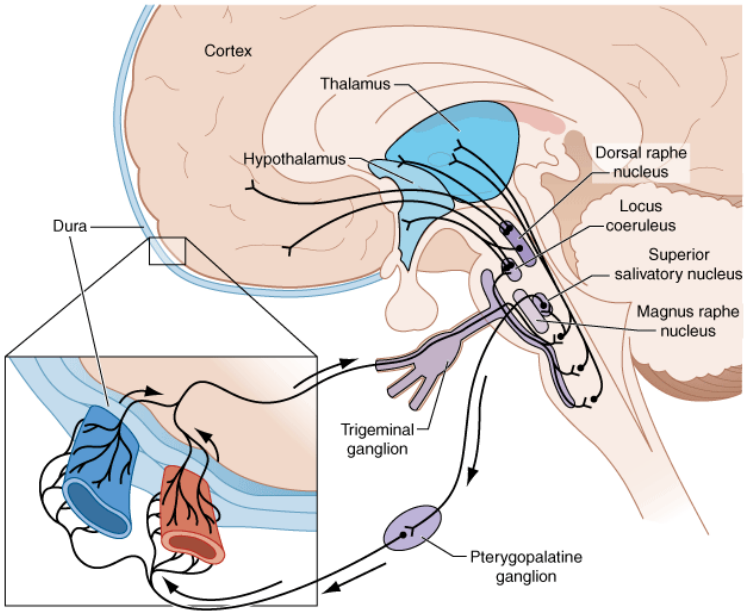

Figure 1.11: **Anatomical structures involved in migraine pathophysiology.** Adapted from Fauci at al. 2008

ganglia (TG) and trigeminal nucleus caudalis (TNC), which in turn send signals rostrally to brain structures involved in processing the pain, leading to the generation of the headache (Figure 1.11 and 1.12).

1.3.6 Genetic basis

Heritability of migraine Migraine is a multifactorial disorder, hence both environmental and genetic factors contribute to its susceptibility [88]. Genetic predisposition to migraine is suggested by a significantly higher risk of migraine in first degree relatives compared to the general population (RR = 3.6–3.7), an estimated heritability of 34%–57% [25, 89] and a higher proband-wise concordance rate in monozygotic twins (34–40%) than in dizygotic twins (12–28%) [88, 90–92]. The two main types of migraine (MA and MO) can coexist in the same family,

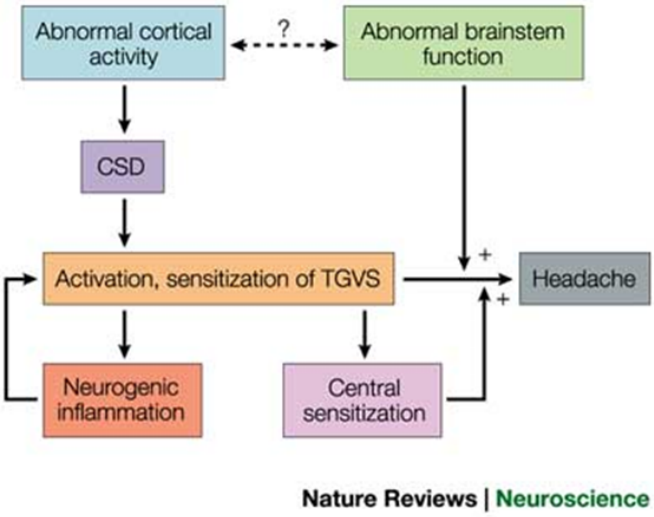

Figure 1.12: **Pathophysiological mechanisms in the generation of migraine headache.** Adapted from Pietrobon at al. 2003 [87]

suggesting a common genetic component [69].

The identification of the genetic basis of migraine is important since it could provide insights into pathogenesis and lead to the identification of new targets for both acute and preventive treatments.

Familial Hemiplegic Migraine The only type of migraine for which the genetic bases have been identified is hemiplegic migraine (HM) [25, 93]. HM is a rare form of MA, in which the aura is characterized by motor weakness of variable intensity (Table 1.4) [58]. Mutations in three genes, two encoding ion-channel subunits (*CACNA1A* and *SCN1A*) and one encoding a Na^+/K^+-ATPase subunit (*ATP1A2*) have been described in several familial and a few sporadic HM cases (Table 1.5) [94–104].

Table 1.4: **International Headache Society diagnostic criteria for Familial hemiplegic migraine (FHM).**

1.2.4	**Familial hemiplegic migraine (FHM)**
A	At least 2 attacks fulfilling criteria B-C.
B	Aura consisting of fully reversible motor weakness and at least one of the following:
	1. Fully reversible visual symptoms including positive features (ie. flickering lights, spots or lines) and/or negative features (ie. loss of vision)
	2. Fully reversible sensory symptoms including positive (ie. pins and needles)and/or negative features (ie. numbness)
	3. Fully reversible dysphasic speech disturbance
C	At least two of the following:
	1. At least one aura symptom develops gradually over 5 minutes
	and/or different aura symptoms occur in succession over $\geq$ 5 minutes
	2. Each aura symptom lasts $\geq$ 5 and $\leq$ 24 hours
	3. Headache fulfilling criteria B-D for Migraine without aura
	begins during the aura or follows aura within 60 minutes
D	At least one first or second degree relative has had attacks fulfilling these criteria A-E
E	Not attributed to another disorder

CACNA1A mutations are responsible for more than 50% of all the familial hemiplegic migraine (FHM) cases [105]. *CACNA1A* variants have also been found in few cases of sporadic hemiplegic migraine (SHM) [95,103]. Some *CACNA1A* mutations can cause a severe phenotype, in which HM is associated with epilepsy [106]. Mutations in the same gene have been described in two other neurological disorders: episodic ataxia type 2 (EA2) and spinocerebellar ataxia type 6 (SCA 6) [101, 107]. *CACNA1A* encodes the pore-forming subunit ($\alpha 1$) of the voltage-dependent calcium channel (VDCC) P/Q-type. VDCCs are heteromultimeric complexes which mediate calcium entry into excitable cells. VDCC P/Q-type is the main channel involved in neurotransmitter release at many synapses in the central nervous system [108]. *CACNA1A* mutations have been shown to increase the channel open probability and to slow down the channel inactivation, leading to an increase in the Ca^{++} influx and neurotransmitter release (Figure 1.13) [109–111]. A knock-in mouse model, carrying a human *CACNA1A* mutation, shows a reduced threshold of CSD [112].

SCN1A encodes the α subunit of the neuronal voltage-gated sodium channel, essential for the generation and propagation of action potentials in neurons. *SCN1A* mutations were initially associated to epilepsy syndromes [113–116]. Subsequently, *SCN1A* mutations have been described in pure FHM and SHM [95, 96]. Recently, *SCN1A* mutations have been found in families with co-segregating HM and epilepsy [117, 118]. *SCN1A* mutations have been shown to determine either loss or gain of function features, indicating a complex relationship between clinical phenotype and altered sodium channel function [96, 118, 119].

ATP1A2 encodes the $\alpha 2$ subunit of the Na^+/K^+ transporting ATPase 1. Na/K-ATPases are integral membrane transporters, which maintains the chemical gra-

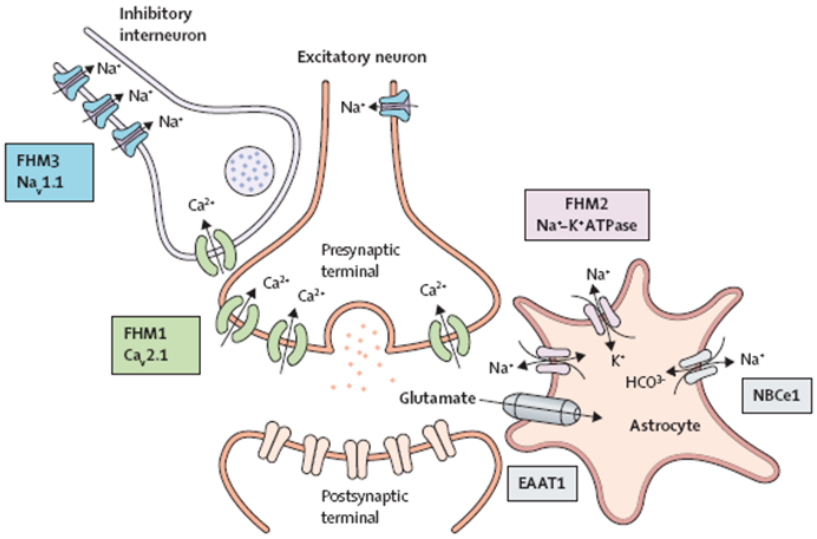

Figure 1.13: **Roles of proteins encoded by genes involved in FHM at a CNS glutamatergic synapse** Adapted from Russell et al. (2011) [126]

dients of Na and K ions across the plasma membrane [120]. These gradients are known to modulate neuron excitability [121]. *ATP1A2* mutations have been described not only in FHM and SHM, but also in more severe cases in which HM is associated with epilepsy or mental retardation [94, 95, 103, 117, 122–124]. *ATP1A2* mutations can lead to a partial or complete loss of activity of the Na^+/K^+-ATPase, which increases neuronal depolarization and might trigger CSD and CSD-like events (Figure 1.12) [125].

Common migraine Genes predisposing to other types of migraine remain unknown, even though several genomic regions have been suggested in linkage studies (Table 1.6), and several genes have been tested in candidate-gene association studies [162–184].

Linkage studies have been successful for localizing regions containing rare ge-

netic variants responsible for mendelian diseases [29, 30]. However, they have had limited success in mapping risk loci in complex diseases, since the genetic susceptibility to complex diseases, such as migraine, is probably conferred not only by rare genetic variants, but also by common variants with moderate effects, for the detection of which linkage studies are underpowered [31–33].

Compared to linkage studies, association studies have higher statistical power to detect common variants that confer modest disease susceptibility [187]. Candidate-gene association studies have identified several variants associated to migraine. However, they have not been replicated. The lack of well-replicated findings is probably due to a limited number of cases and controls tested in most of the studies performed so far (Table 1.7) [173–184].

1.4 This thesis

Migraine is a common neurological disease, affecting 15% of the population in the developed countries. There is evidence that migraine is a multifactorial disorder with complex pathogenesis, which involves both genetic and environmental factors. The genetic susceptibility to migraine is probably conferred both by rare genetic variants with a strong contribution to disease risk and by common genetic variants of small to moderate effect.

To identify common susceptibility variants for migraine, we carried out a large multistage GWAS. In the discovery stage, more than 400,000 SNPs were genotyped in 2748 migraine patients and in 10747 population-matched controls from Finland, Germany and the Netherlands. In the replication stage, the SNP most associated

with migraine was tested in further 3202 migraine cases and 40062 population-matched controls from Denmark, Germany and Iceland. The analysis of the data obtained in the discovery and replication stage of the first GWAS for migraine is discussed in chapter 3.

The hundreds of thousands of SNPs directly assayed in the initial GWAS for migraine represent only a fraction of the millions of SNPs contained in the human genome. Genotype imputation is useful to join together datasets genotyped on different platforms and to evaluate association with a phenotype at variants that are not directly genotyped. In order to identify further susceptibility loci, we carried out a large multistage GWAS of imputed SNPs using as reference 1000 Genomes data (December 2010 release). In the discovery stage, we imputed around 11000000 SNPs in 5403 migraine patients and in 15327 population-matched controls from Finland, Germany and the Netherlands. SNPs that were associated with migraine were carried forward for replication in 3268 case and 2916 control samples from Finland, the Netherlands and Spain. The analysis of the imputed data obtained in the discovery stage and of the typed data obtained in the replication stage is discussed in chapter 4.

Genome-wide association studies (GWAS) have been successful in identifying common variants associated with complex human diseases and traits [194]. The development of next generation sequencing (NGS) is allowing the systematic discovery of rare variants in thousands of samples [55]. The combination of next generation sequencing technologies with efficient methods of sequence capture has

enabled the widespread targeting and analysis of the exome [195–199]. As part of a pilot study aimed at identifying rare variants associated with complex neurological diseases, we sequenced the exomes of five individuals with epilepsy. Analyzing the called variants we realized that genes with a potential impact on the studied phenotype, such as ion channel subunits, were not captured by the used CCDS based capture array. To address this shortcoming, we designed and experimentally tested a more complete set of target regions for the human exome, based on the GENCODE annotation [200]. The design and comparison of the GENCODE exome with previously available exome capture sets is described in chapter 5.

To identify rare variants associated with complex diseases, one type of approach, which is currently affordable, consists in sequencing cases from families that have multiple affected individuals (family-based sequencing) [51]. The identification of genes underlying rare familial forms of migraine, such as familial hemiplegic migraine (FHM), can improve our knowledge of the molecular mechanisms underlying common migraine. In order to identify novel coding variants contributing to the genetic basis of FHM, we sequenced the exome of 88 cases from 44 families with several affected. The analysis of the obtained sequencing data is discussed in chapter 6.

Table 1.5: **Ion transportation genes and familial hemiplegic migraine.**

	FHM1	FHM2	FHM3
Chromosome location	19p13	1q23	2q24
Gene	*CACNA1A*	*ATP1A2*	*SCN1A*
Protein	Pore-forming alpha1 subunit of neuronal CaV2.1 (P/Q type) voltage-gated calcium channels	Catalytic alpha2 subunit of a glial and neuronal sodiumpotassium pump	Pore-forming alpha1 subunit of neuronal NaV1.1 voltage-gated sodium channels
Type of mutation	More than 30 different missense mutations [101, 124, 127–146] large-scale deletions [158] deletion in the promoter [159]	More than 60 different missense mutations [94, 95, 99, 117, 122, 123, 147–154] other types of rare mutations: small deletions, truncating, frameshift, loss of stop codon [99]	Five missense mutations [96, 155–157]
Penetrance[a]	6789% [124, 160]	6387% [124, 161]	100% [96, 155–157]

[a]Penetrance was calculated on the basis of results from cited studies. The table has been adapted from Russell et al. (2011) [126]
FHM, familial hemiplegic migraine

Table 1.6: **Loci identified in linkage studies**

	Chromosome	Study population	Reference
Migraine	1q31	Australian	[166]
Migraine	3q	Australian	[167]
MO	4q21	Icelandic	[164]
MA	4q24	Finnish	[172]
Migraine	5q21	Australian	[170]
Migraine	6p12.2–q21.1	Swedish	[165]
Migraine	10q22–q23	Finnish and Australian	[163]
MA	11q24	Canadian	[185]
MO	14q21.2–q22.3	Italian	[186]
Migraine	18p11	Australian	[167]
Migraine	18q12	Finnish	[162]
Migraine	19p13	Australian	[169]
Migraine	Xq	Australian	[168]

Table 1.7: **Candidate gene association studies for migraine**

Gene	Chromosome	SNP	Sample size cases/controls	Population	Phenotype	P	Reference
Estrogen receptor 1 (*ESR1*)	6q25.1	rs2228480	484/484	Australian	[a]Migraine	8×10^{-6}	[173]
		rs2228480	[b]286/142	Spanish	[a]Migraine	NS	[182]
		rs1801132	[b]286/142	Spanish	[a]Migraine	0.008	[182]
		rs1801132	484/484	Australian	[a]Migraine	NS	[188]
		rs1801132	898/900	Finnish	MA	NS	[176]
Progesterone receptor (*PGR*)	11q22-23	PROGINS Alu insertion	275/275	Australian	[a]Migraine	0.02	[189]
			[c]300/300	Australian	[a]Migraine	0.019	[189]
Insulin receptor (*INSR*)	19p13.3-13.2	rs2860172	313/466	North American	[a]Migraine	0.041	[178]
			275/275	[c]Australian	[a]Migraine	0.17	[178]
		rs2860172	270/280	German	MA	0.11	[180]
Methylentetrahydrofolate reductase (*MTHFR*)	1p36.22	rs1801133	168/269	Australian	MA	0.017	[177]
		rs1801133	230/204	Spanish	Migraine	NS	[181]
		rs1801133	898/900	Finnish	MA	NS	[176]
		rs1801133	186/493	Portuguese	Migraine	NS	[190]
Dopamine beta–hydroxylase (*DBH*)	9q34	rs1611115	170/176	Australian	[a]Migraine	0.004	[191]
			[c]245/279	Australian	[a]Migraine	0.013	[191]
		rs2097629	650/650	German	MA	0.0012	[192]
Dopamine receptor D2 (*DRD2*)	11q23	rs7131056	650/650	German	MA	0.0018	[192]
		rs2283265	263/274	Spanish	Migraine	0.003	[193]
			[c]259/287	Spanish	Migraine	NS	[193]
Solute carrier family 6 member 3 (*SLC6A3*)	5p15.3	rs40184	650/650	German	MA	0.0082	[192]
Low-density lipoprotein receptor (*LDR*)	19p13.2	(TA)n in exon 18	360/200	Italian	[a]Migraine	0.010	[179]
		(TA)n in exon 18	244/244	Australian	[a]Migraine	0.742	[175]

[a]Migraine without and with aura.
[b]Analysis limited to female samples
[c]Replication cohort