Uncovering the role of common and rare variants in migraine

Maria Stella Calafato, MD

Supervisor: Professor Aarno Palotie
Second supervisor: Professor John Todd
Sponsor: Professor David Lomas

Wellcome Trust Sanger Institute
University of Cambridge

Corpus Christi College

This dissertation is submitted for the degree of Doctor Philosophy

December 2011

To my dear parents

Matteo Calafato and Nina Tuccari

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation is not substantially the same as any that they may have submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution, except for Chapter 3, the content of which has been submitted by Verneri Anttila for his PhD at the University of Helsinki. This dissertation does not exceed the word limit (60000 words) set by the Biology Degree Committee.

Abstract

Migraine is a paroxysmal disorder of the nervous system. In order to uncover the genetic architecture underlying migraine, we performed a genome-wide association study (GWAS) of typed variants, a GWAS of imputed variants, and a pilot whole exome sequencing of familial migraine samples. In the GWAS of typed variants, a SNP (single nucleotide polymorphism) on chromosome 8q22.1 reached genome-wide significance in 2748 migraine patients and 10747 population-matched controls. The association was replicated in a further 3202 cases and 40062 population-matched controls. Expression quantitative trait (eQTL) analysis revealed the SNP to be a regulator of astrocyte elevated gene 1 (*AEG-1*). To identify further susceptibility loci for migraine, we carried out a GWAS of imputed SNPs using as reference 1000 Genomes project data (December 2010 release). Testing more than 11000000 SNPs in 5403 migraine patients and 15327 population-matched controls, six loci reached genome-wide significance. In the replication phase, consisting of 3268 cases and 2916 controls, three loci reached the Bonferroni corrected replication threshold. Of these, two loci had been previously identified (*TRPM8* and *LRP1*) and one was a newly identified locus (*C7orf10*). Whole exome sequencing is potentially an effective tool to identify coding variants underlying human diseases. We designed an extended set of baits (GENECODE exome) for capturing the entire human exome. The extended set allowed the coverage of additional 5594 genes and 10.3 Mb compared to the available CCDS-based sets. In order to identify rare variants contributing to migraine, whole-exome sequencing of 88 cases from 44 families with familial hemiplegic migraine (FHM) was performed. On average, we called 22169 variants per exome and we found 31 shared rare functional variants per family. In one family (family 1), we identified a missense variant in *CACNA1A* (rs121908212), which had been previously described as causing FHM. In another family (family 2), we detected a splice-site variant in *EAAT1*. Mutations in this gene had been previously found in a form of episodic ataxia associated with migraine and alternating hemiplegia (EA6). The functional impact of the identified splice-site *EAAT1* variant has still to be verified.

Acknowledgements

I wish to thank Prof. Aarno Palotie for supervising my work and for having given me the possibility to grow scientifically and personally. During the time spent at the Sanger, he was for me like a second father, noticing and supporting me in difficult moments. Moreover, he has been for me a good example of group leader, from whom I learnt how to work in multi-center collaborative projects.

I would like to thank Prof. John Todd, whose support has been essential to progress in my PhD. Prof John Todd has been closely keeping track of progress and he has been very supportive in moments in which things needed to be moved forward.

I need to thank immensely Prof. David Lomas, who gave me the opportunity to do this PhD and who has been always keeping track of my progress.
Dr. Jeffery Barrett, Dr Eli Zeggini, Dr Ines Barroso and Dr Carl Anderson for their suggestions. They have always been available when I needed suggestions and discuss my projects. A particular thank to Eija Hamalainen, whose support was essential. She has not only been great in organizing and performing the Sequenom follow-up for the GWAS projects, but also she has been giving me encouragement throughout this study. She is also the one who I have to thank for having pushed me to cycle, given that since I started cycling in Cambridge my life changed com-

pletely. A big thank to Dr Johannes Kettunen, Dr. Kati Kristiansson and Dr. Kate Morley for their advice; to Verneri Anttila for the contribution to the quality checks and data analysis during the initial GWAS.

Most of this project was a multi-center collaboration. I would like to thank Emmanouli Dermitzakis' group for performing the expression study, Carol Scott for the exome variants calling and for the interesting discussions, Alison Coffey and Felix Kokocinski for the great work done during the development of the GENECODE exome, Mikko Muona for performing the transcription experiments and all the collaborators for collecting the study samples. Finally, I would like to thank everyone in team 128, team 'awesome', and at the Sanger Institute for all the help they gave me.

Contents

1 Introduction **2**

1.1 The Human Genome 2

1.1.1 What is a genome? 2

1.1.2 Human genetic variations 2

1.1.3 The Human Genome Project 5

1.1.4 The ENCODE Project 5

1.1.5 The HapMap Project 6

1.1.6 The 1000 Genomes Project 8

1.2 Investigating the role of genetic in complex diseases 9

1.2.1 Recurrence risk ratio and heritability 9

1.2.2 Identification of causative genes 10

1.3 Migraine 18

1.3.1 Clinical features 18

1.3.2 Classification and diagnosis 20

1.3.3 Epidemiology 20

1.3.4 Comorbid disorders 23

1.3.5 Pathogenesis 25

1.3.6 Genetic basis 26
1.4 This thesis 31

2 Material and methods **37**

2.1 Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1 37
2.1.1 Study sample 37
2.1.2 **Genotyping** 42
2.1.3 **Quality control** 42
2.1.4 **Statistical analysis** 43
2.1.5 **Imputation** 44
2.1.6 **eQTL analysis** 44
2.1.7 **URLs** 45
2.2 Imputation of sequence variants to identify susceptibility loci for migraine 46
2.2.1 Study samples 46
2.2.2 Genotyping 46
2.2.3 Quality control 48
2.2.4 Imputation 49
2.2.5 Post imputation quality control 49
2.2.6 Statistical analysis 50
2.2.7 URLs 51
2.3 The GENCODE exome: sequencing the complete human exome . . 51
2.3.1 Bait design 51
2.3.2 Samples 52

2.3.3 Sequnce capture and sequencing . . . 53
2.3.4 Sequence alignment and variant calling . . . 54
2.4 Exome sequencing in Familial Hemiplegic migraine . . . 55
2.4.1 Familial Hemiplegic Migraine samples . . . 55
2.4.2 Control exomes . . . 55
2.4.3 Exome library construction . . . 56
2.4.4 Library capture and sequencing . . . 56
2.4.5 Exome data analysis . . . 57
2.4.6 Family 2 variants validation . . . 57
2.4.7 Family 2 complementary DNA (cDNA) analysis . . . 58
2.4.8 Family 3 linkage analysis . . . 58

3 Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1 59

3.1 Introduction . . . 59
3.2 Results . . . 61
3.2.1 Discovery stage . . . 61
3.3 Conditional and haplotype analysis . . . 65
3.4 Replication stage . . . 68
3.5 eQTL analysis . . . 69
3.6 Discussion . . . 73

4 Imputation of SNPs to identify susceptibility loci for migraine 77

4.1 Introduction . . . 77
4.2 Results . . . 79
4.2.1 Initial imputation run . . . 79

4.2.2 Discovery stage 81
4.3 Replication stage 84
4.3.1 Discussion 86

5 GENCODE exome **91**
5.1 Introduction 91
5.2 Results 93
5.2.1 The GENCODE exome features 93
5.2.2 The GENCODE exome performance 97
5.3 Discussion 103
5.4 Notes 106

6 Exome sequencing in Familial Hemiplegic Migraine **107**
6.1 Introduction 107
6.2 Results 109
6.2.1 Whole exome capture of 88 FHM cases 109
6.2.2 Potentially pathogenic variants underlying FHM 111
6.2.3 Family 1: known causal *CACNA1A* mutation 113
6.2.4 Family 2: *EAAT1* mutation 117
6.2.5 Family 3: Integration of linkage analysis and whole exome sequencing 120
6.3 Discussion 121

7 Concluding remarks **126**

List of Figures

1.1 Types of human genetic variants 3

1.2 Organization of the ENCODE project 7

1.3 Comparison of linkage with association analysis for detecting genetic effects 12

1.4 Published GWAS 14

1.5 Published GWAS 15

1.6 Genetic variants frequency and diseases susceptibility 16

1.7 Strategies for identifying rare variants 17

1.8 Migraine attack phases. 19

1.9 Prevalence of migraine in adults of different countries. 23

1.10 Prevalence of different headaches in different age categories. 24

1.11 Anatomical structures involved in migraine pathophysiology. 26

1.12 Pathophysiological mechanisms in the generation of migraine headache. 27

1.13 Roles of proteins encoded by genes involved in FHM at a CNS glutamatergic synapse 30

3.1 Genome-wide P-values for the discovery phase 62

3.2 Quantile-quantile plots of the GWAS discovery phase. 63

3.3 Association signals and recombination rates for the chromosome 8q22.1 locus 65
3.4 Linkage disequilibrium between pair of SNPs at the chromosome 8q22.1 locus 66
3.5 Box-plot of the expression values for AEG1/MTDH based on the rs1835740 genotype 72

4.1 Genotype imputation 78
4.2 Genome-wide P-values for the initial imputation run 80
4.3 Quantile-quantile plots of the initial imputation run. 80
4.4 Genome-wide P-values for the discovery phase 82
4.5 Quantile-quantile plots of the GWAS discovery phase. 84
4.6 Locus-specific association plot: chromosome 2q37.1 87
4.7 Locus-specific association plot: chromosome 12q13.3 88
4.8 Locus-specific association plot: chromosome 7p14.1 90

5.1 Comparison of exon and transcript coverage of the available CCDS-based exome sets and the GENCODE exome set with three current reference gene sets 94
5.2 Cumulative distribution of base coverage for HapMap samples . . . 100
5.3 Cumulative distribution of base coverage for the clinical samples . . 101
5.4 Coverage achieved by the GENECODE based set 105

6.1 Genes with shared rare functional variants in three or more families 113
6.2 Genes with possibly damaging missense, splice site or nonsense variants in two or more families 114

6.3 Pedigree of family 1 115
6.4 *CACNA1A* missense mutations causing hemiplegic migraine (HM) . 116
6.5 Pedigree of family 2 117
6.6 The splicing code 119
6.7 Pedigree of family 3 120

List of Tables

1.1 Estimates of heritability 10

1.2 Migraine classification 21

1.3 Migraine diagnostic criteria 22

1.4 FHM diagnostic criteria 28

1.5 Ion transportation genes and familial hemiplegic migraine 34

1.6 Migraine linkage studies 35

1.7 Migraine candidate-gene association studies 36

2.1 Study samples and gentyping platforms 41

2.2 Quality control 43

2.3 Study samples and gentyping platforms 47

2.4 Quality control: Illumina arrays 48

2.5 Quality control: Sequenom genotyping 49

3.1 Summary results of the GWAS discovery phase for rs1835740 64

3.2 Conditional analysis 67

3.3 Haplotype analysis 68

3.4 Summary results of the GWAS replication phase for rs1835740 . . . 70

3.5 SNPs correlated with AEG-1expression levels 71

3.6 Correlation between rs1835740 and gene expression levels 72
3.7 Summary results of the GWAS discovery phase 76

4.1 Summary results of the discovery stage 83
4.2 Summary results of the replication stage 85
4.3 Summary results of the discovery and replication stages 86

5.1 Comparison of the three different exome capture sets 93
5.2 Exon and transcript coverage of the three different exome caputure sets 95
5.3 Comparison of the three different exome caputure sets with the GENECODE design target 96
5.4 Assessment of repeat and low-complexity coverage of the three exome sets. 96
5.5 Bait/probe covered CTR (Capture Target Regions) regions assessed using a uniqueness mask 98
5.6 Mapping statistics for clinical and HapMap samples using GENCODE and Agilent CCDS exome captures. 99
5.7 Variant calling statistics for clinical and HapMap samples using GENCODE and Agilent CCDS exome captures. 102

6.1 Mapping statistics. 110
6.2 Variant calling statistics 112
6.3 Rare functional variants shared by the two cases of each family . . . 112
6.4 Regions with a LOD score greater than zero. 122
6.5 Rare functional variants shared by the two cases of family 2 123

List of abbreviations

ASP	affected sib pair
ATP1A2	ATPase, Na+/K+ transporting, alpha 2 polypeptide
CACNA1A	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
CDH	Chronic daily headache
ChIP	Chromatin immunoprecipitation
ChIP-Seq	chromatin immunoprecipitation followed by sequencing
CMH	Cochran-Mantel-Haenszel
CNP	Copy number polymorphism
CNS	Central nervous system
CNV	Copy number variant
CSD	Cortical spreading depression
CVD	Cardiovascular disease
DBH	deCODE Migraine Questionnaire

DMQ Dopamine beta–hydroxylase

DOE Department of Energy

DRD2 Dopamine receptor D2

EA Episodic ataxia

EAAT1 Excitatory amino acid transporter 1 gene

ECRs Expressed cluster regions

ELSI Ethical, legal and social issue

ENCODE Encyclopedia of DNA Elements

ESR1 Estrogen receptor 1

FHM Familial hemiplegic migraine

GRR Genotype relative risk

GWAS Genome-wide association study

HBCS Helsinki birth cohort study

HM Hemiplegic migraine

ICHD International Classification of Headache Disorders

IHS International Headache Society

INSR Insulin receptor

LCLs Lymphoblastoid cell lines

MA	Migraine with aura
MA/MO	Migraine with and without aura
MAF	Minor allele frequency
MDS	Multidimensional scaling
MO	Migraine without aura
MTHFR	Methylentetrahydrofolate reductase
NGS	Next-generation sequencing
NHGRI	National Human Genome Research Institute
PFO	Patent foramen ovale
PBMCs	Peripheral blood mononuclear cells
PGR	Progesterone receptor
SCA	spinocerebellar ataxia
SCN1A	Sodium channel, voltage-gated, type I, alpha subunit
SHM	Sporadic hemiplegic migraine
SLC6A3	Solute carrier family 6 member 3
SNP	Single nucleotide polymorphism
SV	Structural variant
TF	Transcription factor

TG	Trigeminal ganglia
TNC	Trigeminal nucleus caudalis
TSC	The SNP Consortium
TTH	Tension-type headache
VDCC	Voltage-dependent calcium channel

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

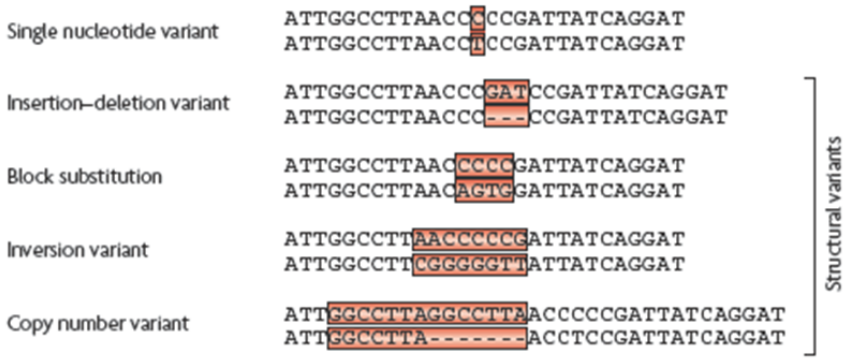

Figure 1.1: **Types of human genetic variants.**Adapted from Frazer et al. (2009) [5].

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-

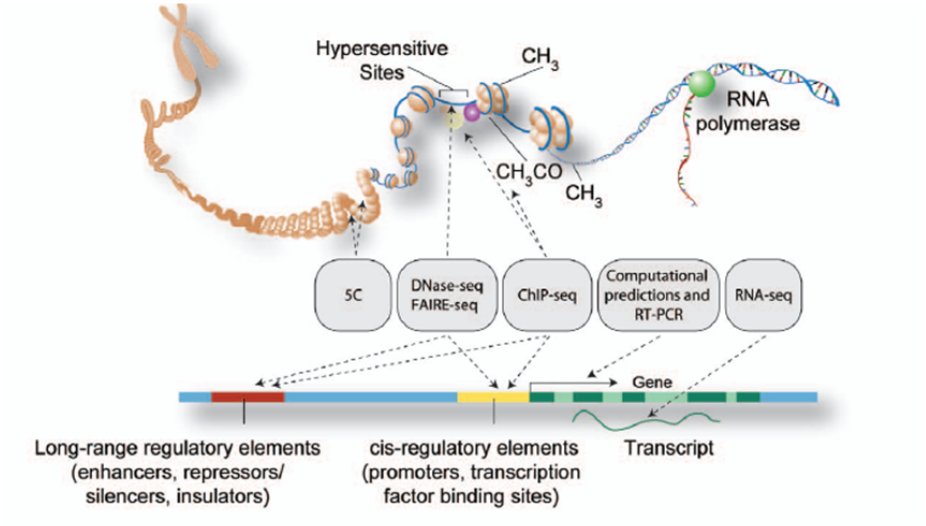

Figure 1.2: **Organization of the ENCODE project**
Representation of the methods used to detect functional elements (gray boxes). Adapted from The ENCODE Project Consortium et al. (2011) [18]

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

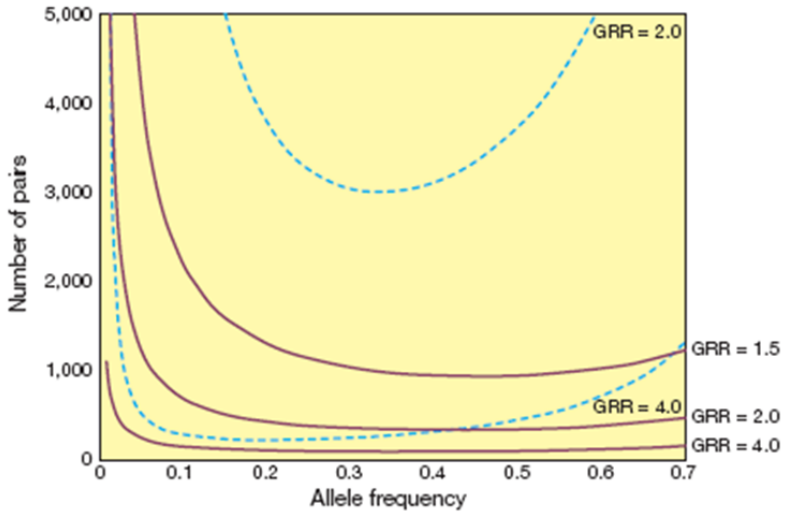

Figure 1.3: **Comparison of linkage with association analysis for detecting genetic effects**
Linkage analysis (dashed lines) is based on affected sib pairs (ASPs) considering a completely linked and informative marker. Association analysis (continuous lines) is based on casecontrol pairs. A multiplicative 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

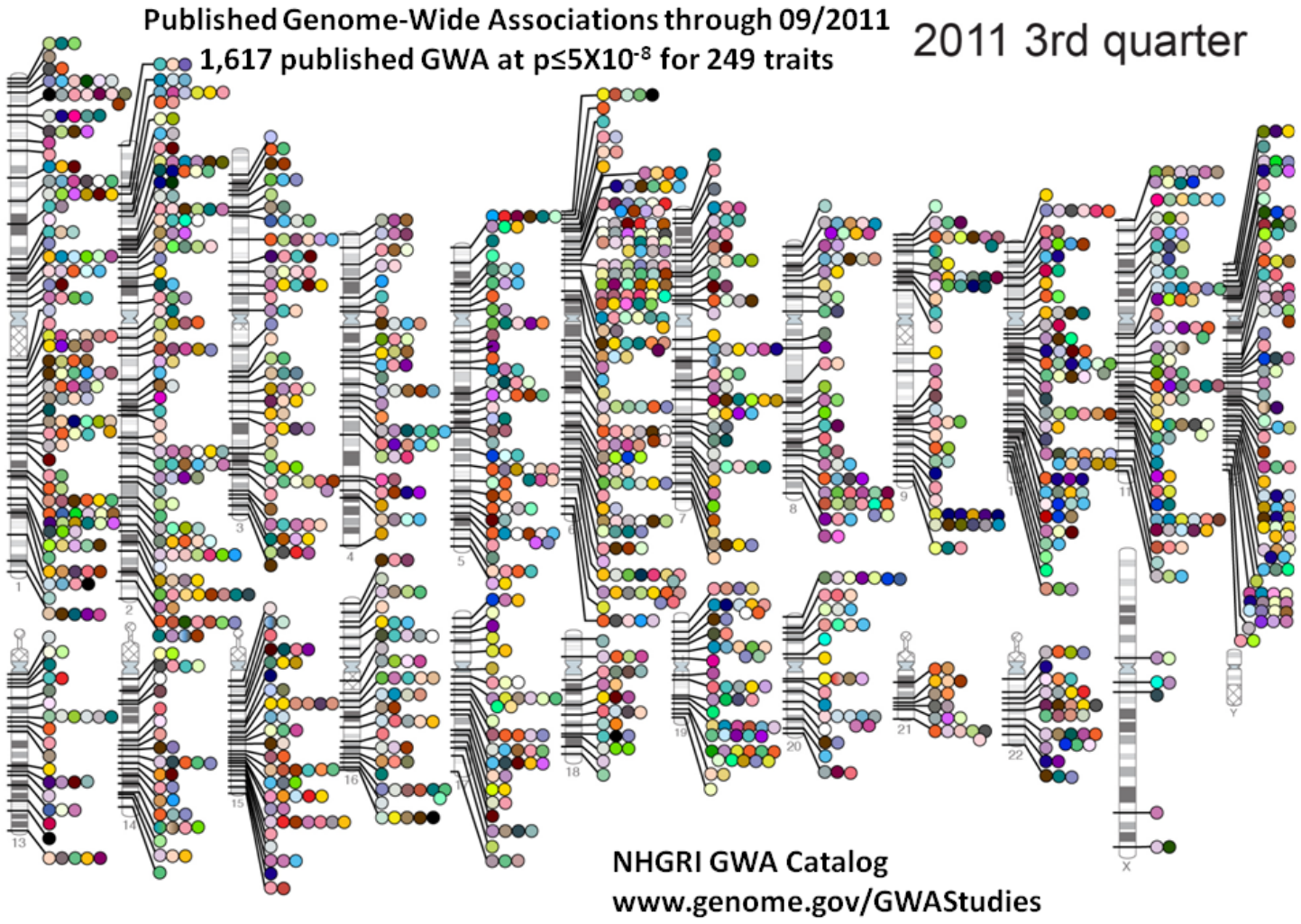

Figure 1.4: **Published GWAS.** Adapted from http://www.genome.gov/gwastudies/.

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

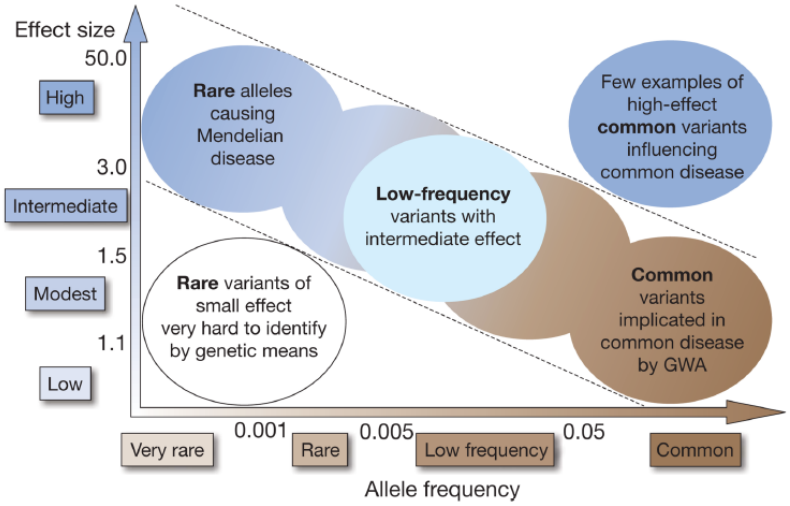

Figure 1.6: **Genetic variants frequency and diseases susceptibility**
GWAS performed so far have mainly identified variants of frequency higher than 5% and modest effect size. Higher penetrance and lower frequency variants could be probably identified in the future thanks to the advances in sequencing 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

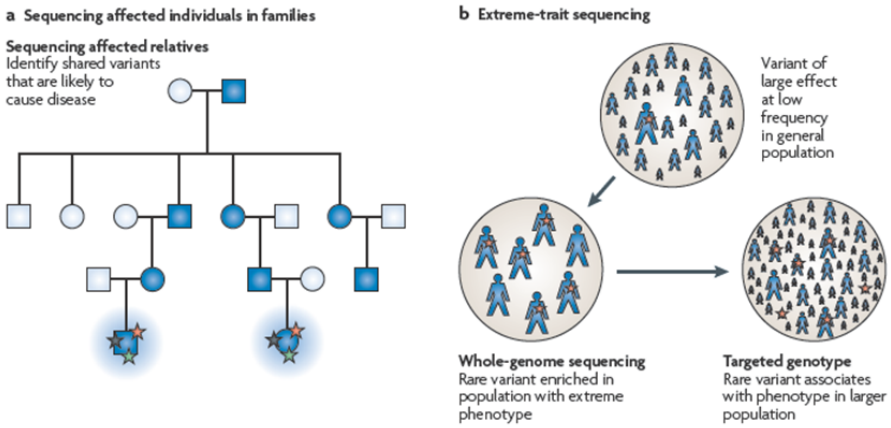

Figure 1.7: **Strategies for identifying rare variants**
In (a) shaded individuals are affected, and sequencing individuals at one or both ends of a trait distribution. In (b) the size of the individual represents the severity of the phenotype. Adapted from Cirulli et al. (2010) [51]

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-

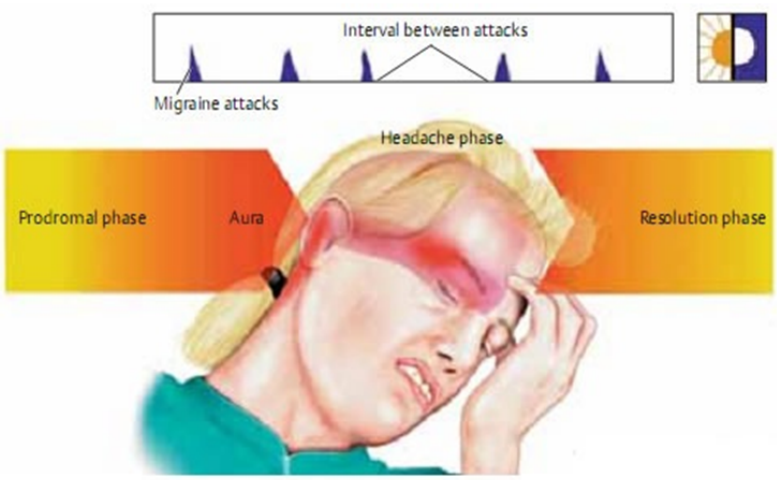

Figure 1.8: **Migraine attack phases.** Adapted from Rohkamm et al (2004)

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

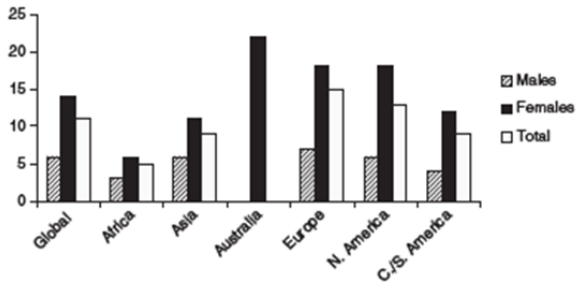

Figure 1.9: **Prevalence of migraine in adults of different countries.**
Adapted from Stovner at al. 2007 [66]

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]

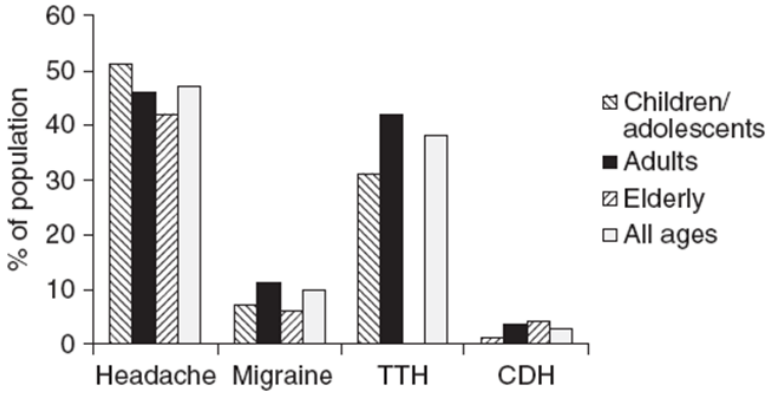

Figure 1.10: **Prevalence of different headaches in different age categories.** TTH, Tension-type headache; CDH, chronic daily headache. Adapted from Stovner at al. 2007 [66]

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

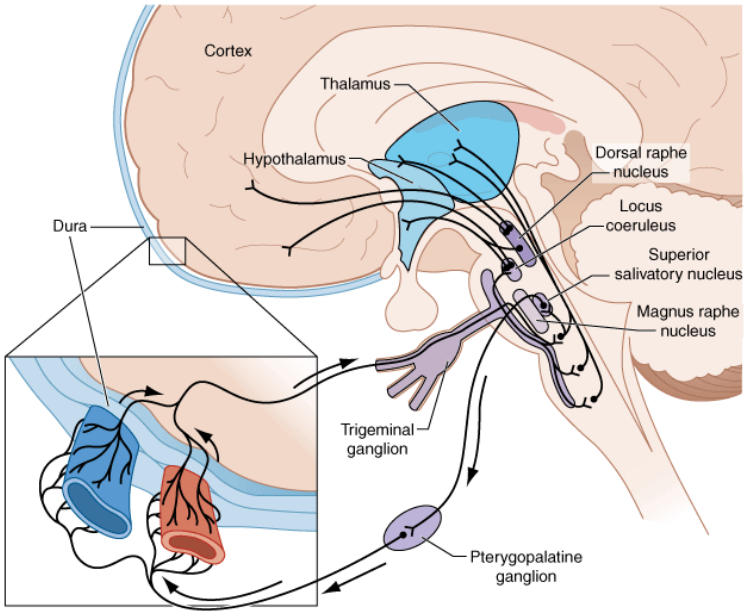

Figure 1.11: **Anatomical structures involved in migraine pathophysiology.** Adapted from Fauci at al. 2008

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,

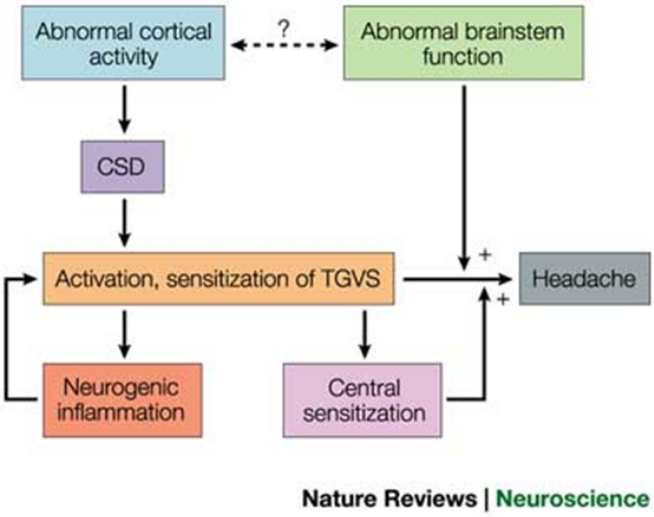

Figure 1.12: **Pathophysiological mechanisms in the generation of migraine headache.** Adapted from Pietrobon at al. 2003 [87]

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-

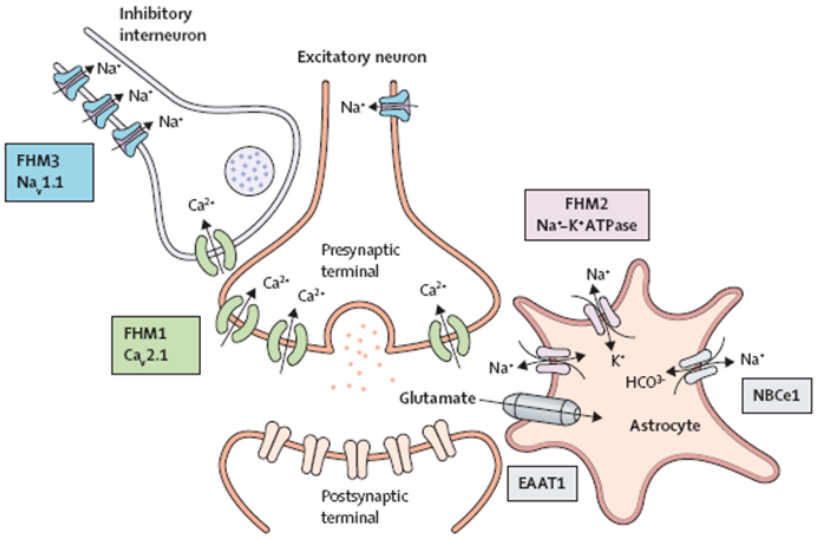

Figure 1.13: **Roles of proteins encoded by genes involved in FHM at a CNS glutamatergic synapse** Adapted from Russell et al. (2011) [126]

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

Chapter 2

Material and methods

2.1 Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1

2.1.1 Study sample

Discovery stage The discovery stage of the GWAS included cases from headache clinics in Finland, Germany and the Netherlands. All the cases were diagnosed as having MA and MA/MO according to the revised criteria of the IHS by experienced neurologists (ICHD-II) (Table 2.1) [58].

In Finland, 1124 Finnish MA and MA/MO patients were recruited. Each one of the recruited cases had at least three family members with migraine. Patients were examined by a neurologist and filled in the validated Finnish Migraine Specific Questionnaire for Family Studies [201] . In cases of insufficient or conflicting

information, a follow-up interview was conducted by telephone. All the diagnosis were made by the same headache specialist (Dr. Mikko Kallela).

In Germany, patient recruitment was done at two sites, in Kiel and in Munich. At the Pain Clinic in Kiel, a total of 994 German MA and MA/MO patients were recruited and the records were maintained at the Universities of Bonn and Cologne. The detailed migraine anamnesis was obtained either by face-to-face or telephone interviews standardized using a comprehensive migraine questionnaire. The second German set of 282 MA and MA/MO cases was recruited at the Klinikum Grohadern of the Ludwig-Maximilians-University, Munich. The patients filled in a German translation of the Finnish Migraine Specific Questionnaire for Family Studies [201] . Whenever the information was insufficient or conflicting, an additional telephone interview was performed.

In the Netherlands, 879 MA and MA/MO patients were available from the clinic-based Leiden University Migraine Neuro Analysis (LUMINA) study. Self-reported migraineurs were recruited via the project's website. A set of screening questions, previously validated in a population-based study, was used to identify patients with migraine [69]. Participants meeting specific screening criteria were asked to complete an extended questionnaire focusing on signs and symptoms of migraine and aura. Diagnoses were made using an algorithm based on the IHS criteria (ICHD-II) [58] and validated by a semi-structured telephone interview performed by either experienced study physicians or well-trained medical students. A subset of the patients was asked to participate upon visiting the outpatient clinic.

Population-matched control samples were obtained from previously genotyped studies (for links to studies, see URL section 2.1.7). Finnish controls originated from the Helsinki Birth Cohort study (1881) and the Health2000 study (2173).

German controls were obtained from the KORA S4/F4 study (840), the HNR study (380) and the PopGen study (677) [202,203]. In addition, 444 controls were obtained from Illumina iControlDB by querying white of European origin samples genotyped on the Illumina 550K platform on June 30th, 2008 and filtering these samples based on stratification as observed from multidimensional scaling plots of all existing German samples, and keeping those identified as being of German descent. Dutch controls were obtained from the Rotterdam study I [204].

Replication stage The replication phase of the study consisted of four independent cohorts from Denmark, Iceland, Germany and the Netherlands (Table 2.1).

In Denmark, 1365 patients, selected from the Danish National Patient Register and case files maintained at neurological clinics, took part in a screening telephone interview. If a proband was diagnosed with MA, according to the ICHD-I criteria, relatives were interviewed by telephone, to identify further migraine cases (Dr. Malene Kirchmann or Dr Anne Hauge). The Danish MA replication sample comprised 825 subjects of which 776 were successfully genotyped. Of these, 483 patients suffered from only MA attacks and 293 from both MA and MO attacks. Danish MO patients (305) were selected from case files at the Danish Headache Center and the MO diagnosis was made according the ICHD-II criteria in an extensive semi-structured telephone interview performed by trained physicians. In addition 81 MO subjects were identified during recruitment of the MA cases. Of the 386 MO patients recruited, 340 were successfully genotyped.

The Icelandic replication samples were obtained from three sources: a list of migraine patients provided by two neurologists (401 participants), responses to an

advertisement in the newsletter of the Icelandic Migraine Society (137 participants) and responses to a brief screening questionnaire mailed to a random sample of 20,000 Icelanders, aged 18-50 years living in the Reykjavik area. All Icelandic participants were asked to answer the comprehensive validated deCODE Migraine Questionnaire (DMQ) 2 or 3, based on the ICHD-II criteria [58]. The reliability of the MA and MO diagnosis, made using on the DMQ, was assessed via a physician-conducted interview. In total, the Icelandic sample consisted of 567 MO patients, and 333 MA patients either with or without MO attacks (MA and MA/MO).

The German replication cohort included 837 MO cases from the Department of Neurology of the Ludwig-Maximilians-University, Munich, Germany. The MO diagnosis was made as described for the MA Munich patient sample.

The Dutch replication sample included 349 Dutch MA and MA/MO patients recently recruited in the clinic-based Leiden University Migraine Neuro Analysis (LUMINA) study. The diagnosis and classification followed the same procedure as in the initial Dutch MA sample.

For each replication cohort, the group providing the cases supplied a matched control dataset. The controls for the Danish and Icelandic replication cohorts were provided by deCODE, the German controls were obtained from the MARS study and GlaxoSmithKline, and the Dutch controls were obtained from the Rotterdam study III.

The study was approved by the ethics committee at all participating institutions. All the study subjects gave written informed consent.

Table 2.1: **Study populations and genotyping platforms**

Country	[a]Cases	[a]Controls	Genotyping platform	Source
Discovery phase				
Finland	[b]1124		Illumina Infinium HumanHap610-Quad	Helsinki University Central Hospital
Finland		1881	Illumina Infinium HumanHap670-Quad	Helsinki Birth Cohort Study (HBCS) [205]
Finland		2173	Illumina Infinium HumanHap610-Quad	Health 2000 study [205]
Germany	[b]994		Illumina Infinium HumanHap610-Quad	University Hospital Cologne
Germany	[b]282		Illumina Infinium HumanHap610-Quad	Ludwig-Maximilians-University, Munich
Germany		840	Illumina Infinium HumanHap550-Duo	KORA study
Germany		380	Illumina Infinium HumanHap550-Duo	Heinz Nixdorf Recall (HNR) study [203]
Germany		677	Illumina Infinium HumanHap550-Duo	PopGen study [206]
Germany		444	Illumina Infinium HumanHap550-Duo	Illumina iControlDB
The Netherlands	[b]879		Illumina Infinium HumanHap610-Quad	Leiden University Migraine Neuro Analysis
The Netherlands		5974	Illumina Infinium HumanHap550-Duo	Rotterdam Study I [204]
Replication phase				
Denmark	[b]825		Centaurus platform	Glostrup Hospital and Danish Headache Center
Denmark	[c]386		Centaurus platform	Glostrup Hospital and Danish Headache Center
Denmark		459	Centaurus platform	deCODE
Denmark		894	Illumina HumanHap650 BeadArray ^{TM}	deCODE
Iceland	[b]333		Illumina Infinium Human317K/370K/610K/1M	deCODE
Iceland	[c]567		Illumina Infinium Human317K/370K/610K/1M	deCODE
Iceland		35221	Illumina Infinium Human317K/370K/610K/1M	deCODE
Germany	[c]837		Illumina Infinium HumanHap610-Quad	Ludwig-Maximilians-University, Munich
Germany		1406	Illumina Infinium HumanHap550-Duo	MARS and Glaxo-SmithKline [207]
The Netherlands	[c]349		Illumina Infinium HumanHap610-Quad	Leiden University Migraine Neuro Analysis
The Netherlands		2082	Illumina Infinium HumanHap550-Duo	Rotterdam Study III

[a]Number of genotyped samples
[b]MA, migraine with aura
[c]MO, migraine without aura

2.1.2 Genotyping

In the discovery stage, cases were genotyped using Illumnina Infinium BeadChips 550K or 610K (Table 2.1) and genotypes were called using the Illuminus algorithm at the Wellcome Trust Sanger Institute [208]. Population-matched controls were drawn from population-based cohorts previously genotyped (Table 2.1).

In the Danish replication cohort, rs1835740 was genotyped in all cases and 459 controls using the Centaurus platform (Nanogen Inc., San Diego, CA, USA). Additional 904 controls were genotyped at deCODE genetics using the Illumina HumanHap650 BeadArray (Table 2.1).

The Icelandic cases and controls were genotyped using the Illumina HumanHap 317K, 370K, 610K or 1M at deCODE genetics.

The Dutch replication cohort was genotyped using the TaqMan technology (Applied Biosystems, Life Technologies) at Leiden University Medical Center.

The German replication cases were genotyped using Illumina HumanHap 610K array at the Institute of Human Genetics at the Helmholtz Zentrum, Munich (Table 2.1).

2.1.3 Quality control

Quality control filtering was performed as detailed in Table 2.2 using PLINK version 1.06 [209]. SNP filtering steps were done separately for each data collection and repeated on the merged data to produce a data set containing only SNPs shared across all the initial data sets, leaving a total of 429912 SNPs for analysis.

Table 2.2: **Quality control**

Sample filtering
Excluded samples:
1. Missing genotype rate >0.03
2. One sample from each pair of related samples
4. Population outliers detected using multidimensional scaling (MDS) plots
5. Samples with inconsistencies between reported gender and genotype-determined gender
6. Outliers for heterozygosity
SNP filtering (it was done after sample filtering)
Excluded SNPs with:
1. Genotype missing rates >0.03
2. Hardy-Weinberg P values $<1 \times 10^{-6}$
3. Minor allele frequencies <0.01

2.1.4 Statistical analysis

In the GWAS discovery phase the Cochran-Mantel-Haenszel (CMH) test implemented was used to test for allelic association on the post quality control data stratified on the basis of the population. Inflation of this statistic was assessed using the genomic control approach. The test inflation factor (λ) was calculated by dividing the median of the test statistics by the expected median from a chi-square distribution with 1 d.f. SNPs with P-values $\leq 5 \times 10^{-8}$ were considered to be genome-wide significant [49]. In the conditional analysis, logistic regression conditioning on each one of the SNPs was used. Significant residual association was defined as P ≤ 0.05. Pairwise r2 values for the same pairs of SNPs were calculated. Haplotype analysis was done using logistic regression. All the analysis of the data in the discovery phase were performed using PLINK version 1.06. To exclude long-range LD for the identified variant, we used the program ssSNPer to demonstrate that no SNP within a 5-Mb window had high LD to rs1835740 in HapMap Phase

II data [210]. Long-range LD analysis was kindly performed by Dr Verneri Anttila.

In the replication case-control cohorts, association between migraine and rs1835740 was analyzed using standard chi-square test. The CMH test was used for the analysis of the combined data of the two stages.

SNPs with P-values $\leq 5 \times 10^{-8}$ in the discovery phase and P-values ≤ 0.05 in the replication phase were considered to have confirmed association to migraine. The statistical analysis was performed in collaboration with Dr Verneri Anttila.

2.1.5 Imputation

For each cohort, imputation of the untyped markers in the two Mb region around rs1835740 was carried out using IMPUTE v2 with the recommended options [211]. Haplotypes from the 1000 Genomes Project (August 2009 release) and HapMap Phase 3 were used as reference panels.

2.1.6 eQTL analysis

This analysis was kindly performed by Emmanouli Dermitzakis group. Obtained data were looked into in order to determine whether rs1835740 acts as an expression quantitative locus (eQTL) [212].

2.1.7 URLs

Control population URLs:

Finland-Health2000 study, http://www.nationalbiobanks.fi

Finland-Helsinki Birth Cohort study, http://www.nationalbiobanks.fi

Germany-KORA S4/F4 study, http://www.helmholtz-muenchen.de/kora

Germany-PopGen study, http://www.popgen.de

Germany-HNR study, http://www.recall-studie.uni-essen.de/recall_info.html

Illumina iControlDB, http://www.illumina.com

The Netherlands-Rotterdam I and III studies, http://www.epib.nl/research/ergo.htm

the Netherlands-Lumina study, http://www.lumc.nl/hoofdpijn

Other URLs:

International Headache Genetics Consortium, http://www.headachegenetics.org

ssSNPer, http://gump.qimr.edu.au/general/daleN/ssSNPer/

GWAS plotter, http://www.broadinstitute.org/node/555

HapMap Phase 2 and 3 data, http://www.hapmap.org

2.2 Imputation of sequence variants to identify susceptibility loci for migraine

2.2.1 Study samples

Discovery stage The discovery stage of the GWAS included cases from headache clinics in Finland, Germany and the Netherlands (Table 2.3).

All the cases were diagnosed as having MA (migraine with aura only), MA/MO (migraine with and without aura) and MO (migraine without aura) according to the revised criteria of the IHS by experienced neurologists (ICHD-II) [58].

Replication phase Finnish, Spanish and Dutch cases were diagnosed as having MO according to the revised criteria (ICHD-II) [58]. For each replication cohort, the group providing the cases supplied a matched control dataset (Table 2.3) The study was approved by the ethics committee at all participating institutions. All the study subjects gave written informed consent.

2.2.2 Genotyping

In the discovery and in the first replication stage, cases were genotyped using Illumnina Infinium BeadChips 550K or 610K (Table 2.3) and genotypes were called using the Illuminus algorithm at the Wellcome Trust Sanger Institute [208]. Population-matched controls were drawn from population-based cohorts previously genotyped (Table 2.3).

Table 2.3: **Study populations and genotyping platforms**

Country	[a]Cases	[a]Controls	Genotyping platform	Source
Discovery phase				
Finland	[b]1124		Illumina Infinium HumanHap610-Quad	Helsinki University Central Hospital
Finland		1881	Illumina Infinium HumanHap670-Quad	Helsinki Birth Cohort Study (HBCS) [205]
Finland		2173	Illumina Infinium HumanHap610-Quad	Health 2000 study [205]
Germany	[b]994		Illumina Infinium HumanHap610-Quad	University Hospital Cologne
Germany	[b]282		Illumina Infinium HumanHap610-Quad	Ludwig-Maximilians-University, Munich
Germany		840	Illumina Infinium HumanHap550-Duo	KORA study
Germany		380	Illumina Infinium HumanHap550-Duo	Heinz Nixdorf Recall (HNR) study [203]
Germany		677	Illumina Infinium HumanHap550-Duo	PopGen study [206]
Germany		444	Illumina Infinium HumanHap550-Duo	Illumina iControlDB
Germany	[c]1208		Illumina Infinium HumanHap610-Quad	Ludwig-Maximilians-University, Munich
Germany		2564	Illumina Infinium HumanHap550-Duo	GSK, MPIPSYKL, KORA
The Netherlands	[b]879		Illumina Infinium HumanHap610-Quad	Leiden University Migraine Neuro Analysis
The Netherlands		5974	Illumina Infinium HumanHap550-Duo	Rotterdam Study I [204]
The Netherlands	[c]1288		Illumina Infinium HumanHap670-Quad	Leiden University Migraine Neuro Analysis
The Netherlands		2157	Illumina Infinium HumanHap550-Duo	Rotterdam Study II
Replication phase				
Finland	[c]875		Sequenom platform	FIMM
Finland		1025	Sequenom platform	Laseri
The Netherlands	[c]1043		Sequenom platform	Leiden University Migraine Neuro Analysis
The Netherlands		910	Sequenom platform	Blood bank
Spain	[c]1350		Sequenom platform	
Spain		981	Sequenom platform	

[a]Number of genotyped samples
[b]MA, migraine with aura
[c]MO, migraine without aura

Table 2.4: **Quality control: Illumina arrays**

Sample filtering
Excluded samples:
1. Missing genotype rate >0.03
2. One sample from each pair of related samples
4. Population outliers detected using multidimensional scaling (MDS) plots
5. Samples with inconsistencies between reported gender and genotype-determined gender
6. Outliers for heterozygosity
SNP filtering (it was done after sample filtering)
Excluded SNPs with:
1. Genotype missing rates >0.03
2. Hardy-Weinberg P values $<1 \times 10^{-6}$
3. Minor allele frequencies <0.01

In the Finnish, Dutch and Spanish replication cohorts, SNPs were genotyped in all cases and controls using the Sequenom platform (Table 2.3).

2.2.3 Quality control

For the data genotyped on the Illumina arrays, quality control filtering was performed as detailed in Table 2.4 using PLINK version 1.06 [209].

For the data genotyped on the Sequenom platform, quality control filtering was performed as detailed in Table 2.5

Table 2.5: **Quality control: Sequenom genotyping**

Sample filtering
Excluded samples:
1. Missing genotype rate >0.1
2. Duplicates
3. Samples with inconsistencies between reported gender and genotype-determined gender

2.2.4 Imputation

For each cohort, imputation of the untyped markers was carried out using IMPUTE2 [211].

In the first imputation run, CEU haplotypes from the 1000 Genomes Project (June 2010 release) and HapMap Phase 3 (February 2009 release) were used as reference panels. In this initial imputatation cases and controls of each data set were imputed separately.

In the second imputation run, 566 EUR haplotypes from the 1000 Genomes Project (December 2010 release) were used as reference panel and in each data set, cases and controls were imputed together.

Reference data were downloaded from the IMPUTE2 website (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html).

2.2.5 Post imputation quality control

SNPs with either an IMPUTE-info score lower than 0.4 or a SNPTEST frequentist-additivite-info score lower than 0.4 were removed from the analysis.

2.2.6 Statistical analysis

In the discovery stage genome wide case-control association analysis was performed using a frequentist association test. In order to take into account for the uncertainty of the imputed genotypes the missing data likelihood score test was used, as implemented in SNPTEST v2 [213].

SNPs with P-values $\leq 5 \times 10^{-8}$ were considered to be genome-wide significant [49]. Meta-analysis results for the case-control analysis were obtained assuming a fixed effect model to combine estimates of the allelic odds ratios and 95% confidence intervals across studies using GWAMA version 2.0.4 [214]. The effective sample size for each data set was used in the meta-analysis.

SNPs from the 29 top regions, containing at least a SNP with a P-values $\leq 5 \times 10^{-6}$ and a typed SNP with a P-values $\leq 1 \times 10^{-4}$, in the discovery stage, were tested for replication in three independent migraine cohorts.

In the replication stage, case-control association was analyzed using logistic regression, as implemented in PLINK version 1.06 [209].

Fixed-effect meta-analysis, as implemented in GWAMA [214], was used to combine the results of the two stages [214]. The effective sample size for each data set was used in the meta-analysis.

SNPs with P-values $\leq 5 \times 10^{-8}$ in the combined discovery and replication stage and P-values $\leq 2.94 \times 10^{-3}$ in the replication stage were considered to have confirmed association to migraine.

2.2.7 URLs

Control population URLs:

Finland-Health2000 study, http://www.nationalbiobanks.fi

Finland-Helsinki Birth Cohort study, http://www.nationalbiobanks.fi

Germany-KORA S4/F4 study, http://www.helmholtz-muenchen.de/kora

Germany-PopGen study, http://www.popgen.de

Germany-HNR study, http://www.recall-studie.uni-essen.de/recall_info.html

Illumina iControlDB, http://www.illumina.com

The Netherlands-Rotterdam I and III studies, http://www.epib.nl/research/ergo.htm

the Netherlands-Lumina study, http://www.lumc.nl/hoofdpijn

Other URLs:

International Headache Genetics Consortium, http://www.headachegenetics.org

GWAS plotter, http://csg.sph.umich.edu/locuszoom/

2.3 The GENCODE exome: sequencing the complete human exome

2.3.1 Bait design

The Genecode exome capture oligonucleotide pool was designed by targeting 288654 exons from 46275 transcripts of 20921 Ensembl protein-coding genes (release 53) and 33621 transcripts of 13772 HAVANA manually annotated protein-coding genes

(database version February 2009), together with additional 1635 miRNA coding genes (Ensembl/miRBase) [215,216]. If the coordinates of the targeted exons overlapped by one or more base pairs, regardless of strand, the overlapping exons were clustered together into expressed cluster regions (ECRs). A 10 bp flank was added on both sides of each ECR. ECRs that overlapped, as result of the added flank, by at least 1 bp were merged. This resulted in a design target of 207108 ECRs, covering 39.3 Mb (35.2 Mb of exonic sequence plus 4.1 Mb of flanking sequence). The coordinates of these ECRs were used for bait design. The baits were designed using the Agilent SureSelect design algorithm using RepeatMasker and WindowMasker defined repeats in order to avoid repetitive regions and to improve uniformity of coverage across the target exons. Three rounds of design were performed. In each successive round of design repeat overlap (0, 20 and 40 bp)was more permissive. After sequencing, the under performing baits were boosted at specific ratios to even out the coverage across all the targets. Depending on the location of baits relative to repeat regions, the boosting was done either by direct replication or by shifting the bait either up or downstream by 30 bp. Doing so, it was possible to design baits targeting 205031 ECRs (99% of the initial GENCODE target). The designed baits covered 47.9 Mb of sequence. Bait design was kindly performed by Felix Kokocinski.

2.3.2 Samples

HapMap samples (NA12878, NA07000, NA19240) were obtained from the Sanger Institute Clinical samples were obtained from Finnish families with a neurological

paroxismal disorder.

2.3.3 Sequnce capture and sequencing

For all the samples, 15 μg of DNA diluted in TE were sheared to 100-400 bp using a Covaris S2 (Covaris, Woburn, MA, USA). The sheared DNA was quantified on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA), and 7.5 μl of COT 1 DNA at 100 ng/ml was added. Each DNA sample was lyophilized in a vacuum concentrator to a pellet and suspended in 3.4 μl of ultrapure water. Following Agilent's SureSelect protocol, 10 μg of sheared DNA were then end repaired, polyA tailed and Illumina-sequencing adapters were ligated to the resulting fragments using the Illumina (San Diego, CA, USA) Paired-End DNA Sample-Prep protocol, except that the gel-size selection step was replaced with a purification using magnetic bead-based solid phase reversible immobilization (SPRI) beads. The capture library was prepared by mixing 5 μl of the oligo capture library, 1.5 μl of ultrapure water and 1 μl of 1:1 dilution of RNase block. Then 500 ng of each sample library was hybridized to the appropriate bait set in PCR plates on a thermocycler at 65°C for 24 h (following the manufacturer's protocol with the modification that no pre-hybridization PCR was performed). The capture was performed according to the manufacturer's protocol with streptavidin-coated Dynal beads (Invitrogen, Paisley, UK), and captured samples were washed three times using SureSelect wash buffers with a series of incubation steps. The samples were cleaned up using Mini Elute columns (Qiagen, Hilden, Germany) and eluted in 50 μl of PCR-grade water. Eluted samples were amplified using a master-mix containing 2 mM $MgCl_2$, 0.2

mM dNTPs, 0.5 μM PE.1, 0.5 μM PE.2 and 3 units of Platinum Pfx DNA Polymerase (Invitrogen) per sample. Samples were aliquoted into three individual wells of a plate and amplified using the following conditions: 94°C for 5 min, followed by 20 cycles of 94°C for 15 s, 58°C for 30 s, 72°C for 30 s and a final extension of 72°C for 5 min. PCR products were purified using SPRI beads before sequencing. For each sample one capture reaction was performed. Captured libraries were sequenced on the Illumina Genome Analyzer 2 platform as paired-end 54-bp reads according to the manufacturer's protocol. Sequence capture and sequencing were kindly performed by Dr Alison Coffey.

2.3.4 Sequence alignment and variant calling

Sequence data were aligned to the human genome (NCBI36) using the MAQ software package v0.7.1 [217]. Base qualities were recalibrated using the Genome Analysis Toolkit v1.0.3540 and duplicate fragments marked using Picard v.1.17. SNPs were called using SAMtools v.0.1.7 and GATK, and the intersection of the resulting calls with a sequence read depth $\geq$ 8x in the target regions were reported. Coverage comparisons of the different target set locations were done using BEDTools v.2.6.0 [218, 219]. Sequence alignment and variant calling was kindly performed by Dr Carol Scott.

2.4 Exome sequencing in Familial Hemiplegic migraine

2.4.1 Familial Hemiplegic Migraine samples

I examined 44 families where there were at least two subjects with a diagnosis of hemiplegic migraine. All the cases were diagnosed as having hemiplegic migraine according to the IHS criteria by experienced neurologists (ICHD-II) [58].
Of the 44 families, forty were obtained from the Department of Neurology at Helsinki University Central Hospital (Helsinki, Finland) and four from Department of Neurology at the Leiden University Medical Center (Leiden, The Netherlands). From each one of the family, two affected subjects were selected for whole exome sequencing.
In families for which DNA of more affected individuals was available, the two individuals who were further apart in the pedigree were chosen for sequencing, in order to minimize the number of shared variants.

The study was approved by the ethics committee at all participating institutions. All the study subjects gave written informed consent.

2.4.2 Control exomes

Shared variants by the two members of each family were screened against variants called in 390 control samples available at the Wellcome Trust Sanger Institute, in order to remove common variants and systematic artifacts.

2.4.3 Exome library construction

Exome libraries were prepared from blood extracted genomic DNA. For each sample, 15 μg of DNA were sheared to 100-400 bp using a Covaris S2 (Covaris, Woburn, MA, USA).

According to Agilent's SureSelect protocol, 10 μg of sheared DNA were then end repaired, polyA tailed and ligated to Illumina-sequencing adapters. Each library was purified using magnetic bead-based solid phase reversible immobilization (SPRI) beads.

2.4.4 Library capture and sequencing

Each library, was hybridized to the bait set (GENCODE exome set) in PCR plates on a thermocycler at 65°C for 24 h. The capture was performed with streptavidin-coated Dynal beads (Invitrogen, Paisley, UK), and captured samples were washed three times using SureSelect wash buffers. The samples were cleaned up using Mini Elute columns (Qiagen, Hilden, Germany) and eluted in 50 μl of PCR-grade water. Eluted samples were amplified using a master-mix containing 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 μM PE.1, 0.5 μM PE.2 and 3 units of Platinum Pfx DNA Polymerase (Invitrogen) per sample. Samples were amplified using the following conditions: 94°C for 5 min, followed by 20 cycles of 94°C for 15 s, 58°C for 30 s, 72°C for 30 s and a final extension of 72°C for 5 min. PCR products were purified using SPRI beads before sequencing. For each sample one capture reaction was performed. Captured libraries were sequenced on the Illumina Genome Analyzer II platform as paired-end 76-bp reads according to the manufacturer's protocol.

Sequence capture and sequencing were kindly performed by Dr Alison Coffey.

2.4.5 Exome data analysis

Sequence data were aligned to the human genome (build37) using the BWA software package v(0.5.6). Base qualities were recalibrated using the Genome Analysis Toolkit v1.0.35 and duplicate fragments marked using Picard v.1.17. Reads with mapping score of zero and potential duplicates were removed. Variants were called using SAMtools v.0.1.7 and GATK, and the intersection of the resulting calls with a sequence read depth $\geq$ 4x in the GENCODE target regions were reported [218]. Variant were filtered out if the read depth was $\leq$ 4x or $\geq$ 1200x, if the consensus quality was $\leq$ 20 or if the base quality was $\leq$ 25. Variants were compared to dbSNP version 132 (via UCSC), 500 exomes (Dec 2010 release) and 1000 Genomes Pilot sites. Variants consequence annotation was added using ensembl version 61. Variants shared by the two members of each family were compared against 390 exomes available at the Wellcome Trust Sanger Institute to remove common variants and systematic artifacts. For non synonymous variant the impact on the function of the encoded protein was predicted using PhdSNP. Sequence alignment and variant calling was kindly performed by Dr Carol Scott.

2.4.6 Family 2 variants validation

Sanger sequencing was used to determine the presence of the identified variants in the other affected members of the family.

2.4.7 Family 2 complementary DNA (cDNA) analysis

BD Vacutainer CPT Cell Preparation Tube with Sodium Citrate was used to separate peripheral blood mononuclear cells (PBMCs). Total RNA from the white blood cells was extracted by using Trizol reagent (Invitrogen) and RNeasy Mini kit (Qiagen). One microgram of RNA was converted to cDNA with iScript cDNA Synthesis Kit (Bio-Rad). To study whether the cDNA population of the mutation carriers contained EAAT1 transcripts incorrectly spliced, PCR primers amplifying cDNA between different EAAT1 exons (exons 2-5, 3-5, 3-6, 3-7, and 2-7) were designed. Gel electrophoresis was performed to separate cDNA molecules based on their size. This analysis was kindly performed in Finland by Mikko Muona.

2.4.8 Family 3 linkage analysis

As part of a previous project, 350 microsatellite markers had been genotyped in 12 members of family 3. Parametric linkage analysis was performed using Merlin software assuming a dominant model.

Chapter 3

Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1

3.1 Introduction

Migraine is a multifactorial disorder and therefore, both environmental and genetic factors contribute to its susceptibility [88]. Genes predisposing to common migraine remain unknown, even though several genomic regions have been implicated in linkage studies [162–172], and several genes have been tested in candidate-gene association studies [173–184]. Linkage studies have had limited success in mapping risk loci in complex diseases, such as migraine [31–33]. Compared to linkage studies, association studies have higher statistical power to detect common variants that confer modest disease susceptibility [187]. Risch (2000) has estimated that for variants with modest genotype relative risks (GRR $\leq$ 2) and intermediate allele

frequency (0.05%-0.50%), linkage analysis are not able to provide any statistical evidence, except in unrealistically large samples. By contrast, case-control association studies, even using a stringent significance level as the one used in GWAS ($P = 5 \times 10^{-8}$), have adequate power to detect variants with GRR as low as 1.5 and with an intermediate allele frequency [34]. Candidate-gene association studies have identified several variants associated to migraine. However, most of the associations have not been replicated. The lack of well-replicated findings could be due to a limited number of cases and controls tested in most of the studies performed so far [173–184]. Genome-wide association studies (GWAS) are association studies in which most of the common variations in the human genome are tested for association with the trait of interest [47]. Compared to candidate-gene association studies, they do not require any assumptions about the causal variants [37]. GWAS are an effective approach to identify common genetic variants which confer disease susceptibility [37, 40, 44–50]. GWAS for migraine have not been performed so far.

To identify common susceptibility variants for migraine, a two stage GWAS was carried out. In the discovery stage, 2748 migraineurs from headache clinics and 10747 population-matched controls from Finland, Germany and the Netherlands were analyzed. In the replication stage, further 3202 cases and 40062 population-matched controls from Iceland, Denmark, The Netherlands and Germany were studied. Owing to the overlap between individuals having migraine with aura (MA) and those having migraine without aura (MO), the following diagnostic subgroups were analyzed: (i) 'all migraine', defined as individuals with migraine irrespective of the subtype; (ii) 'migraine with aura only' (MA), defined as individuals who have only attacks where the aura is present; (iii) 'both migraine with

aura and migraine without aura' (MA/MO), defined as individuals having both attacks with and without aura; (iv) 'migraine without aura only' (MO), defined as individuals with only attacks of migraine without aura.

3.2 Results

3.2.1 Discovery stage

In the discovery stage, more than 500000 SNPs across the genome were genotyped in 3279 European individuals affected by migraine with aura (MA and MA/MO) (1124 Finnish, 1276 Germans, and 879 Dutch), recruited in headache specialized clinics. Diagnoses were made by headache specialists using a combination of questionnaires and individual interviews according to the ICHD-II guidelines [58]. In this stage of the study, we focused our attention on MA, since currently the diagnosis of migraine is based only on clinical features and the presence of aura as diagnostic criteria reduces the possibility to include other headache types among the cases. Population-matched controls (12369) were drawn from population-based cohorts previously genotyped (Helsinki Birth Cohort study, Health2000 study, KORA study, HNR study, PopGen study, Illumina iControlDB and Rotterdam study I)(see Methods).

Study samples had been screened for SNP call rate, presence of population outliers, duplicates and relatedness (see Methods). Overall 2748 cases and 10747 controls passed the quality control filters and remained in the study. After excluding SNPs which did not pass the quality control filters (see Methods), 429912

SNPs were available for analysis.

I performed a Cochran-Mantel-Haenszel (CMH) test (1 degree of freedom) for differences in SNP allele frequencies between cases and controls, stratified by the three different European study samples (Finnish, German and Dutch). The strongest association with migraine, above the threshold for genome-wide significance ($P = 5 \times 10^{-8}$) [49], was found for a locus on chromosome 8q22.1, marked by rs1835740. The minor allele of rs1835740 (A) was significantly associated, with a P-value of 5.38×10^{-9} (OR = 1.23, 95% confidence interval = 1.15-1.32) (Table 3.1, Figure 3.1, 3.2 and 3.3). The pattern of association was similar in all three study cohorts and there was no evidence of heterogeneity of the association (Table 3.1).

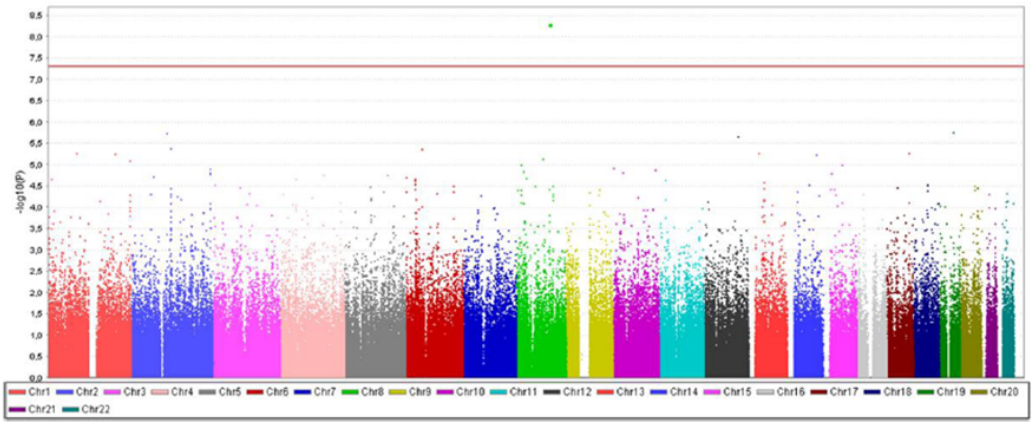

Figure 3.1: **Genome-wide P-values for the discovery phase.** P-values are log transformed (-log10) (y axis) and plotted against chromosomes (x axis). The red line indicates the genome-wide significant threshold. The signal in green above the threshold for genome-wide significance is rs1835740.

Quantile-quantile plot of the distribution of the test statistic for comparison of SNP allele frequencies in cases and controls (1 degree of freedom CMH test) (Figure 3.2) and an estimated genomic inflation factor (λ) of 1.08 suggested a modest overall inflation of P-values.

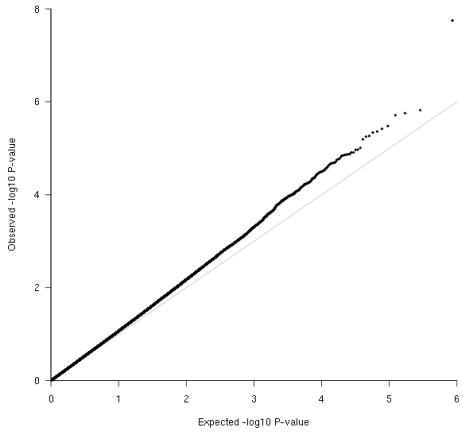

Figure 3.2: **Quantile-quantile plots of the GWAS discovery phase.** Plots of the CMH tests for association between SNP allele frequencies in cases and controls. This shows a slight deviation from the expected distribution (λ= 1.08).

In order to evaluate whether any known untyped marker in the associated region could have explained the signal, the 2-Mb window around rs1835740 was imputed using as reference 1000 Genomes (August 2009 release) and the HapMap3 data. However, no other marker showed evidence of association exceeding the rs1835740 one (Figure 3.3).

Table 3.1: **Summary results of the GWAS discovery phase for the chromosome 8q22.1 locus showing genome-wide significant association with migraine.**

				Finnish			Germans			Dutch			Combined		
Chr	SNP	Position	Alleles minor/major	MAF cases/controls	P	OR	MAF cases/controls	P	OR	MAF cases/controls	P	OR	CMH P	OR	BD P
8	rs2436046	98,232,193	C/T	0.21/0.19	0.0445	1.13	0.22/0.19	0.0005	1.25	0.20/0.18	0.0762	1.14	1.78×10^{-5}	1.18	0.48
8	rs2436047	98,235,049	A/G	0.15/0.16	0.3216	0.93	0.20/0.20	0.8699	0.99	0.22/0.21	0.1793	1.10	0.9529	1.00	0.25
8	rs1835740	98,236,089	A/G	0.26/0.22	0.0005	1.22	0.25/0.22	0.0014	1.20	0.26/0.21	0.0009	1.26	5.38×10^{-9}	1.23	0.79
8	rs982502	98,237,879	T/C	0.24/0.23	0.1085	1.10	0.23/0.20	0.0007	1.24	0.21/0.20	0.1172	1.12	1.34×10^{-4}	1.16	0.33
8	rs1155199	98,246,539	G/A	0.36/0.36	0.9068	1.01	0.42/0.38	0.0079	1.15	0.41/0.39	0.1162	1.10	0.0137	1.08	0.18

Position from human NCBI build 36. MAF, minor allele frequency. OR, Odds ratio. CMH , Cochran-Mantel-Haenszel.
BD, Breslow-Day test for heterogeneous odds ratios between strata.

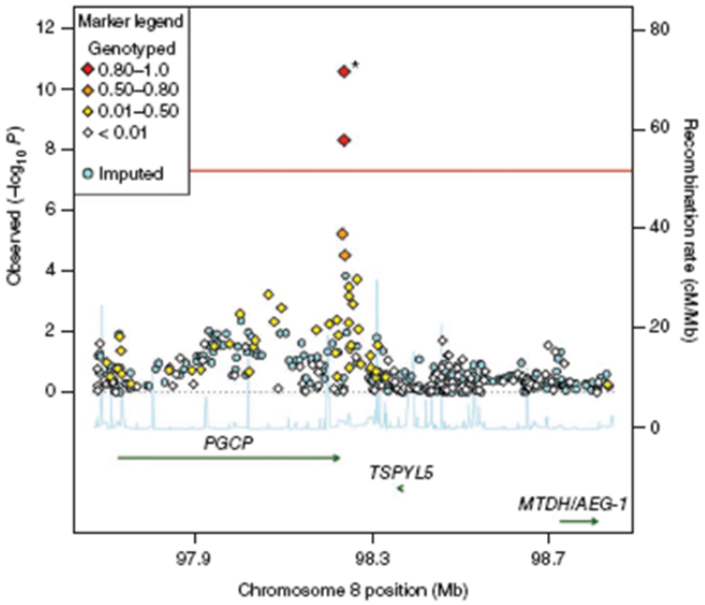

Figure 3.3: **Association signals and recombination rates for the chromosome 8q22.1 locus showing genome-wide significant association with migraine.** The RED line indicates the genome-wide significant threshold ($P = 5{\times}10^{-8}$). Diamonds represent genotyped markers in the region, with colors indicating the extent of linkage disequilibrium (measured in r^2) with the marker rs1835740. Blue circles indicate imputed markers.The signals in red above the threshold for genome-wide significance correspond to rs1835740. For which, P values are shown for both the discovery phase and the meta-analysis of the discovery and replication studies (denoted by asterisk). The blue line shows the recombination rate based on HapMap Phase II data [11] . This figure was generated using the script available at http://www.broadinstitute.org/node/555 modified by Pablo Marin-Garcia and me.

3.3 Conditional and haplotype analysis

According to the HapMap2 recombination maps [11], rs1835740 is located in a region of chromosome 8 between two close recombination hotspots, and an analysis conducted using the ssSNPer program demonstrated that no long range LD to rs1835740 exists within a 5 Mb window [210]. Among the genotyped SNPs in this region, two (rs2436046 and rs982502) showed a P–value $\leq 10^{-3}$ (Table 3.2 and figure 3.4). Based on our data rs2436046 ($r^2 = 0.69$) and rs982502 ($r^2 = 0.59$) are in moderate LD with rs1835740.

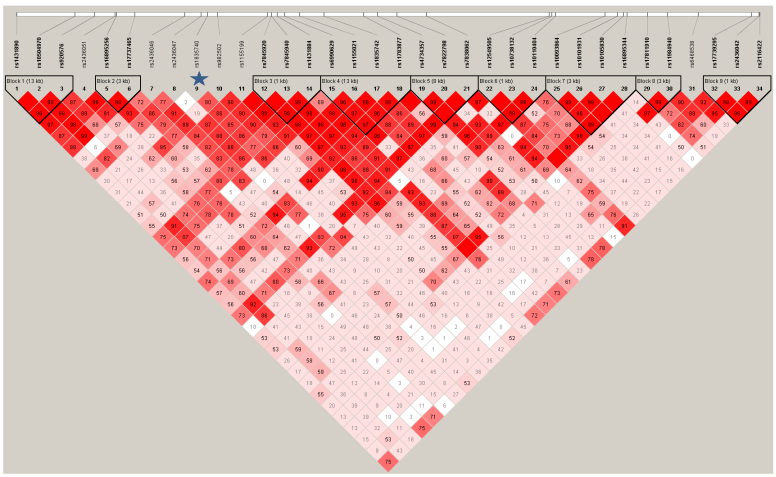

Figure 3.4: **Linkage disequilibrium between pair of SNPs at the chromosome 8q22.1 locus in all subjects included in the discovery phase of the study.** SNPs shown are those which have passed the quality control filters. D′ plot have been produced using Haploview, darker red indicates higher D′ [220]. The most strongly associated SNP (rs1835740) is number 9. The association signal (rs1835740) resides in an intergenic region between two LD blocks (block 2 and 3), defined according to the definition given by Gabriel SB et al. (2002) [221].

To evaluate whether or not the further signals detected in the region were independent from the top SNP association signal, I performed a conditional analysis. The association between SNPs and migraine was tested using the logistic regression with and without conditioning on each one of the SNPs in the associated region with $P \leq 10^{-3}$ (rs2436046, rs1835740 and rs982502). Significant residual association signal was defined as $P \leq 0.05$. Conditioning on rs1835740, no evidence for additional independent signals was found neither for rs2436046 or rs982502 ($P = 0.89$ and $P = 0.47$, respectively) (Table 3.2), suggesting that the moderate association of rs2436046 and rs982502 observed in the CMH test is probably the result of being in LD with rs1835740.

Table 3.2: **Conditional analysis for pair of SNPs.** P-values were calculated using the logistic regression and adjusting for study site.

Chr	SNP A	SNP B	r^2	SNP A P	SNP B P	SNP A given SNP B	SNP B given SNP A
8	rs2436046	rs1835740	0.69	1.78×10^{-5}	5.38×10^{-9}	0.89	1.41×10^{-4}
8	rs982502	rs1835740	0.59	1.34×10^{-4}	5.38×10^{-9}	0.47	4.93×10^{-6}

To determine whether a haplotype-specific effect was present, I inferred haplotypes composed of the five genotyped SNPs in the intergenic region in which rs1835740 resides and tested the association of the inferred haplotypes with migraine before and after conditioning on rs1835740 using the logistic regression. Among the seven haplotypes with frequency higher than 1%, inferred in the region, two showed a modest association signal (haplotype1: $P = 5.72 \times 10^{-5}$, OR=1.18; haplotype7: P = 0.000206, OR = 0.89), which disappeared after conditioning on rs1835740 (haplotype1: P = 0.387, OR = 0.94; haplotype7: P = 0.844, OR = 0.99) (Table 3.3), suggesting that rs1835740 is driving the haplotype association signal.

Table 3.3: **Haplotype frequencies and haplotype-based association test results before and after conditioning for rs1835740**

Haplotype	SNPs					Frequency	OR	P	OR	P
	rs2436046	rs2436047	rs1835740	rs982502	rs1155199				conditioning for rs1835740	
1	C	G	**A**	T	G	0.17	1.18	5.72×10^{-5}	0.94	0.39
2	T	A	G	T	G	0.03	0.88	0.19	0.93	0.45
3	T	A	**A**	C	G	0.04	1.19	0.03	1.01	0.89
4	T	A	G	C	G	0.12	0.98	0.65	1.03	0.64
5	T	G	G	C	G	0.04	0.91	0.27	0.95	0.56
6	C	G	**A**	T	A	0.01	1.24	0.14	1.06	0.72
7	T	G	G	C	A	0.59	0.89	0.0002	0.99	0.84

Marked in bold is the allele more frequently observed in affected individuals compared to controls for the top hit SNP.
Rs2436046, rs2436047, rs1835740, rs982502, rs1155199 are the SNPs forming the haplotypes.
OR, Odds ratio.
P-values were calculated using the logistic regression and adjusting for study site.

3.4 Replication stage

In order to confirm the initial finding, in the replication stage, rs1835740 was analyzed in a further 3202 migraine cases and 40062 population-matched controls from four independent cohorts (Danish, Icelandic, German and Dutch).

In each replication data set the minor allele (A) of rs1835740 was associated with an increased risk of migraine (OR = 1.12–1.23) (Table 3.4). In the overall analysis of the replication stage combined data using the CMH test, the association of rs1853740 was confirmed with a combined P-value of 5.75×10^{-4} (OR = 1.14, 95% confidence interval = 1.06–1.22).

Combining the data of the two stages of the study, convincing evidence for association of rs1835740 with migraine susceptibility was observed, with a combined

P of 1.69×10^{-11} (OR = 1.18, 95% confidence interval = 1.13-1.24). The effect size was similar in the two stages and there was no heterogeneity in the OR estimates among different cohorts in any of the stages (Table 3.4).

3.5 eQTL analysis

Non-coding DNA variants have been shown to modulate gene expression and transcript levels have been correlated to complex traits [222,223]. To evaluate a possible regulatory role of rs1835740 variants on gene expression, Emmanouli Dermitzas' group investigated the effect of this marker genotype on the expression of genes within a 2-Mb window, using a whole-genome gene expression profiling and association analysis with SNPs in three different cell types: fibroblasts, lymphoblastoid cell lines (LCLs) and primary T-cells obtained from the umbilical cord of 75 individuals of Western European origin. These subjects were typed using the Illumina 550K SNP array and transcript abundance for 17945 genes was measured using the Illumina WG-6 v3 expression array [212].

Spearman correlation was used to test for association in cis between SNP genotypes and gene expression levels in each cell type. For each transcript 10000 permutations of expression values relative to SNP genotypes were performed.

This analysis revealed that rs1835740 is an expression quantitative locus (eQTL) for astrocyte elevated gene 1 (*AEG-1*) in LCLs (Table 3.5). At the significance permuted threshold of 0.001, among the 394,651 SNPs tested only rs1835740 was found to be an eQTL for *AEG-1*, with the migraine risk genotype (AA) associated

Table 3.4: **Summary results of rs1835740 in the GWAS discovery and replication phase**

Study group		[a]Cases	[a]Controls	MAF Cases	MAF Controls	OR (95% CI)	P	Heterogeneity P
Discovery phase								
Finnish	Migraine	1064	3513	0.258	0.221	1.22 (1.09–1.37)	0.00045	
Germans	Migraine	1029	2317	0.251	0.216	1.22 (1.08–1.38)	0.00142	
Dutch	Migraine	655	4917	0.255	0.212	1.26 (1.10–1.44)	0.00088	
Combined	Migraine	2748	10747	0.255	0.216	[b]1.23 (1.15–1.32)	[b]5.38×10^{-9}	0.79
	MA only	589	10747	0.267	0.216	[b]1.33 (1.16–1.53)	[b]3.07×10^{-5}	
	MA/MO	2159	10747	0.251	0.216	[b]1.21 (1.12–1.30)	[b]2.69×10^{-6}	
Replication phase								
Danish	Migraine	1116	1353	0.232	0.208	1.15 (0.99–1.34)	0.069	
	MA only	483	1353	0.253	0.208	1.29 (1.05–1.58)	0.015	
	MA/MO	293	1353	0.206	0.208	0.99 (0.79–1.26)	0.951	
	MO only	340	1353	0.225	0.208	1.11 (0.99–1.34)	0.333	
Icelandic	Migraine	900	35,221	0.229	0.202	1.18 (1.04–1.33)	0.010	
	MA only	137	35221	0.255	0.202	1.36 (1.02–1.81)	0.038	
	MA/MO	196	35221	0.209	0.202	1.05 (0.81–1.35)	0.726	
	MO only	567	35221	0.230	0.202	1.18 (1.02–1.38)	0.029	
Dutch	Migraine	349	2082	0.238	0.218	1.12 (0.93–1.35)	0.250	
	MA only	212	2082	0.236	0.218	1.11 (0.87–1.40)	0.406	
	MA/MO	137	2082	0.241	0.218	1.14 (0.85–1.51)	0.382	
Germans	MO	837	1,406	0.240	0.224	1.08 (0.93–1.24)	0.321	
Combined	Migraine	3202	40062	0.234	0.204	[c]1.14 (1.06–1.22)	[c]5.75×10^{-4}	0.82
	MA only	832	38656	0.249	0.203	[c]1.24 (1.08–1.42)	0.002	
	MA/MO	626	38656	0.215	0.203	[c]1.05 (0.90–1.21)	0.533	
	MO only	1744	37980	0.232	0.203	[c]1.12 (1.03–1.23)	0.010	
All combined	Migraine	5950	50809	0.243	0.206	[d]1.18 (1.13–1.24)	[d]1.69×10^{-11}	0.64
	MA only	1421	49403	0.256	0.206	[d]1.29 (1.17–1.41)	[d]6.98×10^{-8}	
	MA/MO	2785	49403	0.243	0.206	[d]1.17 (1.09–1.25)	[d]1.09×10^{-5}	
	MO only	1744	37980	0.232	0.203	[d]1.12 (1.03–1.23)	0.010	

MAF, minor allele frequency. OR, Odds ratio. CMH, Cochran-Mantel-Haenszel.
[a]Number of samples after quality control filtering
[b]OR and P value of CMH test used to combine data sets of the discovery phase
[c]OR and P value of CMH used to combine data sets of the replication phase
[d]OR and P value of the CMH used to combine data sets of the discovery and replication phase

with the highest expression levels (Table 3.5 and figure 3.5).

No correlation between AEG-1 transcripts levels and rs1835740 was found in the other two cell types tested.

These findings suggest that the chromosome 8q22.1 migraine associated SNP is the main genetic determinant of *AEG-1* expression in LCLs. Even if it has been show that peripheral blood can be used for brain-related eQTL mapping further studies will be needed to confirm that rs1835740 is the main *AEG-1* eQTL in the nervous system [224].

None of the other transcripts from the region showed a significant association to rs1835740 in our data set (Table 3.6).

Table 3.5: **SNPs correlated with AEG-1 expression levels.** Correlation between AEG-1 expression and SNP genotypes in lymphoblastoid cell lines (LCLs) has been tested using Sperman rank correlation (SRC). 10000 permutations of AEG-1 expression values relative to SNP genotypes were performed. SNPs for which permuted P-values ≤ 0.01 were obtained are reported. Nominal P-values are shown. Marker rs1835740 is the only significant eQTL detected for AEG-1.

SNP	Gene	SNP position	Gene start	Distance	P
rs11783750	AEG-1	98865219	98725583	-139636	0.00187
rs10105830	AEG-1	98307895	98725583	417688	0.00042
rs1835740	AEG-1	98236089	98725583	489494	0.00004[a]
rs7845920	AEG-1	98247132	98725583	478451	0.00147

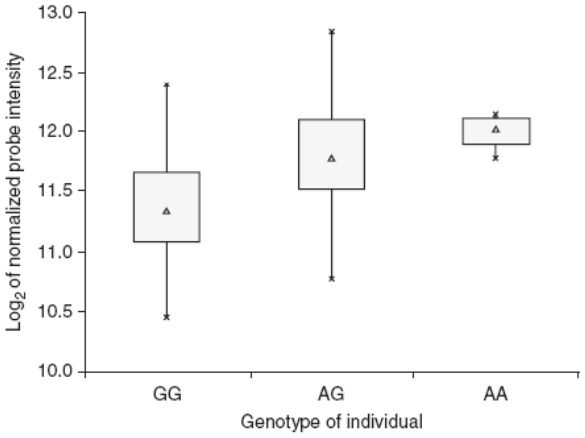

Figure 3.5: **Box-plot of the expression values for AEG1/MTDH based on the rs1835740 genotype** Normalized expression levels of AEG1/MTDH in lymphoblastoid cell lines are shown. In each group, the small pyramid indicates the median, the shaded area represents the lower and upper quartiles, and the crosses show the minimum and maximum values in the expression data

Table 3.6: **Correlation between rs1835740 and gene expression levels.**

SNP	Gene	Strand	SNP position	Gene start	Distance	P
rs1835740	UQCRB	-	98236089	97311911	-924178	0.00132
rs1835740	AEG-1	+	98236089	98725583	489494	0.00004[a]
rs1835740	HRSP12	-	98236089	99183743	947654	0.00287

3.6 Discussion

In the first GWAS for migraine, an associated locus was identified on chromosome 8q22.1. The most significantly associated marker rs1835740 maps to an intergenic region less than 500 kb away from: plasma glutamate carboxypeptidase precursor (*PGCP*), testis-specific Y-encoded-like protein 5 (*TSPY-like protein 5*) and metadherin/astrocyte elevated gene-1 protein (*MTDH/AEG-1*).

Given the available biological data, even though limited, two of these genes (*PGCP* and *MTDH/AEG-1*) seem good candidate genes for migraine. *PGCP* encodes a protein with significant homology to human members of the peptidase family M28, such as glutamate carboxypeptidase II (GCPII) and as other members of the family, it has been shown to have glutamate carboxypeptidase activity. In humans, *PGCP* has been detected principally in blood plasma, where its role has not been identified yet. However, given its observed activity, it has been suggested that PGCP could be involved in the hydrolysis of glutamate from circulating peptides, leading to an increase of plasma glutamate [225]. Several evidences, such as higher concentrations of plasma glutamate in migraine patients compared to controls [226–229] and the induction of migraine through excessive oral intake of glutamate in susceptible individuals [230], suggest a possible role of peripheral glutamate in migraine pathophysiology. These observations make *PGCP* an interesting potential candidate gene for migraine.

MTDH/AEG-1 was originally identified as a HIV-1 and TNF-α inducible gene in primary human foetal astrocytes [231–233]. But for a role in development and progression of different types of cancer, it has been shown that ectopic expression of *MTDH/AEG-1* inhibits excitatory amino acid transporter 2 (*EAAT2/GLT1*) pro-

moter activity [231]. EAATs are trans-membrane proteins involved in the clearance of glutamate from the synaptic cleft. *EAAT2*, expressed by astrocytes, is responsible for most of the glutamate clearance in the brain [234–236]. Several evidences support a role of glutamate in the initiation and propagation of CSD, which is considered to be the most likely pathophysiologic mechanism underlying migraine [82, 86, 237, 238]. MTDH/AEG-1 could increase CSD susceptibility, reducing the glutamate clearance at the synapses via the inhibition of *EAAT2* expression. Using a whole-genome expression profiling assay, we have demonstrated that the migraine associated SNP (rs1835740) is an eQTL for *AEG-1*, with the migraine risk genotype (AA) associated with highest *AEG-1* expression levels. Therefore, it can be hypothesized that the rs1835740 migraine risk variant (A) influences migraine susceptibility increasing *AEG-1* expression and consequently leading to a reduction in *EAAT2* expression, an accumulation of glutamate in the synaptic cleft and finally to an increase in CSD susceptibility. It is worth noting that mutations in the functionally related EAAT1 transporter have been identified in other paroxismal disorders of the nervous system, such as episodic ataxia 6, providing further support for the likely existence of a link between EAAT transporters and episodic disorders of the nervous system [150, 239].

These findings could have important clinical impact, since they support a role of glutamate in migraine pathogenesis and glutamate receptors activity can be inhibited by antagonists. A recent study by Peeters et al. (2007) showed that a glutamate receptor blocker, approved for clinical use, memantine, was able to decrease CSD events induced by potassium chloride in rats [240]. Moreover, pilot clinical studies conducted to evaluate the efficacy of memantine, as preventive therapy in patients with frequent and refractory migraine, obtained promising re-

sults [241,242]. Therefore, memantine could be a potential new agent for migraine preventive treatment.

It is likely that more loci, which confer susceptibility to migraine have to be found. The proportion of genetic variance explained by the rs1835740 variant was estimated to be between 1.5% and 2.5%, depending on the heritability estimate used, and the population attributable risk was estimated to be 10.7% [34]. After exclusion of SNPs at the top associated region, several SNPs showed a $P \leq 10^{-5}$ (Table 3.7). It is likely that some of these will be identified as true associated variants analyzing larger data sets.

In conclusion, in this first GWAS a locus on chromosome 8q22.1 associated with migraine susceptibility was identified and it was shown that the most strongly associated SNP has an allelic-specific effect on *AEG-1* expression. Understanding the mechanisms through which this variant increases migraine susceptibility could improve our knowledge of migraine pathogenesis and could lead to new approaches to treat and prevent the disease. Since the cases enroled in this study were mainly selected from specialized headache clinics, subsequent studies are needed to evaluate the contribution of rs1835740 in population-based migraine cohorts. In these cohorts the migraine spectrum is more heterogeneous and possibly explained by a different combination of genetic susceptibility variants. Even if in the present study, the effect of rs1835740 was found to be larger in individuals with migraine with aura (MA and MA/MO) than in those with migraine without aura (MO), further studies are needed to confirm the role of the variant in different migraine subgroups. Since the identified variant explains only a small fraction of the overall migraine genetic variance, further GWAS, perhaps with different ascertainment schemes, will be needed to identify additional loci.

Table 3.7: **Summary results of the GWAS discovery phase for the regions of the genome showing P-value $\leq 5 \times 10^{-5}$**

Chr	SNP	Position	Alleles minor/major	MAF	P	OR (95% CI)	Location	GENE
1	rs12084862	244269837	A/G	0.27	8.20×10^{-6}	1.17 (1.09–1.25)	intergenic	*SMYD3*
2	rs17528324	118572626	A/G	0.09	4.13×10^{-6}	1.27 (1.15–1.41)	intergenic	*INSIG2*
2	rs17862920	234492734	T/C	0.09	1.26×10^{-5}	0.78 (0.69–0.87)	intergenic	*TRPM8*
6	rs2038761	2625766	A/G	0.31	2.02×10^{-5}	0.87 (0.81–0.93)	intergenic	*MYLK4*
6	rs6456880	29071227	G/T	0.39	2.18×10^{-5}	0.87 (0.82–0.93)	intergenic	*ZNF311*
6	rs7753655	49644523	G/A	0.30	4.29×10^{-6}	0.85 (0.80–0.91)	intergenic	*AL590244.2*
8	rs10888075	13804790	A/C	0.15	1.04×10^{-5}	1.21 (1.11–1.31)	intergenic	*SGCZ*
8	rs10111769	21003036	G/A	0.31	1.49×10^{-5}	1.15 (1.08–1.23)	intergenic	
11	rs2042600	19709275	T/G	0.48	2.28×10^{-5}	0.88 (0.82–0.93)	intergenic	*NAV2*
13	rs3794331	44951545	C/A	0.07	2.7×10^{-5}	1.28 (1.14–1.43)	intergenic	*COG3*
15	rs473422	56453633	A/G	0.35	1.03×10^{-5}	0.86 (0.82–0.92)	intergenic	*AQP9*

For each region the SNP with the strongest signal has been reported. P-values were obtained using the Cochran-Mantel-Haenszel test. Position from human NCBI build 36. MAF, minor allele frequency. OR, Odds ratio. CI, confidence interval.

Chapter 4

Imputation of SNPs to identify susceptibility loci for migraine

4.1 Introduction

To identify common susceptibility variants for migraine, we carried out a GWAS, described in the previous chapter. This study provided evidence of association for a SNP on chromosome 8q22.1 (rs1835740). Expression quantitative trait (eQTL) analysis revealed this SNP to be a key regulator of astrocyte elevated gene 1 *AEG–1* in lymphoblastoid cell lines [243]. A subsequently published population-based GWAS has identified other three risk loci for migraine (chromosome 1p36.23, chromosome 2q37.1, chromosome 12q13.3) [244].

The hundreds of thousands of SNPs directly assayed 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. The

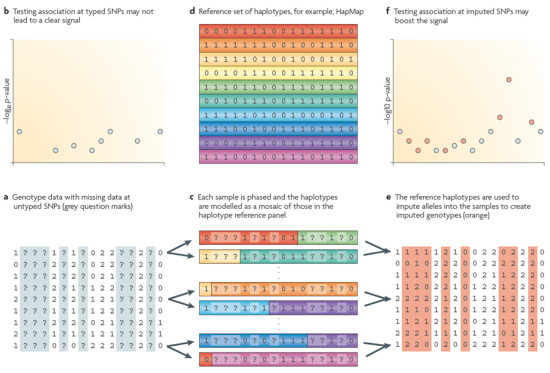

Figure 4.1: **Genotype imputation.** [245]

term imputation means predict genotypes of SNPs, which have not been directly assayed, in a sample using a reference panel of haplotypes including a much larger number of SNPs (Figure 4.1).

Genotype imputation tools involve phasing the typed SNPs in each individual of the study. Then these haplotypes are compared to the haplotypes of the reference panel and missing genotypes are predicted after matching haplotypes with the reference ones. A probability distribution over the possible genotypes is produced for each one of the imputed genotypes [245]. It has been shown that imputation error rate decreases as the minor allele frequency and the size of the reference panel increase [211, 245].

In order to identify novel risk loci for migraine, I have imputed untyped SNPs in migraine cases and population-matched controls from Finland, Germany and the Netherlands, using the 566 haplotypes of 1000 Genomes project (December 2010 release) as reference. The results obtained from the imputed data were replicated in independent migraine case and population-matched controls from Finland, the Netherlands and Spain.

4.2 Results

4.2.1 Initial imputation run

In an initial imputation run, 3279 European individuals affected by migraine with aura (MA) only or by migraine with and without aura (MA/MO) (1124 Finnish, 1276 Germans, and 879 Dutch) and 12369 population-matched controls (Helsinki Birth Cohort study, Health2000 study, KORA study, HNR study, PopGen study, Illumina iControlDB and Rotterdam study I) were included (see Methods).

Study samples had been screened for SNP call rate, presence of population outliers, duplicates and relatedness (see Methods). Overall 2948 cases and 10747 controls passed the quality control filters and remained in the study.

After excluding SNPs which did not pass the quality control filters (see Methods), around 7000000 untyped SNPs were imputed separately in cases and controls of each cohort using the software IMPUTE2 and 1000 Genomes plus HapMap III data as reference [211].

Genotyped and imputed SNPs were tested for association with migraine using a score test, as implemented in SNPTEST v2 [213]. The results of the association tests across the three cohorts (Finnish, German and Dutch) were combined using a fixed effect meta-analysis, as implemented in GWAMA version 2.0.4 [214]. This led to the identification of 62 loci that surpassed the threshold for genome-wide significance ($P = 5 \times 10^{-8}$) (Figure 4.2) [49]. However, quantile-quantile plot of the distribution of the test statistic suggested an overall inflation of P-values (λ= 1.38)(Figure 4.3).

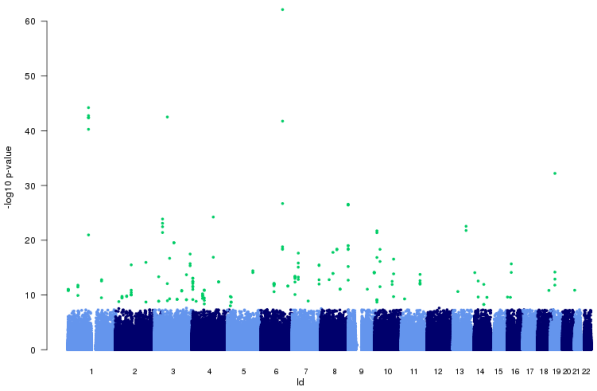

Figure 4.2: **Genome-wide P-values for the initial imputation run.** P-values are log transformed (-log10) (y axis) and plotted against chromosomes (x axis). The signal in green are the ones above the threshold for genome-wide significance.

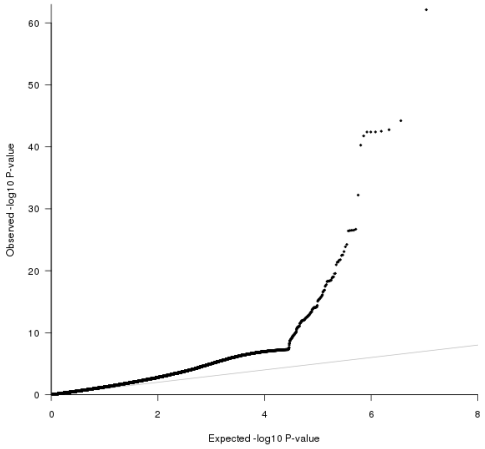

Figure 4.3: **Quantile-quantile plots of the initial imputation run** Plots of the fixed effect meta-analysis results of the initial imputation run.

4.2.2 Discovery stage

Since the number of genome-wide significant loci seemed excessively high, I thought that bias could have been introduced by imputing, in each population, cases and control separately. Therefore, it was decided to repeat the imputation in merged sets of cases and controls for each population. In the meantime a new release of 566 European haplotypes was released by the 1000 Genomes project and, hence, it was decided to use this new set as reference for the new imputation run, since the higher number of reference haplotypes would have improved the imputation accuracy. Moreover, two other migraine data sets became available, including 2490 migraine without aura cases (MO) (1208 German and 1282 Dutch) and 4580 population-matched controls. Therefore, since the two main types of migraine (MA and MO) seem to share a common genetic component, we decided to include them in the discovery stage of our study, to increase the power of detection of migraine risk loci [69].

The discovery stage included 5403 European individuals affected by migraine, of which 2748 were part of our previous GWAS. Diagnoses were made by headache specialists using a combination of questionnaires and individual interviews according to the ICHD-II guidelines [58]. Population-matched controls (15327) were drawn from previously genotyped population-based cohorts previously genotyped (see Methods). Study samples had been screened for SNP call rate, presence of population outliers, duplicates and relatedness (see Methods).

After excluding SNPs which did not pass the quality control filters (see Methods), around 11000000 untyped SNPs were imputed in each cohort using the software IMPUTE2 and 566 European haplotypes from the 1000 Genomes project

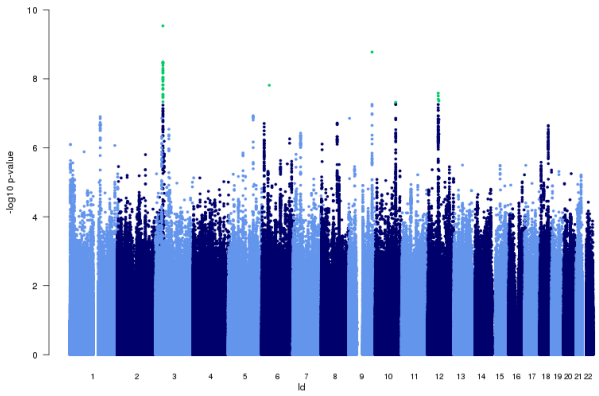

Figure 4.4: **Genome-wide P-values for the discovery phase.** P-values are log transformed (-log10) (y axis) and plotted against chromosomes (x axis). The signal in green are the ones above the threshold for genome-wide significance.

(December 2010 release) as reference [211].

Genotyped and imputed SNPs were tested for association with migraine using a score test, as implemented in SNPTEST v2, to take into account the uncertainty of the imputed genotypes [213]. The results of the association tests across the three cohorts (Finnish, German and Dutch) were combined using a fixed effect meta-analysis, as implemented in GWAMA version 2.0.4 [214].

Six loci that surpassed the threshold for genome-wide significance ($P = 5 \times 10^{-8}$) were identified (Figure 4.4 and Table 4.1) [49]. The genome-wide significant SNPs had the same direction of allelic effect in all the study cohorts. Two were previously identified loci (chromosome 2q37.1 and chromosome 12q13.3) and four were newly identified loci. Quantile-quantile plot of the distribution of the test statistic suggested a modest overall inflation of P-values (λ= 1.09) (Figure 4.5).

Table 4.1: **Summary results of the discovery stage**

				Finnish MA and MA/MO (1064/3513)[a]			Germans MA and MA/MO (1029/2317)[a]			Dutch MA and MA/MO (820/4917)[a]			German MO (1208/2564)[a]			Dutch MO (1282/2016)[a]			Meta-analysis (5403/15327)[a]	
Chr	Position	SNP	Alleles minor/major	MAF cases/controls	OR (95% CI)	P	MAF cases/controls	OR (95% CI)	P	MAF cases/controls	OR (95% CI)	P	MAF cases/controls	OR (95% CI)	P	MAF cases/controls	OR (95% CI)	P	OR (95% CI)	P
1	3319777	rs4471209	A / G	0.36 / 0.34	1.05 (0.95 – 1.16)	3.34×10^{-1}	0.31 / 0.27	1.22 (1.09 – 1.37)	4.10×10^{-4}	0.31 / 0.28	1.13 (1.01 – 1.27)	2.45×10^{-2}	0.30 / 0.27	1.17 (1.05 – 1.30)	2.66×10^{-3}	0.29 / 0.28	1.07 (0.96 – 1.19)	2.16×10^{-1}	1.13 (1.08 – 1.19)	2.93×10^{-6}
1	4185770	rs6679136	T / C	0.33 / 0.34	0.95 (0.86 – 1.06)	3.62×10^{-1}	0.36 / 0.38	0.94 (0.84 – 1.04)	2.05×10^{-1}	0.34 / 0.39	0.79 (0.71 – 0.88)	1.49×10^{-5}	0.34 / 0.38	0.86 (0.78 – 0.95)	1.61×10^{-3}	0.37 / 0.39	0.91 (0.82 – 1.00)	5.27×10^{-2}	0.88 (0.84 – 0.93)	7.98×10^{-7}
1	15532130	rs10927720	T / C	0.20 / 0.19	1.08 (0.95 – 1.22)	2.23×10^{-1}	0.25 / 0.20	1.27 (1.13 – 1.44)	1.27×10^{-4}	0.23 / 0.21	1.15 (1.01 – 1.30)	3.01×10^{-2}	0.24 / 0.21	1.19 (1.06 – 1.34)	2.76×10^{-3}	0.21 / 0.21	1.04 (0.92 – 1.18)	5.01×10^{-1}	1.15 (1.08 – 1.21)	2.75×10^{-6}
1	74118855	rs61767191	C / T	0.06 / 0.04	1.30 (1.05 – 1.62)	8.61×10^{-3}	0.06 / 0.05	1.29 (1.03 – 1.61)	9.11×10^{-3}	0.05 / 0.04	1.16 (0.90 – 1.49)	1.63×10^{-1}	0.07 / 0.05	1.27 (1.04 – 1.55)	4.68×10^{-3}	0.04 / 0.04	1.17 (0.92 – 1.50)	1.43×10^{-1}	1.36 (1.20 – 1.55)	1.31×10^{-6}
1	115677183	rs2078371	C / T	0.15 / 0.13	1.21 (1.06 – 1.39)	5.78×10^{-3}	0.11 / 0.11	0.98 (0.83 – 1.15)	7.66×10^{-1}	0.12 / 0.11	1.12 (0.95 – 1.32)	1.64×10^{-1}	0.12 / 0.11	1.06 (0.91 – 1.23)	4.33×10^{-1}	0.13 / 0.10	1.36 (1.17 – 1.59)	8.60×10^{-5}	1.15 (1.07 – 1.23)	1.54×10^{-4}
1	156456301	rs3790455	C / T	0.43 / 0.42	1.04 (1.15 – 0.94)	4.32×10^{-1}	0.35 / 0.33	1.10 (1.22 – 0.98)	1.01×10^{-1}	0.36 / 0.33	1.11 (1.24 – 1.00)	5.56×10^{-2}	0.38 / 0.33	1.24 (1.37 – 1.12)	3.63×10^{-5}	0.37 / 0.32	1.22 (1.35 – 1.10)	1.72×10^{-4}	1.14 (1.20 – 1.09)	1.26×10^{-7}
1	231361110	rs11590715	T / C	0.07 / 0.06	1.17 (0.97 – 1.42)	7.21×10^{-2}	0.06 / 0.05	1.30 (1.04 – 1.62)	8.55×10^{-3}	0.07 / 0.06	1.21 (0.98 – 1.49)	4.53×10^{-2}	0.07 / 0.05	1.40 (1.15 – 1.71)	1.24×10^{-4}	0.06 / 0.06	1.12 (0.91 – 1.38)	2.26×10^{-1}	1.32 (1.18 – 1.48)	8.58×10^{-7}
2	145222038	rs13403907	G / A	0.18 / 0.15	1.16 (1.02 – 1.32)	1.65×10^{-2}	0.13 / 0.10	1.28 (1.09 – 1.51)	1.04×10^{-3}	0.13 / 0.11	1.17 (1.00 – 1.37)	3.35×10^{-2}	0.12 / 0.11	1.15 (0.99 – 1.34)	4.43×10^{-2}	0.13 / 0.11	1.11 (0.95 – 1.29)	1.59×10^{-1}	1.20 (1.12 – 1.30)	1.58×10^{-6}
2	234821445	rs11892538	C / G	0.11 / 0.14	0.81 (0.70 – 0.94)	3.30×10^{-3}	0.14 / 0.17	0.81 (0.70 – 0.93)	2.86×10^{-3}	0.15 / 0.17	0.85 (0.73 – 0.98)	2.07×10^{-2}	0.14 / 0.17	0.82 (0.71 – 0.94)	2.43×10^{-3}	0.14 / 0.17	0.80 (0.70 – 0.92)	9.92×10^{-4}	0.81 (0.76 – 0.86)	2.92×10^{-10}
2	241447428	rs4676436	A / C	0.14 / 0.13	1.10 (0.96 – 1.27)	1.71×10^{-1}	0.13 / 0.11	1.24 (1.06 – 1.45)	7.48×10^{-3}	0.12 / 0.11	1.18 (1.00 – 1.38)	4.59×10^{-2}	0.13 / 0.11	1.20 (1.04 – 1.40)	1.42×10^{-2}	0.14 / 0.11	1.27 (1.09 – 1.48)	1.62×10^{-3}	1.20 (1.12 – 1.29)	6.46×10^{-7}
3	30480085	rs6790925	T / C	0.41 / 0.39	1.09 (0.99 – 1.21)	7.48×10^{-2}	0.37 / 0.35	1.11 (0.99 – 1.23)	6.20×10^{-2}	0.40 / 0.38	1.11 (1.00 – 1.24)	5.13×10^{-2}	0.38 / 0.35	1.16 (1.05 – 1.29)	2.88×10^{-3}	0.41 / 0.36	1.20 (1.09 – 1.33)	3.36×10^{-4}	1.14 (1.08 – 1.19)	1.39×10^{-7}
3	67144706	rs4311165	C / G	0.17 / 0.20	0.81 (0.71 – 0.92)	1.35×10^{-3}	0.24 / 0.27	0.87 (0.77 – 0.98)	2.04×10^{-2}	0.25 / 0.26	0.93 (0.82 – 1.05)	2.30×10^{-1}	0.24 / 0.27	0.87 (0.77 – 0.97)	1.01×10^{-2}	0.25 / 0.27	0.87 (0.78 – 0.97)	1.28×10^{-2}	0.87 (0.82 – 0.92)	2.89×10^{-7}
5	75495849	rs6453209	T / A	0.46 / 0.50	0.85 (0.93 – 0.77)	8.84×10^{-4}	0.45 / 0.47	0.91 (1.01 – 0.82)	6.49×10^{-2}	0.47 / 0.48	0.96 (1.06 – 0.86)	3.94×10^{-1}	0.48 / 0.49	0.93 (1.03 – 0.85)	1.30×10^{-1}	0.43 / 0.47	0.85 (0.93 – 0.76)	8.61×10^{-4}	0.89 (0.93 – 0.85)	1.41×10^{-6}
5	127722107	rs77050147	C / G	0.10 / 0.09	1.24 (1.05 – 1.45)	1.11×10^{-2}	0.10 / 0.08	1.27 (1.06 – 1.52)	8.43×10^{-3}	0.12 / 0.10	1.24 (1.05 – 1.46)	9.28×10^{-3}	0.09 / 0.08	1.16 (0.98 – 1.39)	8.78×10^{-2}	0.12 / 0.09	1.25 (1.07 – 1.47)	5.37×10^{-3}	1.25 (1.15 – 1.35)	1.18×10^{-7}
6	12908747	rs13197912	T / A	0.30 / 0.29	1.06 (0.95 – 1.17)	3.09×10^{-1}	0.37 / 0.36	1.08 (0.97 – 1.21)	1.44×10^{-1}	0.37 / 0.35	1.09 (0.97 – 1.21)	1.32×10^{-1}	0.40 / 0.35	1.23 (1.12 – 1.36)	3.73×10^{-5}	0.38 / 0.33	1.21 (1.09 – 1.34)	2.42×10^{-4}	1.14 (1.09 – 1.20)	1.98×10^{-7}
6	39177971	rs873690	C / G	0.09 / 0.07	1.26 (1.05 – 1.50)	3.25×10^{-3}	0.09 / 0.08	1.19 (0.99 – 1.44)	1.50×10^{-2}	0.08 / 0.07	1.10 (0.90 – 1.34)	2.18×10^{-1}	0.09 / 0.08	1.20 (1.01 – 1.42)	1.01×10^{-2}	0.09 / 0.07	1.30 (1.09 – 1.56)	1.81×10^{-4}	1.37 (1.23 – 1.52)	1.54×10^{-8}
6	97039665	rs2971609	G / A	0.20 / 0.18	1.15 (1.02 – 1.30)	2.44×10^{-2}	0.26 / 0.24	1.11 (0.98 – 1.25)	8.59×10^{-2}	0.23 / 0.23	1.01 (0.89 – 1.14)	9.24×10^{-1}	0.27 / 0.24	1.19 (1.07 – 1.33)	1.51×10^{-3}	0.27 / 0.23	1.22 (1.08 – 1.36)	7.17×10^{-4}	1.14 (1.08 – 1.21)	2.34×10^{-6}
6	143288832	rs1041655	A / C	0.41 / 0.42	0.93 (0.84 – 1.02)	1.29×10^{-1}	0.37 / 0.41	0.88 (0.79 – 0.98)	1.58×10^{-2}	0.35 / 0.38	0.87 (0.78 – 0.98)	1.79×10^{-2}	0.37 / 0.41	0.86 (0.78 – 0.95)	1.35×10^{-3}	0.36 / 0.39	0.90 (0.81 – 1.00)	3.04×10^{-2}	0.88 (0.84 – 0.93)	5.51×10^{-7}
6	153997563	rs6937294	C / T	0.43 / 0.45	1.08 (1.19 – 0.98)	1.11×10^{-1}	0.40 / 0.43	1.12 (1.24 – 1.01)	3.11×10^{-2}	0.38 / 0.42	1.20 (1.34 – 1.08)	5.28×10^{-4}	0.41 / 0.43	1.07 (1.18 – 0.97)	1.85×10^{-1}	0.37 / 0.40	1.14 (1.26 – 1.03)	9.10×10^{-3}	1.12 (1.18 – 1.07)	2.46×10^{-6}
7	17014115	rs17136988	G / T	0.05 / 0.06	0.84 (0.68 – 1.04)	1.03×10^{-1}	0.08 / 0.10	0.75 (0.62 – 0.91)	1.48×10^{-3}	0.07 / 0.08	0.83 (0.67 – 1.01)	4.84×10^{-2}	0.09 / 0.10	0.87 (0.73 – 1.03)	8.08×10^{-2}	0.07 / 0.09	0.80 (0.66 – 0.96)	8.66×10^{-3}	0.80 (0.73 – 0.88)	1.55×10^{-6}
7	40466200	rs4379368	T / C	0.16 / 0.14	1.17 (1.02 – 1.33)	2.14×10^{-2}	0.12 / 0.11	1.15 (0.97 – 1.35)	9.87×10^{-2}	0.12 / 0.11	1.17 (1.00 – 1.38)	5.24×10^{-2}	0.13 / 0.10	1.29 (1.11 – 1.50)	7.58×10^{-4}	0.13 / 0.11	1.20 (1.03 – 1.40)	1.58×10^{-2}	1.20 (1.12 – 1.29)	3.76×10^{-7}
8	4391037	rs17070498	C / T	0.10 / 0.11	0.90 (0.77 – 1.06)	2.01×10^{-1}	0.12 / 0.15	0.79 (0.67 – 0.92)	1.79×10^{-3}	0.13 / 0.15	0.86 (0.74 – 1.01)	5.99×10^{-2}	0.13 / 0.16	0.84 (0.73 – 0.97)	1.59×10^{-2}	0.12 / 0.14	0.82 (0.71 – 0.95)	6.92×10^{-3}	0.84 (0.79 – 0.90)	7.83×10^{-7}
8	81379656	rs368280	A / C	0.19 / 0.23	0.81 (0.72 – 0.92)	8.82×10^{-4}	0.25 / 0.28	0.85 (0.75 – 0.95)	4.89×10^{-3}	0.27 / 0.27	0.96 (0.86 – 1.09)	5.40×10^{-1}	0.25 / 0.28	0.86 (0.77 – 0.96)	5.80×10^{-3}	0.26 / 0.29	0.87 (0.78 – 0.98)	1.68×10^{-2}	0.87 (0.82 – 0.92)	1.93×10^{-7}
9	3943997	rs7024842	T / C	0.26 / 0.24	1.12 (1.01 – 1.26)	3.47×10^{-2}	0.34 / 0.30	1.18 (1.06 – 1.32)	3.02×10^{-3}	0.36 / 0.32	1.18 (1.06 – 1.32)	2.35×10^{-3}	0.33 / 0.30	1.15 (1.04 – 1.27)	9.03×10^{-3}	0.35 / 0.33	1.08 (0.98 – 1.20)	1.26×10^{-1}	1.15 (1.09 – 1.21)	1.40×10^{-7}
9	119252629	rs6478241	A / G	0.39 / 0.37	1.11 (1.22 – 1.00)	4.96×10^{-2}	0.39 / 0.36	1.12 (1.25 – 1.01)	3.55×10^{-2}	0.40 / 0.37	1.14 (1.26 – 1.02)	1.84×10^{-2}	0.42 / 0.36	1.26 (1.39 – 1.14)	5.45×10^{-6}	0.40 / 0.37	1.16 (1.29 – 1.05)	4.11×10^{-3}	1.16 (1.21 – 1.10)	1.68×10^{-9}
10	105039048	rs1163084	T / C	0.45 / 0.47	0.92 (1.02 – 0.84)	1.02×10^{-1}	0.50 / 0.48	0.92 (1.02 – 0.83)	9.48×10^{-2}	0.50 / 0.47	0.89 (0.99 – 0.80)	2.55×10^{-2}	0.51 / 0.47	0.87 (0.96 – 0.79)	3.99×10^{-3}	0.51 / 0.46	0.81 (0.89 – 0.73)	2.93×10^{-5}	0.88 (0.92 – 0.84)	4.79×10^{-8}
12	53831472	rs12820532	A / G	0.15 / 0.14	1.12 (0.98 – 1.29)	7.90×10^{-2}	0.17 / 0.15	1.14 (0.99 – 1.32)	4.16×10^{-2}	0.19 / 0.17	1.16 (1.01 – 1.32)	2.26×10^{-2}	0.18 / 0.14	1.28 (1.13 – 1.46)	6.42×10^{-5}	0.19 / 0.17	1.18 (1.04 – 1.35)	6.73×10^{-3}	1.21 (1.13 – 1.29)	2.61×10^{-8}
12	57527283	rs11172113	C / T	0.37 / 0.40	0.90 (0.82 – 1.00)	4.43×10^{-2}	0.39 / 0.41	0.92 (0.83 – 1.02)	1.20×10^{-1}	0.38 / 0.42	0.85 (0.76 – 0.94)	2.18×10^{-3}	0.36 / 0.40	0.86 (0.77 – 0.95)	1.90×10^{-3}	0.38 / 0.41	0.86 (0.77 – 0.95)	2.68×10^{-3}	0.88 (0.83 – 0.92)	4.38×10^{-8}
18	6767965	rs4798494	T / C	0.33 / 0.36	1.13 (1.25 – 1.02)	1.95×10^{-2}	0.34 / 0.37	1.14 (1.27 – 1.02)	2.21×10^{-2}	0.32 / 0.34	1.07 (1.20 – 0.96)	2.32×10^{-1}	0.33 / 0.36	1.12 (1.24 – 1.01)	2.82×10^{-2}	0.32 / 0.35	1.16 (1.29 – 1.04)	5.52×10^{-3}	1.12 (1.18 – 1.07)	2.64×10^{-6}
18	43706491	rs28532950	T / C	0.40 / 0.37	1.13 (1.02 – 1.25)	1.43×10^{-2}	0.45 / 0.41	1.15 (1.03 – 1.27)	9.51×10^{-3}	0.46 / 0.43	1.14 (1.02 – 1.26)	1.64×10^{-2}	0.46 / 0.42	1.17 (1.06 – 1.29)	1.20×10^{-3}	0.44 / 0.43	1.06 (0.96 – 1.18)	2.14×10^{-1}	1.13 (1.08 – 1.19)	2.26×10^{-7}

[a](cases/controls)

Position from human NCBI build 37. MAF, minor allele frequency. OR, odds ratio for the minor allele. CI, confidence interval.

MA, migraine with aura. MA/MO, migraine with and without aura. MO, migraine without aura.

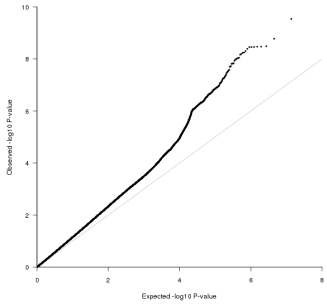

Figure 4.5: **Quantile-quantile plots of the GWAS discovery phase.** Plots of the fixed effect meta-analysis results in the MA discovery phase.

4.3 Replication stage

In the replication stage, SNPs from the top twenty nine loci were genotyped in 3268 migraine case and 2916 control European samples (Finland, The Netherlands and Spain) Of these six had at least one SNP that surpassed the threshold for genome-wide significance ($P = 5 \times 10^{-8}$) and 23 had at least one SNP with a P value lower than 5×10^{-6}.

Among the seventeen SNPs successfully genotyped, three reached the Bonferroni corrected replication threshold ($P \leq 2.94 \times 10^{-3}$): rs11892538 (OR=0.77 , 95% CI = 0.69 – 0.84, $P = 2.74 \times 10^{-5}$) rs4379368 (OR= 1.21, 95% CI = 1.08 – 1.35, $P = 8.68 \times 10^{-4}$) and rs11172113 (OR=0.86 , 95% CI = 0.79 – 0.92, $P = 3.66 \times 10^{-5}$). The effect estimate for rs11892538, rs4379368 and rs11172113 were concordant in direction among all replication cohorts with the discovery cohorts (Table 4.2). All the three SNPs reached genome-wide significance ($P \leq 5 \times 10^{-8}$) in a meta-analysis combining all cohorts (discovery and replication cohorts) (Table 4.3).

Table 4.2: **Summary results of the replication stage**

				Finnish (875/1025)[a]			Dutch (1043/910)[a]			Spanish (1350/981)[a]			Meta-analysis (3268/2916)[a]	
Chr	Position	SNP	Alleles minor/major	MAF cases/controls	OR(95% CI)	P	MAF cases/controls	OR(95% CI)	P	MAF cases/controls	OR(95% CI)	P	OR(95% CI)	P
1	3319777	rs4471209	A / G	0.32 / 0.35	0.88	6.92×10^{-2}	0.29 / 0.28	1.03	0.64	0.30 / 0.29	1.03	0.62	0.98	0.63
1	15532130	rs10927720	T / C	0.20 / 0.20	0.99	0.94	0.23 / 0.22	1.03	0.68	0.22 / 0.23	0.95	0.46	0.99	0.79
2	145222038	rs13403907	G / A	0.17 / 0.16	1.03	0.76	0.13 / 0.13	0.98	0.82	0.13 / 0.12	1.10	0.30	1.04	0.50
2	234821445	rs11892538	C / G	0.13 / 0.16	0.77	5.85×10^{-3}	0.17 / 0.21	0.80	5.20×10^{-3}	0.16 / 0.21	0.74	6.35×10^{-5}	0.77	2.74×10^{-5}
2	241447428	rs4676436	A / C	0.14 / 0.14	1.00	0.99	0.12 / 0.12	0.99	0.94	0.10 / 0.10	1.10	0.35	1.03	0.62
3	67144706	rs4311165	C / G	0.19 / 0.20	0.91	0.24	0.26 / 0.27	0.95	0.48	0.28 / 0.30	0.92	0.18	0.93	0.06
5	127722107	rs77050147	C / G	0.09 / 0.09	1.04	0.72	0.10 / 0.11	0.98	0.86	0.11 / 0.09	1.26	2.48×10^{-2}	1.09	0.14
6	12908747	rs13197912	T / A	0.30 / 0.29	1.02	0.80	0.37 / 0.38	0.97	0.60	0.40 / 0.37	1.13	4.63×10^{-2}	1.04	0.28
6	39177971	rs873690	C / G	0.08 / 0.07	1.03	0.79	0.06 / 0.05	1.05	0.72	0.05 / 0.05	0.95	0.70	1.01	0.89
6	143288832	rs1041655	A / C	0.42 / 0.43	0.97	0.60	0.38 / 0.37	1.07	0.32	0.39 / 0.38	1.05	0.43	1.03	0.46
7	40466200	rs4379368	T / C	0.17 / 0.13	1.31	2.75×10^{-3}	0.13 / 0.11	1.20	0.07	0.09 / 0.09	1.09	0.39	1.21	8.68×10^{-4}
8	4391037	rs17070498	C / T	0.10 / 0.11	0.91	0.38	0.15 / 0.13	1.19	0.06	0.16 / 0.16	1.00	0.96	1.03	0.55
8	81379656	rs368280	A / C	0.20 / 0.20	0.97	0.69	0.28 / 0.25	1.14	0.07	0.35 / 0.36	0.93	0.26	1.00	0.94
9	119252629	rs6478241	A / G	0.41 / 0.36	1.20	6.36×10^{-3}	0.39 / 0.38	1.03	0.69	0.45 / 0.44	1.05	0.38	1.09	2.32×10^{-2}
10	105039048	rs1163084	T / C	0.49 / 0.48	1.02	0.79	0.50 / 0.47	1.12	0.08	0.50 / 0.50	0.99	0.83	0.97	0.33
12	57527283	rs11172113	C / T	0.38 / 0.39	0.97	0.63	0.37 / 0.43	0.77	8×10^{-5}	0.32 / 0.36	0.85	7.61×10^{-3}	0.86	3.66×10^{-5}
18	43706491	rs28532950	T / C	0.38 / 0.38	1.01	0.87	0.45 / 0.42	1.16	2.20×10^{-2}	0.43 / 0.43	0.97	0.61	1.04	0.28

[a](cases/controls)

Position from human NCBI build 37. MAF, minor allele frequency. OR, odds ratio for the minor allele. CI, confidence interval.

MA, migraine with aura. MA/MO, migraine with aura and without aura. MO, migraine without aura.

Table 4.3: **Summary results of the discovery and replication stages**

				Discovery stage (5403/1537)[a]		Replication stage (3268/2916)[a]		Discovery and replication stages (8671/18243)[a]		
Chr	Position	SNP	Alleles minor/major	Meta-analysis OR (95% CI)	P	Meta-analysis OR (95% CI)	P	Meta-analysis OR (95% CI)	P	Gene
2	234821445	rs11892538	C / G	0.81 (0.76 – 0.86)	2.92×10^{-10}	0.77 (0.69 – 0.84)	2.74×10^{-5}	0.80 (0.76 – 0.84)	3.67×10^{-17}	*TRPM8*
7	40466200	rs4379368	T / C	1.20 (1.12 – 1.29)	3.76×10^{-7}	1.21 (1.08 – 1.35)	8.68×10^{-4}	1.20 (1.13 – 1.28)	1.36×10^{-9}	*C7orf10*
12	57527283	rs11172113	C / T	0.88 (0.83 – 0.92)	4.38×10^{-8}	0.86 (0.79 – 0.92)	3.66×10^{-5}	0.87 (0.84 – 0.91)	5.06×10^{-10}	*LRP1*

[a](cases/controls)
Position from human NCBI build 37.
MAF, minor allele frequency.
OR, odds ratio for the minor allele.
CI, confidence interval.

4.3.1 Discussion

In the first GWAS of imputed SNPs for migraine three loci associated with migraine were identified: one on chromosome 2q37.1 (rs11892538), one on chromosome 7p14.1 (rs4379368) and one on chromosome 12q13.3 (rs11172113) (Figure 4.6, 4.7 and 4.8). Two of the three loci (chromosome 2q37.1 and chromosome 12q13.3) had been already identified as associated with migraine in a previous study [244].

On chromosome 2q37.1, the most significantly associated marker rs11892538 maps to an intergenic region less than 5 kb away from the transient receptor potential cation channel 8 gene *TRPM8*. The second closest gene, encoding for Holliday junction recognition protein *HJURP*, maps 58.2 kb away from rs11892538. Given the current knowledge, *TRPM8* could be involved in migraine pathogenesis. *TRPM8* is a cold and menthol modulated ion channel with a role in the detection of cold in the mammals [246]. *TRPM8* is expressed in subpopulations of sensory

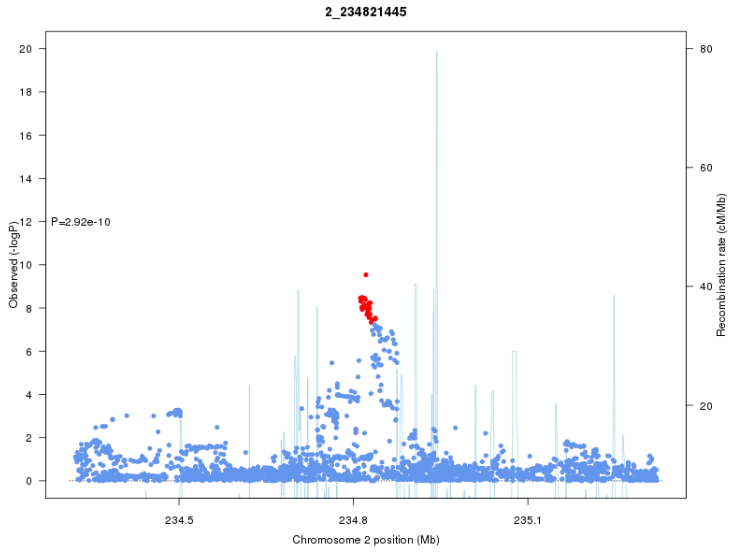

Figure 4.6: **Locus specific association plot: chromosome 2q37.1** The region +/- 500 kb around the most strongly associated SNP is shown. The diamond represents the most strongly associated SNP. P values are shown for the discovery stage. The blue line shows the recombination rate based on HapMap Phase II data. SNP and gene position are based on built 37.

neurons [247]. There is evidence suggesting that *TRPM8* may play a role in inflammatory and neuropathic pain [248]. Given that the migraine headache has some features in common with inflammatory and neuropathic pain, it is possible that *TRPM8* may play a role in its pathogenesis [249]. There is evidence suggesting that the in vivo antagonism of TRPM8 constitutes a possible strategy for treating neuropathic pain [250].

On chromosome 12q13.3, the most significantly associated marker rs11172113 maps to the first intron of low density lipoprotein receptor-related protein gene *LRP1*. *LRP1* is a cell surface receptor member of the low-density lipoprotein (LDL)-receptor family [251,252]. It is expressed in the vasculature, central nervous system, macrophages and adipocytes [253]. *LRP1* seems to play a role in various biological processes including lipoprotein metabolism [253]. Boucher et al. (2003)

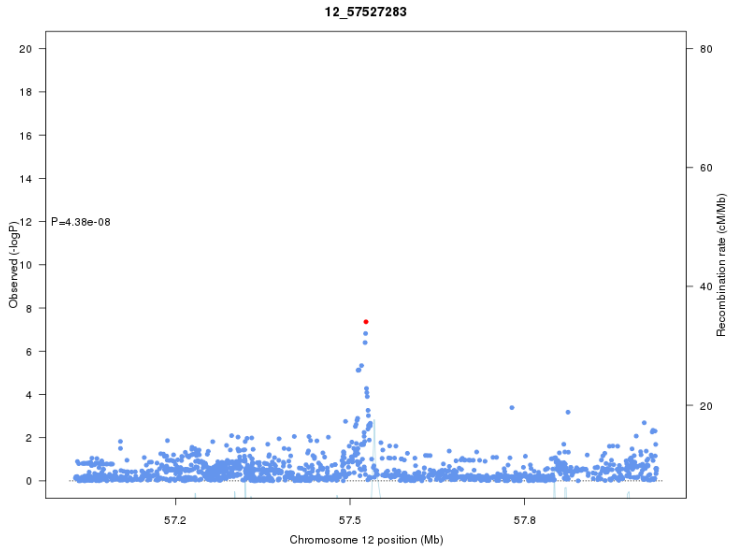

Figure 4.7: **Locus specific association plot: chromosome 12q13.3** The region +/- 500 kb around the most strongly associated SNP is shown. The diamond represents the most strongly associated SNP. P values are shown for the discovery stage. The blue line shows the recombination rate based on HapMap Phase II data. SNP and gene position are based on built 37.

have shown that inactivation of *LRP1* in vascular smooth muscle cells of mice leads to marked susceptibility to cholesterol-induced atherosclerosis [254]. Liu et al. (2010), performing neuronal lrp1 knockout in mice, have shown that the levels of glutamate receptors are reduced in lrp1 knockout neurons and that they are partially rescued by restoring neuronal cholesterol [255]. Glutamate is the main excitatory neurotransmitter in the central nervous system. Data from animal and human studies support a role of glutamate in the pathophysiology of migraine [256]. The second closest gene in the region, signal transducer and activator of transcription 6 (*STAT6*), maps 23.1 kb away from rs11172113. *STAT6* is a member of the STAT family of transcription factors, which plays a role in differentiation and function of T helper 2 (Th2) cells [257].

On chromosome 7p14.1, the most significantly associated marker rs4379368

maps to the dermal papilla derived protein 13 gene *C7orf10*. *C7orf10* is a peroxisomal glutaryl-CoA oxidase [258]. Mutations in this gene have been associated with glutaric aciduria type III, characterized by abnormal amounts of urinary glutaric acid [258]. Bennett et al. (1991) described a lack of peroxisomal glutaryl-CoA oxidase activity in a 1-year-old girl with failure to thrive and hematologic evidence of thalassemia. Sherman et al. (2008) reported three children homozygous for a nonsynonymous variant in C7orf10, who excreted large quantities of glutarate in the urine and remained healthy during a 15 years follow-up period [259]. The second closest gene to rs4379368 , encoding for cell division cycle 2-like 5 *CDC2L5*, maps 331 kb away. This gene encodes for a member of the cyclin-dependent serine/threonine protein kinase family. Members of cyclin-dependent serine/threonine protein kinase family have an important role in cell cycle control. The exact function of the protein encoded by *CDC2L5* has not been defined yet, but it has been suggested that it may have a role in mRNA splicing regulation [260].

In conclusion, in this first GWAS of imputed SNPs for migraine, three loci (one on chromosome 2q37.1, one on chromosome 7p14.1 and one on chromosome 12q13.3), associated with migraine were identified. Two of these three loci (one on chromosome 2q37.1 and chromosome 12q13.3) had already been found associated with migraine in a recent population based study [244]. The most significantly associated marker on chromosome 2q37.1 (rs11892538), maps 5 kb away from the *TRPM8* gene, which encodes for a ion channel with a role in pain pathogenesis [248]. On chromosome 12q13.3, the most significantly associated marker (rs11172113) maps to the first intron of the *LRP1* gene. *LRP1* has been shown to modulate the neural glutamate receptor levels, and therefore, the association of

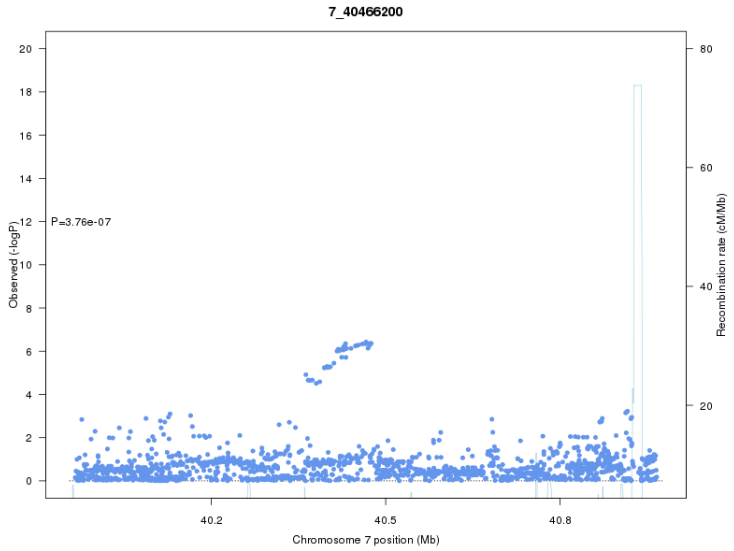

Figure 4.8: **Locus specific association plot: chromosome 7p14.1** The region +/- 500 kb around the most strongly associated SNP is shown. The diamond represents the most strongly associated SNP. P values are shown for the discovery stage. The blue line shows the recombination rate based on HapMap Phase II data. SNP and gene position are based on built 37.

LRP1 with migraine provides further support to the role of glutamate in migraine pathogenesis [255]. The functional role of the new third locus on chromosome 7p14.1 is not currently definable. Future functional studies on the role of genes present in the locus (*C7orf10* and *CDC2L5*) will be provide an understanding of its functional link with migraine. Larger GWAS, currently underway, will allow the identification of further genetic variants underlying the pathophysiology of migraine.

Chapter 5

GENCODE exome

5.1 Introduction

Genome-wide association studies have been successful in identifying common variants associated with complex human diseases and traits [194]. However, in most of the cases the portion of the heritability explained by these variants is modest [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].

The development of next generation sequencing (NGS) has allowed the systematic discovery of rare variants in thousands of samples [55]. The cost of whole-genome sequencing has been falling dramatically over the last couple of years, however it is still too expensive to be applied to large scale genomic studies aimed at identifying variants associated with complex diseases. Currently, the combination of next generation sequencing technologies with efficient methods of sequence capture has enabled the widespread targeting of the exome [195–199]. Exome re-

sequencing constitutes an effective tool for discovering coding variants underlying monogenic diseases and for identifying coding variants associated with complex diseases [261–263].

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.

The two most widely used commercial kits for capturing the exome, available at that time, (NimbleGen Sequence Capture 2.1M Human Exome Array, http://www.nimblegen.com/products/seqcap/ and Agilent SureSelect Human All Exon Kit, http://www.genomics.agilent.com) targeted exons from genes in the consensus coding sequence (CCDS) consortium database, in addition to a selection of miRNAs and non-coding RNAs [264]. Although the CCDS database contains a high-quality set of consistently annotated protein-coding genes, many annotated genes, with solid evidence of transcription, are not part of this set yet. In addition, in the CCDS database only 21% of the genes have alternative spliced variants included.

To address this shortcoming, a more complete set of target regions for the human exome, based on the GENCODE annotation was designed and experimentally tested [200]. The GENCODE is part of the Encode project and responsible for the annotation and experimental validation of gene loci on the human genome.

Table 5.1: **Comparison of the three different exome capture sets**

	[a]NimbleGen CCDS	[b]Agilent CCDS	GENECODE exome
Number of bait regions	197218	316000	406539
Genome coverage (Mb)	34.1	37.6	[c]47.9 [d](35.2)
ECRs covered [e](%)	150529 (72.7)	164225 (79.3)	205031 (99.0)
Transcripts covered [e](%)	66828 (81.0)	71279 (86.4)	81204 (98.4)
Genes covered [e](%)	28203 (76.5)	30030 (81.5)	35989 (97.7)

[a]NimbleGen Sequence Capture 2.1M Human Exome Array
[b]Agilent SureSelect Human All Exon Kit
[c]Total length of bait regions including flanking regions.
[d]Design target length without flanking regions.
[e]Percentage of the GENCODE exome design target

5.2 Results

5.2.1 The GENCODE exome features

A comparison of the coverage of the bait/oligonucleotide positions of available CCDS-based exome sets and our GENCODE exome set with the GENCODE design target showed an increased coverage of our set (Table 5.1).

The GENCODE exome baits covered 99% of the design target, resulting in additional 59600 exons available for capture, which were not present in either one of the CCDS-based sets [265]. The missing 1% were regions for which reliable bait design was not possible.

A comparison of exon and transcript coverage of the available CCDS based exome sets and the GENCODE exome set with three current reference gene sets (CCDS, RefSeq, Gencode), showed that the GENCODE exome set covered a

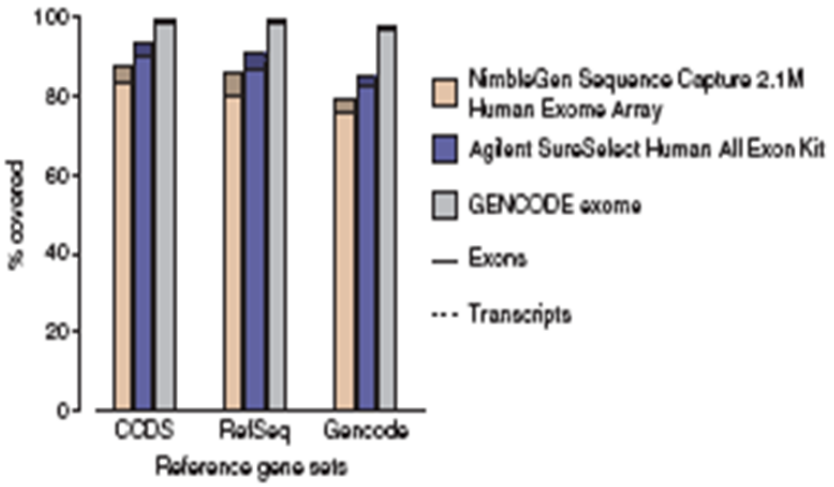

Figure 5.1: **Comparison of exon and transcript coverage of the available CCDS-based exome sets and the GENCODE exome set with three current reference gene sets.** The histograms show the near-complete coverage by the GENCODE exome set of all reference sets.CCDS database March 2010, RefSeq genes March 2010 and GENCODE version 3c.

greater percentage of the reference genes (Table 5.2, figure 5.1). For example, the GENCODE exome set covered an additional 9% of the exons from the CCDS database and 12% of the exons from RefSeq compared to the CCDS-based exome sets.

The content present exclusively in the GENCODE exome set consists of 38933 exome cluster regions, which contain 5594 additional genes of the GENECODE exome design target. Of these additional genes, the 4363 Ensembl-53-based genes include 1881 (43%) genes with an official HGNC identifier, 711 (16%) with an OMIM entry and 1410 (32%) with a Gene Ontology annotation (Table 5.3) [265]. The content of repetitive/low-complexity sequence in the GENCODE exome set is similar to the available CCDS-based exome sets (Table 5.4).

A comparison with a sequence uniqueness mask supports these findings (Table

Table 5.2: **Exon and transcript coverage of the three different exome caputure sets**

Exome set	CCDS exons	CCDS transcripts	RefSeq exons	RefSeq transcripts	GENECODE exons	GENECODE transcripts
NimbleGen Sequence Capture 2.1M Human Exome Array	82.80%	87.19%	79.47%	85.28%	75.29%	79.16%
Agilent SureSelect Human All Exon Kit	90.05%	93.16%	86.39%	90.91%	82.15%	84.67%
GENCODE exome	99.18%	99.63%	98.75%	99.19%	97.31%	96.51%
Additional content of the GENCODE exome	9.12%		12.36%			

Table 5.3: **Comparison of the three different exome caputure sets with the GENECODE design target**

	Genes	Transcripts	Exons
[a]Totals	36853	82522	463778
Nimblegen CCDS [b](%)	28203 (76.5%)	66828 (81.0%)	307866 (66.4%)
Agilent CCDS [b](%)	30030 (81.5%)	71279 (86.4%)	334191 (72.1%)
GENCODE exome [b](%)	35989 (97.7%)	81204 (98.4%)	397856 (85.8%)
Covered by all 3 sets	27828	65943	303483
Covered by Nimblegen CCDS only	10	12	318
Covered by Agilent CCDS only	0	0	0
Covered by Gencode exome only	5594	9052	59600
Covered by Nimblegen and Agilent CCDS only	0	0	0
Covered by Nimblegen CCDS and Gencode exome only	365	873	4065
Covered by Agilent CCDS and Gencode exome only	2202	5336	30708

[a]Referred to the GENECODE design target
[b]Percentage of the GENCODE exome design target

Table 5.4: **Assessment of repeat and low-complexity coverage of the three exome sets.** Repeats and low-complexity regions identified with RepeatMasker (parameters: -nolow -species homo -s), Dust and TRF using Ensembl 53 data.

Exome set	Total base pairs	Base pairs with repeats	Ratios
NimbleGen CCDS	34108810	884080	38.6
Agilent CCDS	37640396	799357	47.1
GENCODE exome	47933967	1303879	36.8

5.5).

The list of 5594 additional genes and regions targeted by the GENCODE exome exclusively, data for the final GENCODE exome and the initial design target is available on our ftp site (http://ftp.sanger.ac.uk/gencode/exome).

The 406539 bait locations are supplied as a Distributed Annotation System data source (das.sanger.ac.uk/das/Exome) and they can be displayed in genome browsers such as Ensembl (version 53; http://tinyurl.com/browse-exome) [265].

5.2.2 The GENCODE exome performance

To evaluate the performance of the GENCODE exome set, DNA from three HapMap samples (NA12878, NA07000 and NA19240) was captured using both the Agilent SureSelect Human All Exon Kit and GENCODE exome baits.

Moreover, to evaluate the performance of the GENCODE exome set using DNA from clinical samples, DNA from seven individuals recruited from a clinical neurological unit was captured using the GENCODE exome baits. Samples were sequenced as described in the methods section. Variants were called using SAMtools v0.1.71 and GATK, the intersection of the resulting calls in the GENECODE target regions (39.3 Mb) with a sequence read depth of $\geq$ 8x was reported and compared to known variants in all the samples.

On average 97% of reads could be successfully mapped back to the genome, 67% of the mapped reads were uniquely mapped and 82% 0f the unique mapped reads derived from the capture target regions (Table 5.6).

The average coverage of the HapMap samples was 73-fold from 9.2 Gb of se-

Table 5.5: **Bait/probe covered CTR (Capture Target Regions) regions assessed using a uniqueness mask** The used uniqueness mask was developed by Heng Li for the 1000 Genomes project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/README_hg36_uniqueness_mask)

Exome set	CTR count	Total target size (bp)	[a]Type 0 bases	[a]Type 1 bases	[a]Type 2 bases	[a]Type 3 bases
Nimblegen CCDS	165637	37640396	0 [b](0%)	1570956 [b](4.17%)	838459 [b](2.23%)	35230981 [b](93.60%)
Agilent CCDS	176159	34108807	17688 [b](0.05%)	1236875 [b](3.63%)	745619 [b](2.19%)	32108625 [b](94.14%)
GENCODE exome	206275	47933967	0 [b](0%)	2811712 [b](5.87%)	1177904 [b](2.46%)	43944351 [b](91.68%)

[a]Sequenceability/uniqueness measures used in the method:
- Type 0 : all 35mers covering this site cannot be mapped back due to "N"s in the reference
- Type 1 : otherwise (>35×0.5 reads are exact repeats)
- Type 2 : if not 3, $\geq 35 \times 0.5$ reads 1-away unique
- Type 3 : $\geq 35 \times 0.5$ reads 2-away unique

[b]Percentage of total target bases

Table 5.6: **Mapping statistics for clinical and HapMap samples using GENCODE and Agilent CCDS exome captures.**

Sample name	Sanger 1	Sanger 2	Sanger 3	Sanger 4	Sanger 5	Sanger 6	Sanger 7	NA12878	NA070000	NA19240	NA12878	NA070000	NA19240
Bait set	GENECODE										Agilent CCDS		
Lane sequenced	3	3	3	3	3	3	3	4	3	3	3	4	4
Reads	105712968	115540748	107420522	101912732	129221672	205597102	203472750	249957386	216080714	175087650	131337304	174698654	206110412
Reads mapped	101784111 [a](96.28%)	112290675 [a](97.19%)	104444274 [a](97.23%)	99877068 [a](98.00%)	126463612 [a](97.87%)	197922085 [a](96.27%)	196009151 [a](96.33%)	237675276 [a](95.09%)	212171108 [a](98.19%)	168827968 [a](96.42%)	123007651 [a](93.66%)	169580062 [a](97.07%)	198938967 [a](96.52%)
Unique reads mapped	77800441 [b](76.44%)	81310815 [b](72.41%)	74873012 [b](71.69%)	74979882 [b](75.07%)	97666019 [b](77.23%)	145520170 [b](73.52%)	143615416 [b](73.27%)	134894523 [b](56.76%)	122795772 [b](57.88%)	77220371 [b](45.74%)	91946948 [b](74.75%)	97248119 [b](57.35%)	117233076 [b](58.93%)
Unique reads mapped to CTR +/- 250bp	59625505 [c](76.64%)	69956576 [c](80.04%)	65714990 [c](87.77%)	63735274 [c](85.00%)	68908856 [c](70.56%)	86495628 [c](59.44%)	86910473 [c](60.52%)	124450772 [c](92.26%)	113776953 [c](92.66%)	71794822 [c](92.97%)	83772118 [c](91.11%)	89399731 [c](91.93%)	99555617 [c](84.92%)
Unique reads mapped to GENCOODE ECRs	40507262 [c](52.07%)	47280119 [c](58.15%)	46480228 [c](62.08%)	43032073 [c](57.39%)	45129213 [c](46.21%)	66486708 [c](67.96%)	66344444 [c](46.20%)	84566996 [c](62.69%)	81625999 [c](66.47%)	51683594 [c](66.93%)	56713564 [c](61.68%)	66088857 [c](67.96%)	69982605 [c](59.70%)
[d]Mean depth in the GENCODE ECRs	46.38	54.08	53.45	49.83	51.02	76.96	76.61	95.37	92.37	58.81	65.08	75.92	80.14
[d]Median depth in the GENCODE ECRs	34	39	35	34	36	55	55	63	62	36	43	47	57
Bases covered $\geq 8x$ in the GENECODE design target	32914133 [e](83.66%)	33102756 [e](84.14%)	32201108 [e](81.85%)	32536951 [e](82.70%)	32220983 [e](81.90%)	35330571 [e](89.80%)	34963040 [e](88.87%)	32415365 [e](82.39%)	32910800 [e](83.65%)	31502723 [e](80.07%)	28721876 [e](73.01%)	28903556 [e](73.47%)	29429202 [e](74.80%)

[a]Percentage of reads
[b]Percentage of reads mapped
[c]Percentage of unique reads mapped
[d]Calculated after duplicate read removal
[e]Percentage of GENECODE exome design target bases
CTR, Capture Target Regions; ECRs, Exome Cluster Regions

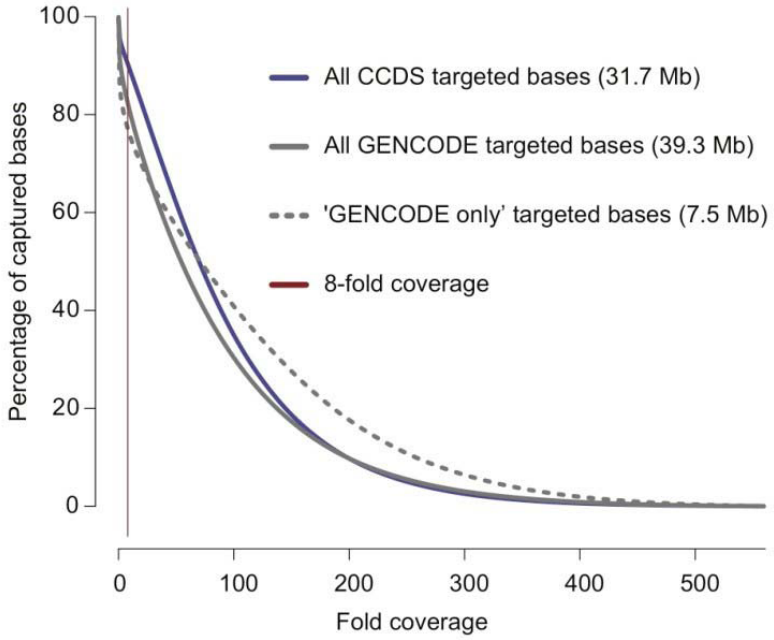

Figure 5.2: **Cumulative distribution of base coverage for HapMap samples.** The blue line represents the cumulative distribution of base coverage for the CCDS based captures. The continuous grey line represents the cumulative distribution of base coverage for the GENCODE exome based captures, and the dashed grey line represents the cumulative distribution of base coverage for the GENCODE exome based captures in the regions covered only by the GENCODE exome baits. The thin red vertical line indicates a coverage of eightfold, which is the coverage commonly required for variant calling.

quence for the CCDS-based captures and 82-fold from 11.5 Gb of sequence for the GENCODE exome captures. The average coverage for the clinical samples, captured only with the GENCODE exome baits, was 58-fold from 7.5 Gb of sequence. The coverage has been calculated only using reads with a mapping quality of 10.

For the HapMap samples, on average, 96% of targeted bases were covered at least once and 90% were covered eightfold or more for the CCDS-based captures. Similar figures were obtained for the the GENCODE exome based captures with 92% of targeted bases covered at least once and 83% covered eightfold or more (Figure 5.2).

For the clinical samples, on average 95% of targeted bases were covered at least once and 88% covered eightfold or more (Figure 5.3).

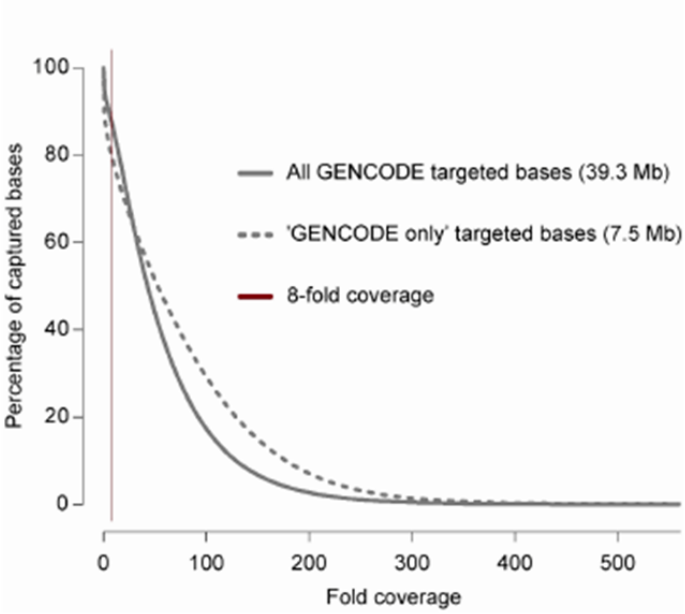

Figure 5.3: **Cumulative distribution of base coverage for the clinical samples.** The continuous grey line represents the cumulative distribution of base coverage in all the targeted regions, and the dashed grey line represents the cumulative distribution of base coverage in the regions covered only by the GENCODE exome baits. The clinical samples were captured only using the GENECODE based Agilent SureSelect Human All Exon Kit. The thin red vertical line indicates a coverage of eightfold, which is the coverage commonly required for variant calling.

The results demonstrate that the GENCODE based captures performed equally to the CCDS based captures. Moreover, considering the regions covered only by the GENECODE baits similar figures were obtained, proving that these regions perform equally to the CCDS regions (Figure 5.2).

Only unique reads mapped to the target were used for variants calling. For the HapMap samples, an average of 22271 variants, of which 2.6% were novel, were called in GENCODE based captures compared with an average of 18554 variants, of which 1.7% were novel, called in the CCDS-based captures (Table 5.7). Variants were defined as novel if they were not present either in dbSNP18 (version 130) or 1000 Genomes project (1000 Genomes Project Consortium, http://www.1000genomes.org, released on 26 March 2010).

Table 5.7: **Variant calling statistics for clinical and HapMap samples using GENCODE and Agilent CCDS exome captures.**

Sample name	Sanger 1	Sanger 2	Sanger 3	Sanger 4	Sanger 5	Sanger 6	Sanger 7	NA12878	NA070000	NA19240	NA12878	NA070000	NA19240
Bait set	GENECODE										Agilent CCDS		
Variants	21170	21529	21052	21445	21124	23612	23276	20780	21513	24520	16732	17014	21915
% dbSNP (version 130)	93.7	93.5	93.7	93.8	92.1	93.5	93.5	96.5	94.1	94.8	98.0	95.2	95.5
% dbSNP and/or 1000 Genomes (26/03/10 pilot 1)	96.3	95.1	96.3	96.1	94.7	96.1	96.1	97.7	97.2	97.3	99.0	97.9	98.1
Heterozigous	12604	13241	12938	13153	13297	14476	14321	12675	13588	16121	10094	10583	14414
Ti/Tv	3.029	2.996	3.036	3.025	2.930	3.021	3.120	3.069	3.112	3.138	3.235	3.258	3.322
% [a] Concordant	99.78	99.79	99.89	99.72	99.72	99.83	$	99.31	99.61	99.16	99.34	99.66	99.17
Synonymous	9196	9249	9072	9191	8948	10220	10111	8480	9207	10568	7979	8133	10528
Non synonymous	8608	8804	8692	8828	8696	9634	9385	8758	8703	9958	6863	6976	8918
Stop gained	86	85	80	89	128	87	95	80	83	99	44	40	51
Variants in the GENCODE only [b] ECRs	5179	5212	5117	5162	5162	5414	5424	5017	5319	5887			

[a] Called variants in the exome captures were compared with Illumina 660K chip genotypes for clinical samples and with the HapMap3 genotypes for HapMap samples.

[b] Excluding flanking regions

Ti, transitions; Tv, transversions. $, missing data. ECRs, Exome Cluster Regions.

An average of 21866 variants, of which 4.2% were novel, were called in the clinical samples. The HapMap samples and the clinical samples had been previously genotyped on arrays, therefore the variants called in the GENECODE based captures were compared to the genotyped SNPs. The concordance rate was found to be 99.7%.

Most of the variants, for which genotypes were discordant between array genotyping and sequencing, were discrepant only in one sample, suggesting that the number of systematic either genotyping or sequencing errors was low.

The 22002 variants found on average in the GENCODE exome captures included, 9006 non-synonymous variants, 9424 synonymous variants and 91 stop-gained variants. Meaning that 268 synonymous variants, 256 non-synonymous variants and 2.6 stop-gained variants were found per megabase of the targeted genomic sequence (35.2 Mb) , corresponding to a total of 626.6 variants per megabase. In the CCDS captured samples among the 18554 coding SNPs found on average, there were 7585 non-synonymous variants, 8880 synonymous variants and 45 stop-gained variants, corresponding to a total of 512 variants per megabase.

5.3 Discussion

The GENCODE gene set used as reference for the design of our ECRs (Exome Cluster Regions) provides a more complete set of targets, as it is the result of the merging of the thorough manual Havana and the genome-wide automatic Ensembl annotation. Both Havana and Ensembl are part of the CCDS consortium, and all

of the CCDS annotated genes have been incorporated into the new target set.

The new set set of target regions for exome capture includes genes potentially relevant for the discovery of disease-associated variants. Among the genes captured by our new expanded set, there are members of well-characterised gene families, which have been associated with important medical conditions. For example, 43 genes, captured only by our set, encode for ion channel subunits. Mutations in ion channel genes have previously been found to cause a range of channelopathies, including arrhythmias and inherited paroxysmal neurological disorders [266]. Seventy genes, encoding for proteins with kinase activity are exclusively present in the GENECODE based set. Members of the protein kinase family have been found commonly mutated in cancer and are considered to be candidate targets for the development of new anticancer therapies, therefore it is important that they are covered in cancer sequencing studies. [267].

The coverage over two genes targeted only by the GENCODE based set, *ABCB11* and *XPC*, (Figures 5.4 a and b) demonstrates that we have been able to design baits for genes not represented in the previous CCDS based sets and that these genes are efficiently captured and uniformly covered. Each exon is covered on average more than eightfold, which is most commonly required depth for variant calling. *ABCB11* and *XPC* are medically relevant genes, since they have already been associated with diseases, and provide examples of candidate disease genes that are missing from the existing CCDS based exome capture sets.

The use of the GENCODE based baits has already allowed the identification of a pathogenic mutation in a gene causing an autosomal recessive Dwarfism syndrome, which would not have been discovered using a standard CCDS-based sets [268].

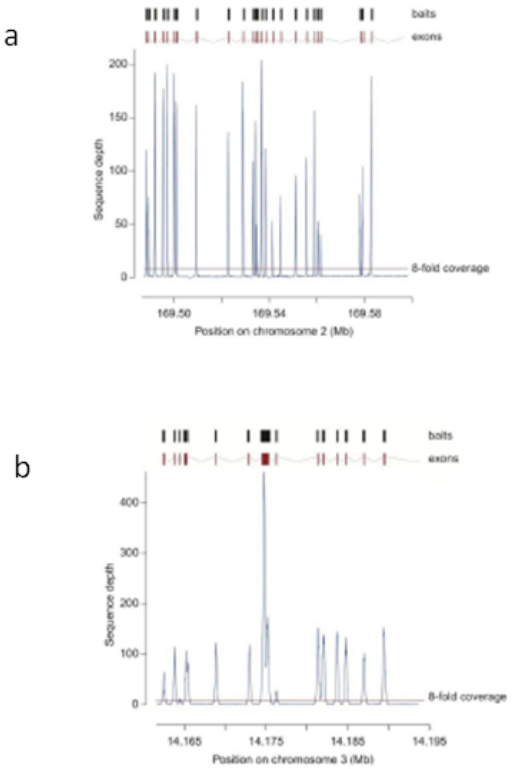

Figure 5.4: **Coverage achieved by the GENECODE based set.** Detailed view of the average depth in the seven clinical samples across two genes that are unique to the GENCODE based set: (a) *ABCB11* and (b) *XPC*. In the upper part of each panel, the positions of the GENCODE exome baits are represented by dark boxes above the gene structure (adapted from the Ensembl genome browser) in red. The increased sequencing depth of the eighth exon of *XPC* is due to high coverage of this larger exon by eight different baits, whereas the other smaller exons are covered by one or two baits. The horizontal thin red line indicates a coverage of eightfold.

The advent of the GENCODE exome represents a substantial improvement to the currently available designs for exome capture, allowing the capture of a more complete target. We estimate that we were able to call variants in 84% of the total GENCODE target regions. The fraction of callable exome regions is likely to increase further with improving sequencing technology and sequencing depth. The GENCODE exome design is currently used by the International Cancer Genome Consortium (ICGC; http://www.icgc.org) for their exome sequencing projects, aimed at obtaining a comprehensive description of different tumor types and by the UK10K project (http://www.uk10k.org).

5.4 Notes

Sequencing data have been deposited at the European GenomePhenome Archive (http://www.ebi.ac.uk/ega/) under accession number EGAS00001000016 and the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession ERP000523.

Chapter 6

Exome sequencing in Familial Hemiplegic Migraine

6.1 Introduction

For complex diseases, such as migraine, GWAS have been successful in identifying associated common variants, which influence disease susceptibility [194]. Most of the identified variants have a small effect on the disease risk and the majority of the genetic contribution to the disease remains unexplained [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].

Technological advances in sequencing and bioinformatics approaches have made a reality the identification of rare variants [55]. Over the last decade, the development of new massively parallel sequencing methods, referred to as next-generation sequencing (NGS), has led to reductions in the costs of sequencing [56, 57]. 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 cases and controls [51]. However, to sequence large sample sizes is still to expensive. One type of approach, which is currently affordable, to identify rare variants underlying diseases consists in sequencing cases from families that have multiple affected individuals (family-based sequencing) [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 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. Familial hemiplegic migraine (FHM) is a subtype of migraine with aura (MA), in which the aura is characterized by motor weakness of variable intensity. An autosomal dominant mode of inheritance has usually been observed. FHM is genetically heterogeneous and 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 cases [94, 96, 101]. Mutations in these genes seem to lead to a modification of the activity of the encoded proteins, which increases neuronal depolarization and might trigger cortical spreading depression (CSD) and CSD-like events [96, 109–111, 118, 119, 125]. CSD is a process characterized by a slowly propagating wave of neuronal depolarization, followed by a suppression of neuronal activity [80, 81]. There is evidence that CSD is the pathophysiological substrate of the aura symptoms [79]. In the present study, in order to identify novel coding variants contributing to the genetic basis of FHM, the exome of 88 cases from 44 families with several affected members was sequenced.

6.2 Results

6.2.1 Whole exome capture of 88 FHM cases

In collaboration with clinicians working at the Helsinky University Hospital, 88 FHM cases from 44 families with several members affected by FHM were selected. All the cases were diagnosed as having hemiplegic migraine according to the IHS criteria by experienced neurologists (ICHD-II) [58].

Whole exome sequencing on the 88 FHM cases separately was performed by subjecting the DNA to exome capture using the GENCODE exome capture solution and sequencing the captured regions on a Genome Analyzer II. Pair-ended 76 base-pair (bp) reads were aligned to the reference genome (hgs19). An average of 102 million reads were produced, 100 million reads were mapped to the human reference sequence (hgs19) and 93 million reads were uniquely mapped. Of the uniquely mapped reads an average of 57% mapped to the GENCODE target regions. After removing duplicate reads the average coverage was 78.71X per exome. On average, 98.18% of targeted bases were covered at least once, 94.28% were covered at least four times and 88.97% were covered at least ten times, which is similar to previous studies (Table 6.1) [263].

Among the 88 analyzed samples, a mean of 22169 variants per exome (Table 6.2), of which 21896 (98.8%) were included in dbSNP or in 1000 Genomes Pilot Project (26/03/10 release), was called. These variants included a mean per sample of 21000 coding variants, of which there were 9432 missense substitutions, 9958 synonymous substitutions, 106 premature termination codons, 1504 splice site variants. Consistent with expectation, on average, the ratio of synonymous to non-synonymous variants was 1.06 per exome [263].

Table 6.1: **Mapping statistics.**

	[a]Mean ± s.d.
Reads	101919389 ± 22850891
Reads mapped	99984936 ± 22232259
Unique reads mapped	92652609 ± 17005248
Unique reads mapped to CTR [b]%	66.65 ± 2.18
Unique reads mapped to GENCOODE ECRs [b]%	56.95 ± 2.26
Mean depth in the GENCODE ECRs [c]X	78.71 ± 15.29
Bases covered $\geq 1x$ in the GENECODE ECRs %	98.18 ± 0.31
Bases covered $\geq 4x$ in the GENECODE ECRs %	94.28 ± 0.59
Bases covered $\geq 10x$ in the GENECODE ECRs %	88.97 ± 0.98

[a]Of total 88 samples
[b]Percentage of unique reads mapped
[c]Calculated after duplicate read removal
CTR, Capture Target Regions (51.5 Mb)
GENCODE ECRs, GENCODE Exome Cluster Regions (39.3 Mb)

On average 272 variants per exome were novel. I defined as novel variants not present in the reference genome (hgs19), in the dbSNP database (version 130) and/or 1000 Genomes Pilot Project data (26/03/10 release).

6.2.2 Potentially pathogenic variants underlying FHM

To identify potentially pathogenic variants underlying FHM, in each one of the 44 families, we selected coding variants shared between the affected individuals. Assuming that synonymous variants are less likely to have functional consequences, we selected shared missense, nonsense and splice site variants and we defined them functional. In this initial analysis I focused on rare variants. I defined a variant as rare if not present in the dbSNP database (version 130), 1000 Genomes Pilot Project (26/03/10 release) or 500 Exomes Project, additional 390 exomes sequenced at Wellcome Trust Sanger Institute or if present in any of the existing databases and no allele frequency information was reported. To summarize, for each family we selected rare functional variants shared by both affected.

On average, we identified 31 shared rare functional variants per family. These variants included 27 missense, three splice site variants and one premature termination codons per family (Table 6.3).

I investigated how many genes had one or more rare functional variant shared by the the two affected members of each family in more than one family. In 122 genes there were shared rare functional variants in two or more families. In 21 of the 122 genes, shared rare functional variants were present in three or more families (Figure 6.1).

Table 6.2: **Variant calling statistics.**

	[a]Mean ± s.d.
Variants	22168.66 ± 452.61
dbSNP (version 130)	21639.55 ± 434.74
1000 Genomes Pilot (26/03/10)	21160.26 ± 416.60
dbSNP and/or 1000 Genomes Pilot	21896.20 ± 439.90
Coding variants	21000.74 ± 422.23
Synonymous	9958.13 ± 202.47
Missense	9432.45 ± 205.58
Splice site	1504.38 ± 34.51
Premature termination	105.78 ± 7.27

[a]Of total 88 samples

Table 6.3: **Rare functional variants shared by the two cases of each family**

	[a]Mean ± s.d.
Total	30.66 ± 14.78
Missense	27.20 ± 13.25
Splice site	2.91 ± 2.03
Premature termination	1.41 ± 0.62

[a]Of total 44 families

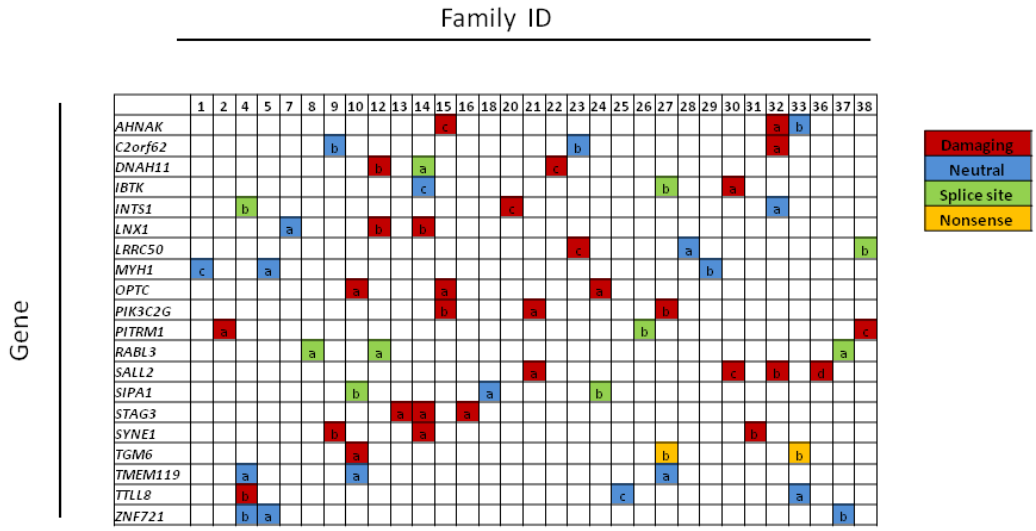

Figure 6.1: **Genes with shared rare functional variants in three or more families.** Letters (a, b, c, d) represent variants. Colors represent functional consequence of the variant as specified in the legend on the left side of the figure. Damaging (red) and neutral (blue) represent the predicted deleteriousness of the missense mutations.

Taking into account the predicted deleteriousness of the missense mutations, of the 122 genes, 47 had possibly damaging missense, splice site or nonsense variants in two or more families (Figure 6.2).

6.2.3 Family 1: known causal *CACNA1A* mutation

Among the rare functional variants shared by the two affected members of each family, I checked whether there were variants in three genes (*CACNA1A*, *SCN1A*) and (*ATP1A2*) which have been already found implicated in FHM [94, 96, 101]. In one family, family 1 (Figure 6.3) a c.1997C>T (rs121908212) nucleotide change resulting in a p.Thr666Met amino acid change in the *CACNA1A* gene was iden-

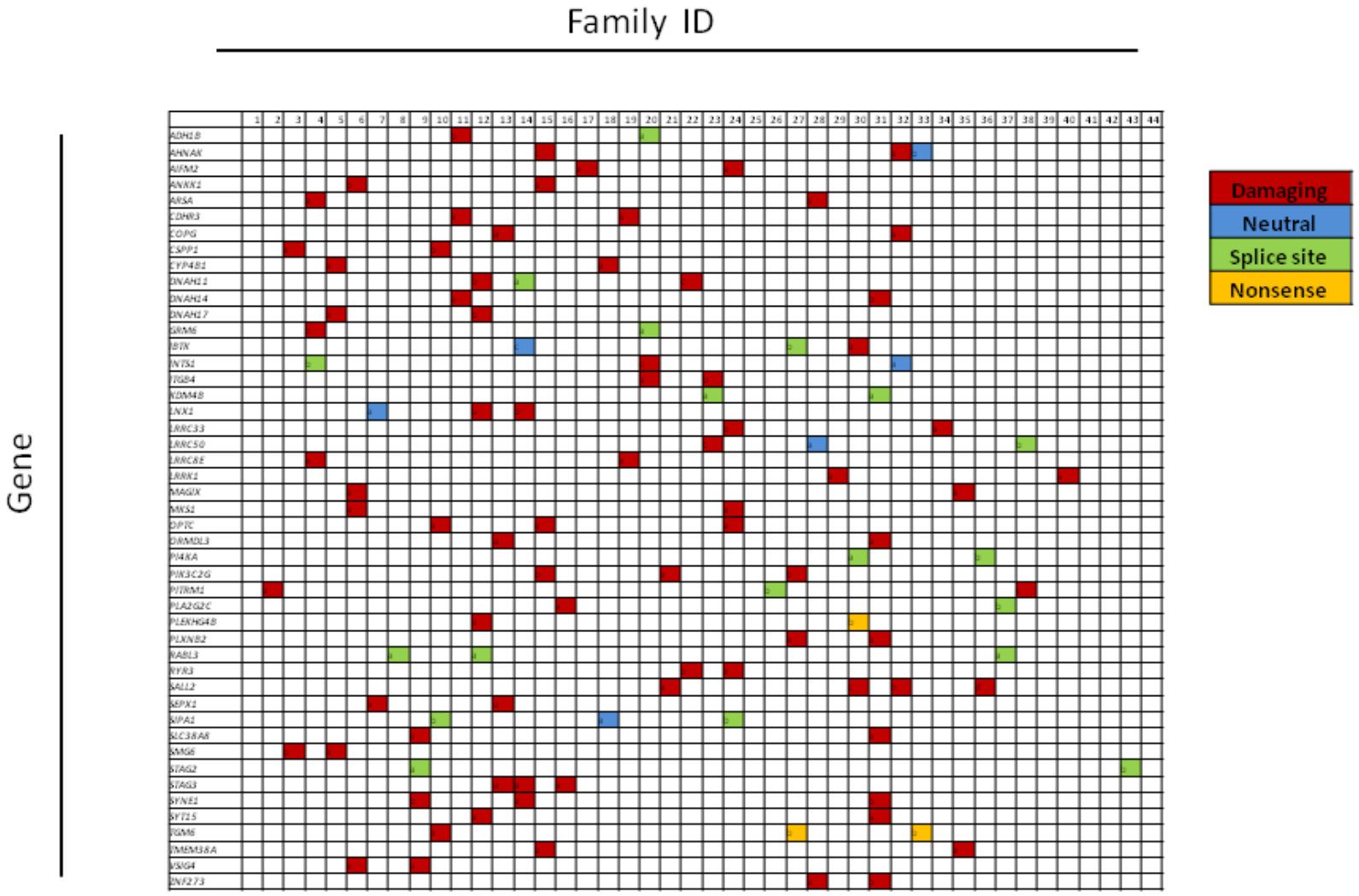

Figure 6.2: **Genes with possibly damaging missense, splice site or nonsense variants in two or more families.** Letters (a, b, c, d) represent variants. Colors represent functional consequence of the variant as specified in the legend on the left side of the figure. Damaging (red) and neutral (blue) represent the predicted deleteriousness of the missense mutations.

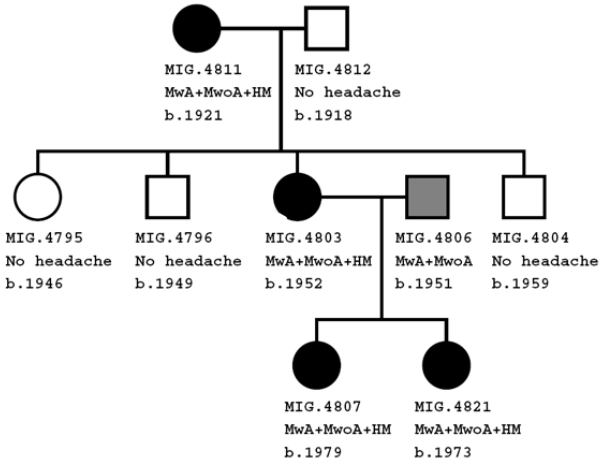

Figure 6.3: **Pedigree of family 1.** Individuals are represented as males (squares), females (circles) and affected (blackened symbols). In family 1, MIG 4821 (proband), MIG 4807 (sister), MIG 4803 (mother) and MIG 4811 (grandmother) suffered from hemiplegic migraine. Exome sequencing was done for MIG 4821 (proband) and MIG 4803 (mother), since DNA for MIG 4811 (grandmother) was not available.

tified (Figure 6.4). This mutation had been previously described as a cause of FHM [101,102,130,131,134]. *CACNA1A* mutations have been found to be 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 have been described to cause a severe phenotype, in which HM is associated with epilepsy or progressive cerebellar ataxia [106,130]. Moreover, 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

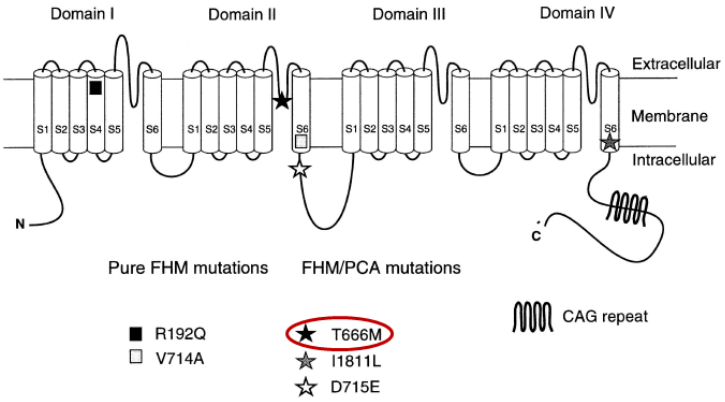

Figure 6.4: *CACNA1A* **missense mutations causing hemiplegic migraine (HM).** A schematic structure of the *CACNA1A* is represented, and positions of some HM causing mutations are marked. A red circle has been drawn around the mutation detected in the 2 cases of family 1. PCA, progressive cerebellar ataxia. Adapted from Ducros et al. (1999) [130].

channel involved in neurotransmitter release at many synapses in the central nervous system [108]. The p.Thr666Met amino acid change has been shown to shift the voltage dependence of activation toward more negative potentials and to slow down the time of recovery from channel inactivation, leading to alterations in the Ca^{++} influx and neurotransmitter release [109]. Variations in neurotransmitter release in the brain might render patients susceptible to migraine attacks [109].

This variant will be genotyped in the other affected individuals of the family to confirm that it co-segregates with the FHM phenotype.

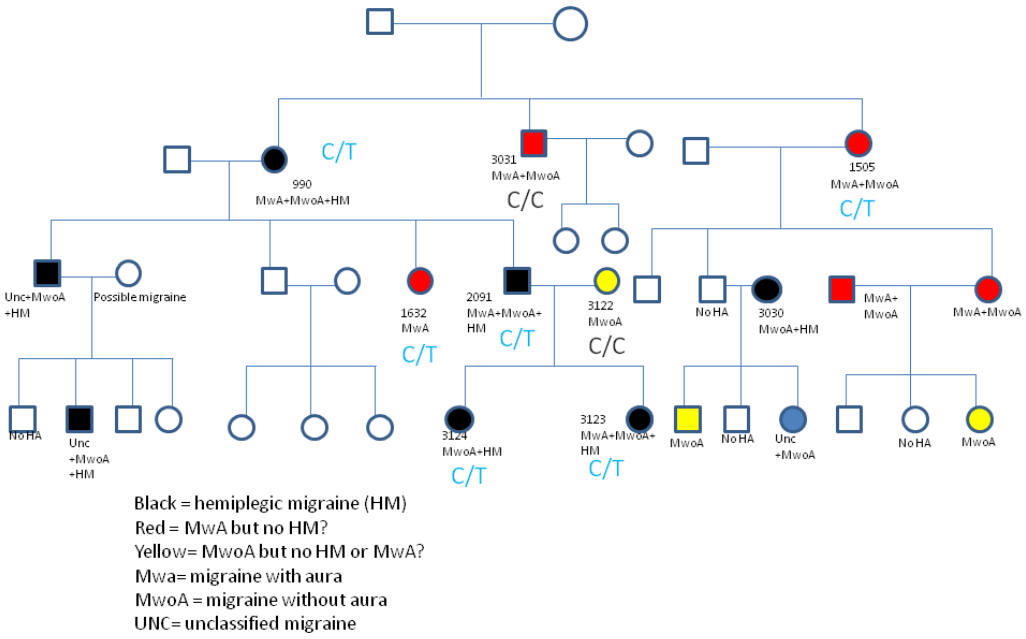

Figure 6.5: **Pedigree of family 2.** Individuals are represented as males (squares), females (circles), FHM cases (blackened symbols), migraine with aura cases (red symbols), migraine without aura cases (yellow symbols) and unclassified (blue symbols). In family 2, patient 2091 (proband), patient 3124 (daughter), patient 3123 (daughter), patient 990 (mother), one brother and one nephew suffered from hemiplegic migraine. Exome sequencing was done for patient 2091 (proband) and patient 990 (mother), since DNA for the daughters, nephew and the brother was initially not available.

6.2.4 Family 2: *EAAT1* mutation

Among the rare functional variants shared by the two affected of each family, we checked whether there were variants in the excitatory amino acid transporter 1 gene (*EAAT1*), since mutations in this gene have been found in a form of episodic ataxia associated with migraine and alternating hemiplegia (EA6) [150]. In one family, family 2 (Figure 6.5), we identified a splice site variant (IVS3-6C>T) in the EAAT1 gene.

Sanger capillary sequencing was performed in other members of family 2 and demonstrated that the variant co-segregated with the FHM phenotype (analysis performed by Mikko Muona). Therefore, it was a plausible causal genetic defect underlying the phenotype. Mutations in cis-acting elements of a gene have been

shown to disrupt the splicing and to be associated with diseases [269]. Splicing is the process whereby the introns in primary messenger RNAs (mRNA) are removed and exons are joined together to produce mature mRNAs. This process requires several cis-acting elements (boundary between exons and introns, branch sites, intronic and exonic enhancers/silencers) located in the intronic and exonic sequences of the gene (Figure 6.6). It has been estimated that a defect in the splicing process is produced by up to 15% of all point mutations underlying genetic diseases [270]. This figure may be an underestimation, since it is based on mutations affecting the known splice sites and our knowledge of the splicing code is still incomplete. Moreover, during the process to identify gene mutations usually intronic sequences are only partially examined, and therefore, mutations in intronic cis-acting splicing regulatory elements may remain undetected [271].

In order to evaluate the functional impact of the identified splice site variant, being *EAAT1* expressed in white blood cells, Mikko Muona tried to detect an aberrantly spliced *EAAT1* transcript in peripheral blood mononuclear cells (PBMCs) obtained from three affected family members. Primer pairs amplifying EAAT1 cDNA between exons 2-5, 3-5, 3-6, 3-7, and 2-7, were designed. The size of the obtained PCR products in the three affected family members and two controls was compared (analysis performed by Mikko Muona). Initially Mikko was able to detect a qualitative difference in EAAT1 splicing between cases and controls. However, these results were not confirmed when additional controls were analyzed. It has to be taken into account that the quantitative effect of the mutation in PBMCs might be relatively mild and it might be difficult to detect with the method used. *EAAT1* (*SLC1A3*) is a member of a family of excitatory amino acid transporters (*EAATs*), which regulate extracellular glutamate concentrations in the central ner-

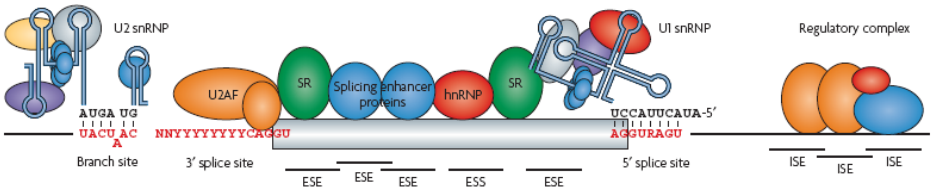

Figure 6.6: **The splicing code** Red indicates consensus splice site sequences at the intronexon boundaries. ESE, exon splicing enhancer. ISE, intron splicing enhancer. Adapted from Wang et al. (2007) [271]

vous system (CNS) [272]. Among the five human *EAATs*, *EAAT1* and *EAAT2*, expressed by astrocytes, have been shown to have a major role on clearance of glutamate released at the synapses during neurotransmission [234–236, 272]. Several lines of evidence support a role of glutamate in the initiation and propagation of cortical spreading depression (CSD), which is considered to be the most likely pathophysiologic mechanism underlying migraine [82, 86, 237, 238]. A reduction in *EAAT1* expression could lead to an accumulation of glutamate in the synaptic cleft and increase in migraine susceptibility.

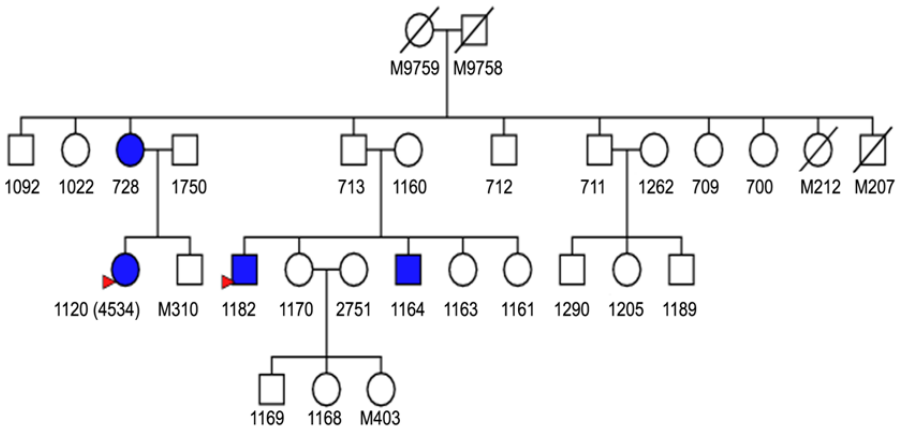

Figure 6.7: **Pedigree of family 3.** Individuals are represented as males (squares), females (circles) and FHM affected (blue symbols). In family 3, patient 1120 (proband), patient 728 (mother), patient 1182 (cousin) and patient 1164 (cousin) suffered from hemiplegic migraine. Exome sequencing was done for patient 1120 (proband) and patient 1182 (cousin). We preferred to sequence the DNA of the cousin compared to the one of the mother to minimize the number of shared variants. Cases for whom exome sequencing was done are marked by an arrow.

6.2.5 Family 3: Integration of linkage analysis and whole exome sequencing

Among the rare functional variants shared by the two affected members of a family, a possible way to narrow down variants of interest is to use linkage analysis data [51]. It has been shown that even modest linkage evidence may help to identify the causative variant [273].

As part of a previous study, members of family 3 had been genotyped for 350 microsatellite markers. Using this data, linkage analysis was performed for the 22 autosome and the X chromosome under the assumption of autosomal dominant inheritance. I found 18 regions with a LOD score greater than zero and, as previously done, we considered these as regions probably containing the causal variant

(Table 6.4) [273]. Of the 10 rare missense variants which were shared by both patients of family 3 (Figure 6.7), five variants were in the genomic regions with a LOD score greater than zero based on linkage analysis. Of these five variants only two were predicted to be damaging to protein function using PolyPhen software analysis. One is a variant in *TBCK* on chromosome 4q24 and the other is a variant in *PITRM1* on chromosome 10p15.2 (Table 6.5).

The two predicted damaging variants will be prioritized for genotyping in other members of the family to evaluate their segregation with the phenotype.

6.3 Discussion

The performed analysis suggests that sequencing of exomes of related individuals with a migraine can lead to the identification of the causative gene (i.e. *CACNA1A*). The *CACNA1A* mutation found in family 1 (Figure 6.3) has already been described in other FHM families. Even if it is a known mutation, it was reassuring to find it among the selected possible causal variants shared by the two affected members of the family, since it confirmed that the used method could lead to narrow the number of possible variants to follow up. According to published data missense mutations in the CACNA1A gene, which encodes for the alpha 1 subunit of a P/Q type voltage-gated calcium channel, account for around half of FHM patients. Among the analyzed 44 families we found CACNA1A mutations only in one family. The discrepancy between the data reported in the literature and our findings maybe due to population differences between previously studied families and our families, which were mainly of Finnish origin.

Table 6.4: **Regions with a LOD score greater than zero.**

Chromosome	Position (bp)	[a]LOD score
1	3484862–48381554	0.88
1	235793755–244166112	0.88
2	11506927–50933915	0.88
2	205127063–241268305	0.88
4	384678–6584947	0.88
4	99334848–117464939	0.88
5	144043648–166976680	0.43
6	144311–20012406	0.69
7	83689381–140112786	0.39
8	2030289–12936149	0.75
9	138616943–140736978	0.88
10	621874–6822254	0.24
12	90723268–116182831	0.26
14	70120123–92302981	0.57
15	27290575–35080297	0.87
16	6172627–31373792	0.24
17	522950–3910723	0.52
20	12484289–43686575	0.85

[a]Highest LOD score in the region
Positions are according to GRCh37/hg19

Table 6.5: **Rare functional variants shared by the two cases of family 2**

Chr	Position (bp)	Reference base	Variant base	Gene	[a]Within linked region	Consequence	[b]Protein prediction
1	43213914	C	T	*LEPRE1*	Yes	Missense	Neutral
4	107171575	C	T	*TBCK*	Yes	Missense	Damaging
6	149893443	C	T	*C6orf72*	No	Missense	Damaging
7	107314634	G	A	*SLC26A4*	Yes	Missense	Neutral
10	3201124	T	A	*PITRM1*	Yes	Missense	Damaging
10	95517902	A	G	*LGI1*	No	Missense	Neutral
11	66833431	A	G	*RHOD*	No	Missense	Neutral
12	123022931	T	C	*KNTC1*	No	Missense	Damaging
16	27788959	A	G	*KIAA0556*	Yes	Missense	Neutral
22	29838017	G	A	*RFPL1*	No	Missense	Neutral

[a]Variant lies within one of the 18 regions with a LOD score greater than zero. The same criteria has been used by Johnson et al. (2010) [273]

[b]Based on the predicted effect of the missense variant on protein function [273]

Positions are according to GRCh37/hg19

In family 2 (Figure 6.5) a splice site mutation in the *EAAT1* gene was found. Mutations in this gene have already been found in a form of episodic ataxia associated with migraine and alternating hemiplegia (EA6) [150]. In the studied samples, it was not possible to clearly determine if the mutation is responsible for the production of aberrantly spliced mRNA. However, it has to be considered that the method used might have been not sensitive enough to detect it or the cell type analyzed (PBMCs) was not the most appropriate. At the same time, the possibility that one of the other variants shared by the two affected subjects is the real underlying cause of the FHM phenotype has to be considered.

A possible way to narrow down possible causal variants, among the ones shared by affected family members is to use linkage analysis data [51]. Johnson et al. have used linkage analysis to identify the causative variant in a family with autosomal dominant amyotrophic lateral sclerosis [273]. Performing exome sequencing in two affected family members 24 possible causative variants were selected, among these four were found to be in regions with a LOD score greater than zero in the linkage analysis and to segregate with the phenotype within the family [273]. The causative role of one of the four was corroborated by finding additional mutation in the same gene in five individuals with amyotrophic lateral sclerosis [273]. Even if the evidence of the utility of the linkage analysis in guiding further analysis of exome sequencing data is still debated, it is still worth exploring. The presented results are preliminary and further follow up of the identified potentially causal variants will be performed.

Although we have sequenced most of the protein-coding exons of the genome, the mutated genes responsible for disease in most of the FHM affected families analyzed have not been identified. This may be due to several factors. Firstly,

distinguishing causative variants from other non causative variants is often difficult, our ability to predict the functional significance of these variants is still very limited. Secondly, although in the studied samples 88.97% of the targeted bases were covered at least ten times, some variants might have been not detected due to regional variation in sequence coverage. Thirdly, in this initial analysis the filters applied for the selection of possible causal variants were quite strict, since variants present in existing databases for which allele frequency information is available were filtered out. When more exome sequencing data from FHM cases and healthy controls will be available, in order to detect not fully penetrant mutations it will be worthwhile to perform case-control association studies. Given the genetic heterogeneity of FHM a larger sample size will be needed to identify causal variants. Fourthly, the exome sequencing, performed in this study, restricts the analysis to a set of exons and splice sites. Therefore, if a causal variant lies in exons not covered by the used exon capture method or in non coding regions, such as regulatory elements, it will not be identified. Moreover, based on the currently available analysis tools, exome sequencing does not allow a good detection of structural variants. Once the costs of the whole genome sequencing reduce sufficiently, it will become the method of choice to identify single nucleotide and structural variants underlying diseases. Fourthly,

Chapter 7

Concluding remarks

In contrast to mendelian disorders, linkage studies had been unsuccessful in identifying risk loci for common diseases, such as migraine [29, 30, 32, 33]. The likely reason is that complex diseases are probably due not only to rare genetic variants, but also to common genetic variants that are incompletely penetrant, 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, which are more difficult and expensive to collect than isolated cases [34].

During the past couple of years substantial advances have been made in understanding the genetic basis of common diseases. The availability of catalogues of common sequence variants, such as HapMap, and the advances in genotyping technologies has made possible to perform GWAS, in which hundreds of thousands 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 that confer moderate susceptibility to common diseases. Since the first GWAS was published, less than ten years ago, hundreds of GWAS have

been completed and have led to the identification of more than 1000 loci associated with common diseases and complex traits [36–41, 44–50].

In contrast with mendelian diseases, it has been found that usually the effect size of each associated variant detected through GWAS was moderate and explained only a small fraction of the phenotypic variance in the population [36, 51–53]. Moreover, GWAS have shown that most common diseases are influenced by a large number of variants [36]. The results of our GWAS, described in Chapters 3 and 4 of this thesis, are in line with the results obtained in other large GWAS, given that we have identified several loci associated with migraine (*AEG1*, *TRPM8*, *LRP1* and *C7orf10*) and variants at these loci have a moderate effect.

By identifying genes associated with common diseases, GWAS are contributing to gain a better understanding of the underlying pathogenesis, which could lead to the development of more specific treatments. In the first GWAS for migraine (Chapter 3), the identification of the association of migraine with astrocyte elevated gene 1 *AEG1*, encoding a protein which inhibits the expression of a glutamate transporter (*EAAT2/GLT1*) in the brain [231], support a role glutamate in the development of migraine [82, 86, 237, 238]. A role of glutamate in the pathogenesis of migraine is also supported by the association of migraine with low density lipoprotein receptor-related protein gene *LRP1* (Chapter 4), since it has been shown that the levels of glutamate receptors are reduced in lrp1 knockout neurons [255]. The evidence supporting a role of glutamate in the development of migraine could have an important therapeutic impact, since glutamate receptors activity can be inhibited by antagonists, such as memantine. The association of migraine with the ion channel *TRPM8* (Chapter 4) suggests that modulators of

neuropathic pain may contribute to the migraine headache. This finding, may also have an important clinical impact since there is evidence suggesting that in vivo antagonism of TRPM8 can reduce neuropathic pain [250].

Even if the results of the GWAS completed so far have improved our understanding of the genetic structure of complex diseases, they often explain only a small proportion of the heritability of common diseases. It is possible that part of the unexplained heritability is due to additional common variants of even smaller effect, which have not been detected yet and they might be discovered increasing the sample size [36]. Other possible sources of the 'missing heritability' are rare variants, copy number variations, gene-gene interactions and perhaps even epigenetic mechanisms [54].

Nowadays, the identification of rare variants underlying disease susceptibility has become possible owing to technological advances in sequencing and bioinformatics approaches [55]. In order to identify rare variants associated with common diseases, the most comprehensive study design will eventually involve sequencing the whole genome in a large number of recruited individuals [51]. Whole-genome sequencing is still too expensive to be applicable to large sample sizes. A strategy which could be pursued, until the cost of whole-genome sequencing drops, is to sequence specific regions, such as the exome [51, 56].

The two kits initially for capturing the exome (NimbleGen Sequence Capture 2.1M Human Exome and Agilent SureSelect Human All Exon Kit) targeted exons from genes in the consensus coding sequence (CCDS) consortium database, in addition to a selection of miRNAs and non-coding RNAs [264]. Although the CCDS database contains a set of consistently annotated protein-coding genes, many genes with solid evidence of transcription and most of the alternative spliced variants

are not included. To address this shortcoming, a more complete set of target regions for the human exome, based on the GENCODE annotation was designed and experimentally tested (Chapter 5) [200]. The extended set (GENCODE exome) covers additional 5594 genes and 10.3 Mb compared with the currently used CCDS-based sets. The additionally covered genes include genes coding for ion channel subunits and protein kinases, which are potential candidate genes for several human diseases.

Two study designs currently used to discover rare variants underlying common diseases are family-based sequencing and extreme-trait sequencing [51]. The first design (family-based sequencing) consists in sequencing cases from families that have multiple affected individuals [51]. The second design (extreme-trait sequencing) consists in sequencing individuals who are at the extreme ends of a phenotype distribution. To identify rare variants contributing to migraine (Chapter 6), whole-exome sequencing of 88 cases from 44 families with multiple individuals affected by familial hemiplegic migraine (FHM) was performed. In family 1, among the rare functional variants shared by the two affected individuals, a missense variant in *CACNA1A* (rs121908212), which had been previously described as causing FHM, was identified [101,102,130,131,134]. In another family, family 2, among the shared functional variants, a splice-site variant in the *EAAT1* gene, which has been previously found mutated in a form of episodic ataxia associated with migraine and alternating hemiplegia (EA6), was detected. However, it was not possible to clearly define the functional effect of this splice-site variant, and therefore, we could not provide any additional supporting evidence that it was the causal variant. In the other 42 families, among the rare functional variants shared (31 on average) by the two affected members of each family, it was difficult to identify

the causal variant. The exome sequencing of several hundreds of FHM cases is currently underway and I hope that obtained data will allow us to identify genes having rare functional variants in more than one family. It has to be considered that studies based on whole exome sequencing are limited by the fact that the analysis is restricted to a set of exons and splice sites. Therefore, if a causal variant lies in exons not covered by the used exon capture method or in non coding regions, such as regulatory elements, it will not be identified. Moreover, the whole exome sequencing does not allow a good detection of structural variants. Once the costs of the whole genome sequencing will reduce, it will become the method of choice to identify single nucleotide and structural variants underlying genetic diseases.

Publications arising from this work

Coffey AJ, Kokocinski F, **Calafato MS**, Scott CE, Palta P, Drury E, Joyce CJ, Leproust EM, Harrow J, Hunt S, Lehesjoki AE, Turner DJ, Hubbard TJ, Palotie A. *The GENCODE exome: sequencing the complete human exome.* Eur J Hum Genet. 2011 Jul;19(7):827-31.

Anttila V, Stefansson H, Kallela M, Todt U, Terwindt GM, **Calafato MS**, Nyholt DR, Dimas AS, Freilinger T, Mller-Myhsok B, Artto V, Inouye M, Alakurtti K, Kaunisto MA, Hamalainen E, de Vries B, Stam AH, Weller CM, Heinze A, Heinze-Kuhn K, Goebel I, Borck G, Gbel H, Steinberg S, Wolf C, Bjrnsson A, Gudmundsson G, Kirchmann M, Hauge A, Werge T, Schoenen J, Eriksson JG, Hagen K, Stovner L, Wichmann HE, Meitinger T, Alexander M, Moebus S, Schreiber S, Aulchenko YS, Breteler MM, Uitterlinden AG, Hofman A, van Duijn CM, Tikka-Kleemola P, Vepslinen S, Lucae S, Tozzi F, Muglia P, Barrett J, Kaprio J, Frkkil M, Peltonen L, Stefansson K, Zwart JA, Ferrari MD, Olesen J, Daly M, Wessman M, van den Maagdenberg AM, Dichgans M, Kubisch C, Dermitzakis ET, Frants RR, Palotie A; International Headache Genetics Consortium. *Genome-wide as-*

sociation study of migraine implicates a common susceptibility variant on 8q22.1. Nat Genet. 2010 Oct;42(10):869-73.

Bibliography

[1] P. D. Turnpenny, *Emery's Elements of Medical Genetics, chapter 2.* Churchill Livingstone, 2007.

[2] L. Kruglyak and D. A. Nickerson, "Variation is the spice of life," *Nat Genet*, vol. 27, no. 3, pp. 234–6, 2001.

[3] International HapMap Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, pp. 931–45, 2004. International Human Genome Sequencing Consortium.

[4] International HapMap Consortium, "A map of human genome variation from population-scale sequencing," *Lancet*, vol. 377, no. 9766, pp. 641–649, 2011. International Parkinson Disease Genomics Consortium.

[5] K. A. Frazer, S. S. Murray, N. J. Schork, and E. J. Topol, "Human genetic variation and its contribution to complex traits," *Nat Rev Genet*, vol. 10, no. 4, pp. 241–51, 2009.

[6] International HapMap Consortium, "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.," *Nature*,

vol. 409, no. 6822, pp. 928–933, 2001. International SNP Map Working Group.

[7] International HapMap Consortium, "The international hapmap project," *Nature*, vol. 426, no. 6968, pp. 789–96, 2003. International HapMap Consortium.

[8] International HapMap Consortium, "A haplotype map of the human genome," *Nature*, vol. 437, no. 7063, pp. 1299–320, 2005. International HapMap Consortium.

[9] International HapMap Consortium, "A second generation human haplotype map of over 3.1 million snps," *Nature*, vol. 449, no. 7164, pp. 851–861, 2007. International HapMap Consortium.

[10] 1000 Genomes Project Consortium, "A map of human genome variation from population-scale sequencing," *Nature*, vol. 467, no. 7319, pp. 1061–73, 2010. 1000 Genomes Project Consortium.

[11] K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu,

M. M. Waye, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F. Olivier, M. S. Phillips, S. Roumy, C. Sallee, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, *et al.*, "A second generation human haplotype map of over 3.1 million snps," *Nature*, vol. 449, no. 7164, pp. 851–61, 2007.

[12] L. Feuk, A. R. Carson, and S. W. Scherer, "Structural variation in the human genome," *Nat Rev Genet*, vol. 7, no. 2, pp. 85–97, 2006.

[13] R. Mills, K. Walter, and . G. Project., "Mapping copy number variation by population-scale genome sequencing," *Nature*, vol. 470, no. 7332, pp. 59–65, 2011.

[14] P. D. Turnpenny, *Emery's Elements of Medical Genetics, chapter 23.* Churchill Livingstone, 2007.

[15] P. Deloukas, G. D. Schuler, G. Gyapay, E. M. Beasley, C. Soderlund, P. Rodriguez-Tome, L. Hui, T. C. Matise, K. B. McKusick, J. S. Beckmann, S. Bentolila, M. Bihoreau, B. B. Birren, J. Browne, A. Butler, A. B. Castle, N. Chiannilkulchai, C. Clee, P. J. Day, A. Dehejia, T. Dibling, N. Drouot, S. Duprat, C. Fizames, S. Fox, S. Gelling, L. Green, P. Harrison, R. Hocking, E. Holloway, S. Hunt, S. Keil, P. Lijnzaad, C. Louis-Dit-Sully, J. Ma, A. Mendis, J. Miller, J. Morissette, D. Muselet, H. C. Nusbaum, A. Peck, S. Rozen, D. Simon, D. K. Slonim, R. Staples, L. D. Stein, E. A. Stew-

art, M. A. Suchard, T. Thangarajah, N. Vega-Czarny, C. Webber, X. Wu, J. Hudson, C. Auffray, N. Nomura, J. M. Sikela, M. H. Polymeropoulos, M. R. James, E. S. Lander, T. J. Hudson, R. M. Myers, D. R. Cox, J. Weissenbach, M. S. Boguski, and D. R. Bentley, "A physical map of 30,000 human genes," *Science*, vol. 282, no. 5389, pp. 744–6, 1998.

[16] M. Olivier, A. Aggarwal, J. Allen, A. A. Almendras, E. S. Bajorek, E. M. Beasley, S. D. Brady, J. M. Bushard, V. I. Bustos, A. Chu, T. R. Chung, A. De Witte, M. E. Denys, R. Dominguez, N. Y. Fang, B. D. Foster, R. W. Freudenberg, D. Hadley, L. R. Hamilton, T. J. Jeffrey, L. Kelly, L. Lazzeroni, M. R. Levy, S. C. Lewis, X. Liu, F. J. Lopez, B. Louie, J. P. Marquis, R. A. Martinez, M. K. Matsuura, N. S. Misherghi, J. A. Norton, A. Olshen, S. M. Perkins, A. J. Perou, C. Piercy, M. Piercy, F. Qin, T. Reif, K. Sheppard, V. Shokoohi, G. A. Smick, W. L. Sun, E. A. Stewart, J. Fernando, Tejeda, N. M. Tran, T. Trejo, N. T. Vo, S. C. Yan, D. L. Zierten, S. Zhao, R. Sachidanandam, B. J. Trask, R. M. Myers, and D. R. Cox, "A high-resolution radiation hybrid map of the human genome draft sequence," *Science*, vol. 291, no. 5507, pp. 1298–302, 2001.

[17] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson,

R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[18] R. M. Myers, J. Stamatoyannopoulos, M. Snyder, I. Dunham, R. C. Hardison, B. E. Bernstein, T. R. Gingeras, W. J. Kent, E. Birney, B. Wold, and G. E. Crawford, "A user's guide to the encyclopedia of dna elements (encode)," *PLoS Biol*, vol. 9, no. 4, p. e1001046, 2011.

[19] International HapMap Consortium, "Integrating common and rare genetic variation in diverse human populations," *Nature*, vol. 467, no. 1038, pp. 52–58, 2010. International HapMap Consortium.

[20] P. R. Burton, M. D. Tobin, and J. L. Hopper, "Key concepts in genetic epidemiology," *Lancet*, vol. 366, no. 9489, pp. 941–51, 2005.

[21] P. D. Turnpenny, *Emery's Elements of Medical Genetics, chapter 9.* Churchill Livingstone, 2007.

[22] D. Clayton, "Prediction and interaction in complex disease genetics: experience in type 1 diabetes.," *PLos Genet.*, vol. 5, no. 7, p. e1000540, 2009.

[23] M. Butler, "Genetics of hypertension. current status.," *J. Med*, vol. 58, no. 3, pp. 175–8, 2010.

[24] S. Ramachandrappa and S. Farooqi, "Genetic approaches to understanding human obesity.," *J. Clin. Invest*, vol. 121, no. 6, pp. 2080–2086, 2011.

[25] E. J. Mulder, C. Van Baal, D. Gaist, M. Kallela, J. Kaprio, D. A. Svensson, D. R. Nyholt, N. G. Martin, A. J. MacGregor, L. F. Cherkas, D. I. Boomsma, and A. Palotie, "Genetic and environmental influences on migraine: a twin study across six countries," *Twin Res*, vol. 6, no. 5, pp. 422–31, 2003.

[26] T. Willemsen, G. nd van Beijsterveldt, C. van Baal, D. Postma, and D. Boomsma, "Heritability of self-reported asthma and allergy: a study in adult dutch twins, siblings and parents.," *Twin Res Hum Genet.*, vol. 11, no. 2, pp. 132–142, 2008.

[27] M. Fischer, B. Mayer, A. Baessler, G. Riegger, J. Erdmann, C. Hengstenberg, and H. Schunkert, "Familial aggregation of left main coronary artery disease and future risk of coronary events in asymptomatic siblings of affected patients.," *Eur Heart J*, vol. 28, no. 20, pp. 2432–7, 2007.

[28] D. Bhugra, "The global prevalence of schizophrenia," *PLoS Med*, vol. 2, no. 5, p. e151, 2005.

[29] S. Menzel, "Genetic and molecular analyses of complex metabolic disorders: genetic linkage," *Ann N Y Acad Sci*, vol. 967, pp. 249–57, 2002.

[30] C. M. Stein and R. C. Elston, "Finding genes underlying human disease," *Clin Genet*, vol. 75, no. 2, pp. 101–6, 2009.

[31] J. H. Barrett, N. A. Sheehan, A. Cox, J. Worthington, C. Cannings, and M. D. Teare, "Family based studies and genetic epidemiology: theory and practice," *Hum Hered*, vol. 64, no. 2, pp. 146–8, 2007.

[32] J. M. Hall, M. K. Lee, B. Newman, J. E. Morrow, L. A. Anderson, B. Huey, and M. C. King, "Linkage of early-onset familial breast cancer to chromosome 17q21," *Science*, vol. 250, no. 4988, pp. 1684–9, 1990.

[33] J. M. Hall, L. Friedman, C. Guenther, M. K. Lee, J. L. Weber, D. M. Black, and M. C. King, "Closing in on a breast cancer gene on chromosome 17q," *Am J Hum Genet*, vol. 50, no. 6, pp. 1235–42, 1992.

[34] R. NJ, "Searching for genetic determinants in the new millennium," *Nature*, vol. 405, no. 6788, pp. 847–56, 2000.

[35] H. J. Cordell and D. G. Clayton, "Genetic association studies," *Lancet*, vol. 366, no. 9491, pp. 1121–31, 2005.

[36] J. N. Hirschhorn and Z. K. Gajdos, "Genome-wide association studies: results from the first few years and potential implications for clinical medicine," *Annu Rev Med*, vol. 62, pp. 11–24, 2011.

[37] J. N. Hirschhorn and M. J. Daly, "Genome-wide association studies for common diseases and complex traits," *Nat Rev Genet*, vol. 6, no. 2, pp. 95–108, 2005.

[38] R. Klein, C. Zeiss, E. Chew, and a. et, "Complement factor h polymorphism in age-related macular degeneration.," *Lancet*, vol. 308, no. 5720, pp. 385–389, 2005.

[39] T. Frayling, N. Timpson, M. Weedon, E. Zeggini, and a. et, "A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity," *Science*, vol. 316, no. 5826, pp. 889–894, 2007.

[40] Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–78, 2007.

[41] J. A. Todd, N. Walker, J. Cooper, D. Smyth, K. Downes, V. Plagnol, and W. T. C. C. Consortium., "Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.," *Nat Gen*, vol. 39, no. 7, pp. 857–864, 2007.

[42] D. Smyth, J. Cooper, R. Bailey, S. Field, O. Burren, L. Smink, C. Guja, C. Ionescu-Tirgoviste, B. Widmer, D. Dunger, D. Savage, N. Walker, D. Clayton, and J. Todd, "A genome-wide association study of nonsynonymous snps identifies a type 1 diabetes locus in the interferon-induced helicase (ifih1) region.," *Nat Gen.*, vol. 38, no. 6, pp. 617–619, 2006.

[43] R. Sladek, G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. Hudson, A. Montpetit, A. Pshezhetsky, M. Prentki, B. Posner, D. Balding, D. Meyre, C. Polychronakos, and P. Froguel, "A genome-wide associa-

tion study identifies novel risk loci for type 2 diabetes.," *Nature*, vol. 445, no. 7130, pp. 881–885, 2007.

[44] K. Bilguvar, K. Yasuno, M. Niemela, Y. M. Ruigrok, M. von Und Zu Fraunberg, C. M. van Duijn, L. H. van den Berg, S. Mane, C. E. Mason, M. Choi, E. Gaal, Y. Bayri, L. Kolb, Z. Arlier, S. Ravuri, A. Ronkainen, A. Tajima, A. Laakso, A. Hata, H. Kasuya, T. Koivisto, J. Rinne, J. Ohman, M. M. Breteler, C. Wijmenga, M. W. State, G. J. Rinkel, J. Hernesniemi, J. E. Jaaskelainen, A. Palotie, I. Inoue, R. P. Lifton, and M. Gunel, "Susceptibility loci for intracranial aneurysm in european and japanese populations," *Nat Genet*, vol. 40, no. 12, pp. 1472–7, 2008.

[45] D. A. Hafler, A. Compston, S. Sawcer, E. S. Lander, M. J. Daly, P. L. De Jager, P. I. de Bakker, S. B. Gabriel, D. B. Mirel, A. J. Ivinson, M. A. Pericak-Vance, S. G. Gregory, J. D. Rioux, J. L. McCauley, J. L. Haines, L. F. Barcellos, B. Cree, J. R. Oksenberg, and S. L. Hauser, "Risk alleles for multiple sclerosis identified by a genomewide study," *N Engl J Med*, vol. 357, no. 9, pp. 851–62, 2007.

[46] J. B. Harley, M. E. Alarcon-Riquelme, L. A. Criswell, C. O. Jacob, R. P. Kimberly, K. L. Moser, B. P. Tsao, T. J. Vyse, C. D. Langefeld, S. K. Nath, J. M. Guthridge, B. L. Cobb, D. B. Mirel, M. C. Marion, A. H. Williams, J. Divers, W. Wang, S. G. Frank, B. Namjou, S. B. Gabriel, A. T. Lee, P. K. Gregersen, T. W. Behrens, K. E. Taylor, M. Fernando, R. Zidovetzki, P. M. Gaffney, J. C. Edberg, J. D. Rioux, J. O. Ojwang, J. A. James, J. T. Merrill, G. S. Gilkeson, M. F. Seldin, H. Yin, E. C. Baechler, Q. Z. Li, E. K. Wakeland, G. R. Bruner, K. M. Kaufman, and J. A. Kelly, "Genome-

wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in itgam, pxk, kiaa1542 and other loci," *Nat Genet*, vol. 40, no. 2, pp. 204–10, 2008.

[47] M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, and J. N. Hirschhorn, "Genome-wide association studies for complex traits: consensus, uncertainty and challenges," *Nat Rev Genet*, vol. 9, no. 5, pp. 356–69, 2008.

[48] J. D. Rioux, R. J. Xavier, K. D. Taylor, M. S. Silverberg, P. Goyette, A. Huett, T. Green, P. Kuballa, M. M. Barmada, L. W. Datta, Y. Y. Shugart, A. M. Griffiths, S. R. Targan, A. F. Ippoliti, E. J. Bernard, L. Mei, D. L. Nicolae, M. Regueiro, L. P. Schumm, A. H. Steinhart, J. I. Rotter, R. H. Duerr, J. H. Cho, M. J. Daly, and S. R. Brant, "Genome-wide association study identifies new susceptibility loci for crohn disease and implicates autophagy in disease pathogenesis," *Nat Genet*, vol. 39, no. 5, pp. 596–604, 2007.

[49] R. Saxena, B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. de Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, M. J. Daly, T. E. Hughes, L. Groop, D. Altshuler, P. Almgren, J. C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H. N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E. K. Speliotes, M. R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia,

M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S. B. Gabriel, G. W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, and S. Purcell, "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels," *Science*, vol. 316, no. 5829, pp. 1331–6, 2007.

[50] R. J. Smith, "Adherence to antiretroviral hiv drugs: how many doses can you miss before resistance emerges?," *Proc Biol Sci*, vol. 273, no. 1586, pp. 617–24, 2006.

[51] E. T. Cirulli and D. B. Goldstein, "Uncovering the roles of rare variants in common disease through whole-genome sequencing," *Nat Rev Genet*, vol. 11, no. 6, pp. 415–25, 2010.

[52] Y. S. Aulchenko, S. Ripatti, I. Lindqvist, D. Boomsma, I. M. Heid, P. P. Pramstaller, B. W. Penninx, A. C. Janssens, J. F. Wilson, T. Spector, N. G. Martin, N. L. Pedersen, K. O. Kyvik, J. Kaprio, A. Hofman, N. B. Freimer, M. R. Jarvelin, U. Gyllensten, H. Campbell, I. Rudan, A. Johansson, F. Marroni, C. Hayward, V. Vitart, I. Jonasson, C. Pattaro, A. Wright, N. Hastie, I. Pichler, A. A. Hicks, M. Falchi, G. Willemsen, J. J. Hottenga, E. J. de Geus, G. W. Montgomery, J. Whitfield, P. Magnusson, J. Saharinen, M. Perola, K. Silander, A. Isaacs, E. J. Sijbrands, A. G. Uitterlinden, J. C. Witteman, B. A. Oostra, P. Elliott, A. Ruokonen, C. Sabatti, C. Gieger, T. Meitinger, F. Kronenberg, A. Doring, H. E. Wichmann, J. H. Smit, M. I. McCarthy, C. M. van Duijn, and L. Peltonen, "Loci influencing lipid levels and coronary heart disease risk in 16 european population cohorts," *Nat Genet*, vol. 41, no. 1, pp. 47–55, 2009.

[53] E. Zeggini, L. J. Scott, R. Saxena, B. F. Voight, J. L. Marchini, T. Hu, P. I. de Bakker, G. R. Abecasis, P. Almgren, G. Andersen, K. Ardlie, K. B. Bostrom, R. N. Bergman, L. L. Bonnycastle, K. Borch-Johnsen, N. P. Burtt, H. Chen, P. S. Chines, M. J. Daly, P. Deodhar, C. J. Ding, A. S. Doney, W. L. Duren, K. S. Elliott, M. R. Erdos, T. M. Frayling, R. M. Freathy, L. Gianniny, H. Grallert, N. Grarup, C. J. Groves, C. Guiducci, T. Hansen, C. Herder, G. A. Hitman, T. E. Hughes, B. Isomaa, A. U. Jackson, T. Jorgensen, A. Kong, K. Kubalanza, F. G. Kuruvilla, J. Kuusisto, C. Langenberg, H. Lango, T. Lauritzen, Y. Li, C. M. Lindgren, V. Lyssenko, A. F. Marvelle, C. Meisinger, K. Midthjell, K. L. Mohlke, M. A. Morken, A. D. Morris, N. Narisu, P. Nilsson, K. R. Owen, C. N. Palmer, F. Payne, J. R. Perry, E. Pettersen, C. Platou, I. Prokopenko, L. Qi, L. Qin, N. W. Rayner, M. Rees, J. J. Roix, A. Sandbaek, B. Shields, M. Sjogren, V. Steinthorsdottir, H. M. Stringham, A. J. Swift, G. Thorleifsson, U. Thorsteinsdottir, N. J. Timpson, T. Tuomi, J. Tuomilehto, M. Walker, R. M. Watanabe, M. N. Weedon, C. J. Willer, T. Illig, K. Hveem, F. B. Hu, M. Laakso, K. Stefansson, O. Pedersen, N. J. Wareham, I. Barroso, A. T. Hattersley, F. S. Collins, L. Groop, M. I. McCarthy, M. Boehnke, and D. Altshuler, "Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes," *Nat Genet*, vol. 40, no. 5, pp. 638–45, 2008.

[54] T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi,

M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher, "Finding the missing heritability of complex diseases," *Nature*, vol. 461, no. 7265, pp. 747–53, 2009.

[55] A. G. Day-Williams and E. Zeggini, "The effect of next-generation sequencing technology on complex trait research," *Eur J Clin Invest*, vol. 41, no. 5, pp. 561–7, 2011.

[56] M. L. Metzker, "Sequencing technologies - the next generation," *Nat Rev Genet*, vol. 11, no. 1, pp. 31–46, 2010.

[57] E. E. Schadt, S. Turner, and A. Kasarskis, "A window into third-generation sequencing," *Hum Mol Genet*, vol. 19, no. R2, pp. R227–40, 2010.

[58] IHS, "The international classification of headache disorders: 2nd edition," *Cephalalgia*, vol. 24 Suppl 1, pp. 9–160, 2004.

[59] A. Ropper and M. Victor, *Adams and Victor's Principles Of Neurology*. Churchill Livingstone, 2007.

[60] L. Kelman, "The premonitory symptoms (prodrome): a tertiary care study of 893 migraineurs," *Headache*, vol. 44, no. 9, pp. 865–72, 2004.

[61] B. K. Rasmussen and J. Olesen, "Migraine with aura and migraine without aura: an epidemiological study," *Cephalalgia*, vol. 12, no. 4, pp. 221–8; discussion 186, 1992.

[62] L. Kelman, "The aura: a tertiary care study of 952 migraine patients," *Cephalalgia*, vol. 24, no. 9, pp. 728–34, 2004.

[63] B. de Vries, R. R. Frants, M. D. Ferrari, and A. M. van den Maagdenberg, "Molecular genetics of migraine," *Hum Genet*, vol. 126, no. 1, pp. 115–32, 2009.

[64] D. R. Nyholt, N. G. Gillespie, A. C. Heath, K. R. Merikangas, D. L. Duffy, and N. G. Martin, "Latent class and genetic analysis does not support migraine with aura and migraine without aura as separate entities," *Genet Epidemiol*, vol. 26, no. 3, pp. 231–44, 2004.

[65] R. A. Purdy, "Migraine with and without aura share the same pathogenic mechanisms," *Neurol Sci*, vol. 29 Suppl 1, pp. S44–6, 2008.

[66] L. Stovner, K. Hagen, R. Jensen, Z. Katsarava, R. Lipton, A. Scher, T. Steiner, and J. A. Zwart, "The global burden of headache: a documentation of headache prevalence and disability worldwide," *Cephalalgia*, vol. 27, no. 3, pp. 193–210, 2007.

[67] R. B. Lipton, M. E. Bigal, M. Diamond, F. Freitag, M. L. Reed, and W. F. Stewart, "Migraine prevalence, disease burden, and the need for preventive therapy," *Neurology*, vol. 68, no. 5, pp. 343–9, 2007.

[68] L. J. Stovner and C. Andree, "Impact of headache in europe: a review for the eurolight project," *J Headache Pain*, vol. 9, no. 3, pp. 139–46, 2008.

[69] L. J. Launer, G. M. Terwindt, and M. D. Ferrari, "The prevalence and characteristics of migraine in a population-based cohort: the gem study," *Neurology*, vol. 53, no. 3, pp. 537–42, 1999.

[70] R. B. Lipton, M. E. Bigal, K. Kolodner, W. F. Stewart, J. N. Liberman, and T. J. Steiner, "The family impact of migraine: population-based studies in the usa and uk," *Cephalalgia*, vol. 23, no. 6, pp. 429–40, 2003.

[71] R. B. Lipton, J. N. Liberman, K. B. Kolodner, M. E. Bigal, A. Dowson, and W. F. Stewart, "Migraine headache disability and health-related quality-of-life: a population-based case-control study from england," *Cephalalgia*, vol. 23, no. 6, pp. 441–50, 2003.

[72] P. Tfelt-Hansen, "Triptans vs other drugs for acute migraine. are there differences in efficacy? a comment," *Headache*, vol. 48, no. 4, pp. 601–5, 2008.

[73] N. M. Ramadan, L. L. Schultz, and S. J. Gilkey, "Migraine prophylactic drugs: proof of efficacy, utilization and cost," *Cephalalgia*, vol. 17, no. 2, pp. 73–80, 1997.

[74] M. M. Bianchin, R. G. Londero, J. E. Lima, and M. E. Bigal, "Migraine and epilepsy: a focus on overlapping clinical, pathophysiological, molecular, and therapeutic aspects," *Curr Pain Headache Rep*, vol. 14, no. 4, pp. 276–83, 2010.

[75] M. E. Bigal, R. B. Lipton, J. Cohen, and S. D. Silberstein, "Epilepsy and migraine," *Epilepsy Behav*, vol. 4 Suppl 2, pp. S13–24, 2003.

[76] Z. Katsarava and C. Weimar, "Migraine and stroke," *J Neurol Sci*, vol. 299, no. 1-2, pp. 42–4, 2010.

[77] M. E. Bigal, T. Kurth, N. Santanello, D. Buse, W. Golden, M. Robbins, and R. B. Lipton, "Migraine and cardiovascular disease: a population-based study," *Neurology*, vol. 74, no. 8, pp. 628–35, 2010.

[78] F. Antonaci, G. Nappi, F. Galli, G. C. Manzoni, P. Calabresi, and A. Costa, "Migraine and psychiatric comorbidity: a review of clinical findings," *J Headache Pain*, vol. 12, no. 2, pp. 115–25, 2011.

[79] M. Lauritzen, "Cortical spreading depression in migraine," *Cephalalgia*, vol. 21, no. 7, pp. 757–60, 2001.

[80] B. Larrosa, J. Pastor, L. Lopez-Aguado, and O. Herreras, "A role for glutamate and glia in the fast network oscillations preceding spreading depression," *Neuroscience*, vol. 141, no. 2, pp. 1057–68, 2006.

[81] R. Marrannes, R. Willems, E. De Prins, and A. Wauquier, "Evidence for a role of the n-methyl-d-aspartate (nmda) receptor in cortical spreading depression in the rat," *Brain Res*, vol. 457, no. 2, pp. 226–40, 1988.

[82] N. Hadjikhani, M. Sanchez Del Rio, O. Wu, D. Schwartz, D. Bakker, B. Fischl, K. K. Kwong, F. M. Cutrer, B. R. Rosen, R. B. Tootell, A. G. Sorensen, and M. A. Moskowitz, "Mechanisms of migraine aura revealed by functional mri in human visual cortex," *Proc Natl Acad Sci U S A*, vol. 98, no. 8, pp. 4687–92, 2001.

[83] M. Denuelle, N. Fabre, P. Payoux, F. Chollet, and G. Geraud, "Posterior cerebral hypoperfusion in migraine without aura," *Cephalalgia*, vol. 28, no. 8, pp. 856–62, 2008.

[84] T. J. Schwedt and D. W. Dodick, "Advanced neuroimaging of migraine," *Lancet Neurol*, vol. 8, no. 6, pp. 560–8, 2009.

[85] T. Sprenger and P. J. Goadsby, "Migraine pathogenesis and state of pharmacological treatment options," *BMC Med*, vol. 7, p. 71, 2009.

[86] H. Bolay, U. Reuter, A. K. Dunn, Z. Huang, D. A. Boas, and M. A. Moskowitz, "Intrinsic brain activity triggers trigeminal meningeal afferents in a migraine model," *Nat Med*, vol. 8, no. 2, pp. 136–42, 2002.

[87] D. Pietrobon and J. Striessnig, "Neurobiology of migraine.," *Nat Rev Neurosci.*, vol. 5, no. 4, pp. 386–398, 2003.

[88] D. Cologno, A. De Pascale, and G. C. Manzoni, "Familial occurrence of migraine with aura in a population-based study," *Headache*, vol. 43, no. 3, pp. 231–4, 2003.

[89] E. A. Schur, C. Noonan, D. Buchwald, J. Goldberg, and N. Afari, "A twin study of depression and migraine: evidence for a shared genetic vulnerability," *Headache*, vol. 49, no. 10, pp. 1493–502, 2009.

[90] C. Lemos, M. J. Castro, J. Barros, J. Sequeiros, J. Pereira-Monteiro, D. Mendonca, and A. Sousa, "Familial clustering of migraine: further evidence from a portuguese study," *Headache*, vol. 49, no. 3, pp. 404–11, 2009.

[91] V. Ulrich, M. Gervil, K. O. Kyvik, J. Olesen, and M. B. Russell, "Evidence of a genetic factor in migraine with aura: a population-based danish twin study," *Ann Neurol*, vol. 45, no. 2, pp. 242–6, 1999.

[92] M. Gervil, V. Ulrich, K. O. Kyvik, J. Olesen, and M. B. Russell, "Migraine without aura: a population-based twin study," *Ann Neurol*, vol. 46, no. 4, pp. 606–11, 1999.

[93] M. L. Honkasalo, J. Kaprio, T. Winter, K. Heikkila, M. Sillanpaa, and M. Koskenvuo, "Migraine and concomitant symptoms among 8167 adult twin pairs," *Headache*, vol. 35, no. 2, pp. 70–8, 1995.

[94] M. De Fusco, R. Marconi, L. Silvestri, L. Atorino, L. Rampoldi, L. Morgante, A. Ballabio, P. Aridon, and G. Casari, "Haploinsufficiency of atp1a2 encoding the na+/k+ pump alpha2 subunit associated with familial hemiplegic migraine type 2," *Nat Genet*, vol. 33, no. 2, pp. 192–6, 2003.

[95] B. de Vries, T. Freilinger, K. R. Vanmolkot, J. B. Koenderink, A. H. Stam, G. M. Terwindt, E. Babini, E. H. van den Boogerd, J. J. van den Heuvel, R. R. Frants, J. Haan, M. Pusch, A. M. van den Maagdenberg, M. D. Ferrari, and M. Dichgans, "Systematic analysis of three fhm genes in 39 sporadic patients with hemiplegic migraine," *Neurology*, vol. 69, no. 23, pp. 2170–6, 2007.

[96] M. Dichgans, T. Freilinger, G. Eckstein, E. Babini, B. Lorenz-Depiereux, S. Biskup, M. D. Ferrari, J. Herzog, A. M. van den Maagdenberg, M. Pusch, and T. M. Strom, "Mutation in the neuronal voltage-gated sodium channel scn1a in familial hemiplegic migraine," *Lancet*, vol. 366, no. 9483, pp. 371–7, 2005.

[97] A. Ducros, A. Joutel, K. Vahedi, M. Cecillon, A. Ferreira, E. Bernard, A. Verier, B. Echenne, A. Lopez de Munain, M. G. Bousser, and E. Tournier-

Lasserve, "Mapping of a second locus for familial hemiplegic migraine to 1q21-q23 and evidence of further heterogeneity," *Ann Neurol*, vol. 42, no. 6, pp. 885–90, 1997.

[98] A. Joutel, M. G. Bousser, V. Biousse, P. Labauge, H. Chabriat, A. Nibbio, J. Maciazek, B. Meyer, M. A. Bach, J. Weissenbach, and et al., "A gene for familial hemiplegic migraine maps to chromosome 19," *Nat Genet*, vol. 5, no. 1, pp. 40–5, 1993.

[99] A. Gallanti, A. Tonelli, V. Cardin, G. Bussone, N. Bresolin, and M. T. Bassi, "A novel de novo nonsense mutation in atp1a2 associated with sporadic hemiplegic migraine and epileptic seizures," *J Neurol Sci*, vol. 273, no. 1-2, pp. 123–6, 2008.

[100] R. Marconi, M. De Fusco, P. Aridon, K. Plewnia, M. Rossi, S. Carapelli, A. Ballabio, L. Morgante, R. Musolino, A. Epifanio, G. Micieli, G. De Michele, and G. Casari, "Familial hemiplegic migraine type 2 is linked to 0.9mb region on chromosome 1q23," *Ann Neurol*, vol. 53, no. 3, pp. 376–81, 2003.

[101] R. A. Ophoff, G. M. Terwindt, M. N. Vergouwe, R. van Eijk, P. J. Oefner, S. M. Hoffman, J. E. Lamerdin, H. W. Mohrenweiser, D. E. Bulman, M. Ferrari, J. Haan, D. Lindhout, G. J. van Ommen, M. H. Hofker, M. D. Ferrari, and R. R. Frants, "Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the ca2+ channel gene cacnl1a4," *Cell*, vol. 87, no. 3, pp. 543–52, 1996.

[102] G. Terwindt, E. Kors, J. Haan, F. Vermeulen, A. Van den Maagdenberg, R. Frants, and M. Ferrari, "Mutation analysis of the cacna1a calcium channel subunit gene in 27 patients with sporadic hemiplegic migraine," *Arch Neurol*, vol. 59, no. 6, pp. 1016–8, 2002.

[103] L. L. Thomsen, E. Oestergaard, A. Bjornsson, H. Stefansson, A. C. Fasquel, J. Gulcher, K. Stefansson, and J. Olesen, "Screen for cacna1a and atp1a2 mutations in sporadic hemiplegic migraine patients," *Cephalalgia*, vol. 28, no. 9, pp. 914–21, 2008.

[104] A. Tonelli, A. Gallanti, A. Bersano, V. Cardin, E. Ballabio, G. Airoldi, F. Redaelli, L. Candelise, N. Bresolin, and M. T. Bassi, "Amino acid changes in the amino terminus of the na,k-adenosine triphosphatase alpha-2 subunit associated to familial and sporadic hemiplegic migraine," *Clin Genet*, vol. 72, no. 6, pp. 517–23, 2007.

[105] M. Wessman, G. M. Terwindt, M. A. Kaunisto, A. Palotie, and R. A. Ophoff, "Migraine: a complex genetic disorder," *Lancet Neurol*, vol. 6, no. 6, pp. 521–32, 2007.

[106] S. Debiais, C. Hommet, I. Bonnaud, M. A. Barthez, S. Rimbaux, F. Riant, and A. Autret, "The fhm1 mutation s218l: a severe clinical phenotype? a case report and review of the literature," *Cephalalgia*, vol. 29, no. 12, pp. 1337–9, 2009.

[107] O. Zhuchenko, J. Bailey, P. Bonnen, T. Ashizawa, D. W. Stockton, C. Amos, W. B. Dobyns, S. H. Subramony, H. Y. Zoghbi, and C. C. Lee, "Autosomal

dominant cerebellar ataxia (sca6) associated with small polyglutamine expansions in the alpha 1a-voltage-dependent calcium channel," *Nat Genet*, vol. 15, no. 1, pp. 62–9, 1997.

[108] F. J. Urbano, M. D. Rosato-Siri, and O. D. Uchitel, "Calcium channels involved in neurotransmitter release at adult, neonatal and p/q-type deficient neuromuscular junctions (review)," *Mol Membr Biol*, vol. 19, no. 4, pp. 293–300, 2002.

[109] R. L. Kraus, M. J. Sinnegger, H. Glossmann, S. Hering, and J. Striessnig, "Familial hemiplegic migraine mutations change alpha1a ca2+ channel kinetics," *J Biol Chem*, vol. 273, no. 10, pp. 5586–90, 1998.

[110] R. L. Kraus, M. J. Sinnegger, A. Koschak, H. Glossmann, S. Stenirri, P. Carrera, and J. Striessnig, "Three new familial hemiplegic migraine mutants affect p/q-type ca(2+) channel kinetics," *J Biol Chem*, vol. 275, no. 13, pp. 9239–43, 2000.

[111] A. Tottene, T. Fellin, S. Pagnutti, S. Luvisetto, J. Striessnig, C. Fletcher, and D. Pietrobon, "Familial hemiplegic migraine mutations increase ca(2+) influx through single human cav2.1 channels and decrease maximal cav2.1 current density in neurons," *Proc Natl Acad Sci U S A*, vol. 99, no. 20, pp. 13284–9, 2002.

[112] A. van den Maagdenberg, D. Pietrobon, T. Pizzorusso, S. Kaja, L. Broos, T. Cesetti, R. van de Ven, A. Tottene, J. van der Kaa, J. Plomp, R. Frants, and M. Ferrari, "A cacna1a knockin migraine mouse model with increased

susceptibility to cortical spreading depression," *Neuron*, vol. 41, no. 5, pp. 701–10, 2004.

[113] L. Claes, J. Del-Favero, B. Ceulemans, L. Lagae, C. Van Broeckhoven, and P. De Jonghe, "De novo mutations in the sodium-channel gene scn1a cause severe myoclonic epilepsy of infancy," *Am J Hum Genet*, vol. 68, no. 6, pp. 1327–32, 2001.

[114] A. Escayg, B. T. MacDonald, M. H. Meisler, S. Baulac, G. Huberfeld, I. An-Gourfinkel, A. Brice, E. LeGuern, B. Moulard, D. Chaigne, C. Buresi, and A. Malafosse, "Mutations of scn1a, encoding a neuronal sodium channel, in two families with gefs+2," *Nat Genet*, vol. 24, no. 4, pp. 343–5, 2000.

[115] C. E. Stafstrom, "Severe epilepsy syndromes of early childhood: the link between genetics and pathophysiology with a focus on scn1a mutations," *J Child Neurol*, vol. 24, no. 8 Suppl, pp. 15S–23S, 2009.

[116] R. H. Wallace, I. E. Scheffer, S. Barnett, M. Richards, L. Dibbens, R. R. Desai, T. Lerman-Sagie, D. Lev, A. Mazarib, N. Brand, B. Ben-Zeev, I. Goikhman, R. Singh, G. Kremmidiotis, A. Gardner, G. R. Sutherland, J. George, A. L., J. C. Mulley, and S. F. Berkovic, "Neuronal sodium-channel alpha1-subunit mutations in generalized epilepsy with febrile seizures plus," *Am J Hum Genet*, vol. 68, no. 4, pp. 859–65, 2001.

[117] M. J. Castro, B. Nunes, B. de Vries, C. Lemos, K. R. Vanmolkot, J. J. van den Heuvel, T. Temudo, J. Barros, J. Sequeiros, R. R. Frants, J. B. Koenderink, J. M. Pereira-Monteiro, and A. M. van den Maagdenberg, "Two novel functional mutations in the na+,k+-atpase alpha2-subunit atp1a2 gene

in patients with familial hemiplegic migraine and associated neurological phenotypes," *Clin Genet*, vol. 73, no. 1, pp. 37–43, 2008.

[118] K. M. Kahlig, T. H. Rhodes, M. Pusch, T. Freilinger, J. M. Pereira-Monteiro, M. D. Ferrari, A. M. van den Maagdenberg, M. Dichgans, and J. George, A. L., "Divergent sodium channel defects in familial hemiplegic migraine," *Proc Natl Acad Sci U S A*, vol. 105, no. 28, pp. 9799–804, 2008.

[119] T. H. Rhodes, C. Lossin, C. G. Vanoye, D. W. Wang, and J. George, A. L., "Noninactivating voltage-gated sodium channels in severe myoclonic epilepsy of infancy," *Proc Natl Acad Sci U S A*, vol. 101, no. 30, pp. 11147–52, 2004.

[120] M. Dobretsov and J. R. Stimers, "Neuronal function and alpha3 isoform of the na/k-atpase," *Front Biosci*, vol. 10, pp. 2373–96, 2005.

[121] J. Cressman, J. R., G. Ullah, J. Ziburkus, S. J. Schiff, and E. Barreto, "The influence of sodium and potassium dynamics on excitability, seizures, and the stability of persistent states: I. single neuron dynamics," *J Comput Neurosci*, vol. 26, no. 2, pp. 159–70, 2009.

[122] M. A. Kaunisto, H. Harno, K. R. Vanmolkot, J. J. Gargus, G. Sun, E. Hamalainen, E. Liukkonen, M. Kallela, A. M. van den Maagdenberg, R. R. Frants, M. Farkkila, A. Palotie, and M. Wessman, "A novel missense atp1a2 mutation in a finnish family with familial hemiplegic migraine type 2," *Neurogenetics*, vol. 5, no. 2, pp. 141–6, 2004.

[123] A. Lebas, L. Guyant-Marechal, D. Hannequin, F. Riant, E. Tournier-Lasserve, and D. Parain, "Severe attacks of familial hemiplegic migraine,

childhood epilepsy and atp1a2 mutation," *Cephalalgia*, vol. 28, no. 7, pp. 774–7, 2008.

[124] L. L. Thomsen, M. Kirchmann, A. Bjornsson, H. Stefansson, R. M. Jensen, A. C. Fasquel, H. Petursson, M. Stefansson, M. L. Frigge, A. Kong, J. Gulcher, K. Stefansson, and J. Olesen, "The genetic spectrum of a population-based sample of familial hemiplegic migraine," *Brain*, vol. 130, no. Pt 2, pp. 346–56, 2007.

[125] N. N. Tavraz, T. Friedrich, K. L. Durr, J. B. Koenderink, E. Bamberg, T. Freilinger, and M. Dichgans, "Diverse functional consequences of mutations in the na+/k+-atpase alpha2-subunit causing familial hemiplegic migraine type 2," *J Biol Chem*, vol. 283, no. 45, pp. 31097–106, 2008.

[126] M. B. Russell and A. Ducros, "Sporadic and familial hemiplegic migraine: pathophysiological mechanisms, clinical characteristics, diagnosis, and management," *Lancet Neurol*, vol. 10, no. 5, pp. 457–70, 2011.

[127] G. M. Terwindt, R. A. Ophoff, J. Haan, R. R. Frants, and M. D. Ferrari, "Familial hemiplegic migraine: a clinical comparison of families linked and unlinked to chromosome 19.dmg rg," *Cephalalgia*, vol. 16, no. 3, pp. 153–5, 1996.

[128] G. M. Terwindt, R. A. Ophoff, J. Haan, M. N. Vergouwe, R. van Eijk, R. R. Frants, and M. D. Ferrari, "Variable clinical expression of mutations in the p/q-type calcium channel gene in familial hemiplegic migraine. dutch migraine genetics research group," *Neurology*, vol. 50, no. 4, pp. 1105–10, 1998.

[129] A. Ducros, C. Denier, A. Joutel, M. Cecillon, C. Lescoat, K. Vahedi, F. Darcel, E. Vicaut, M. G. Bousser, and E. Tournier-Lasserve, "The clinical spectrum of familial hemiplegic migraine associated with mutations in a neuronal calcium channel," *N Engl J Med*, vol. 345, no. 1, pp. 17–24, 2001.

[130] A. Ducros, C. Denier, A. Joutel, K. Vahedi, A. Michel, F. Darcel, M. Madigand, D. Guerouaou, F. Tison, J. Julien, E. Hirsch, F. Chedru, C. Bisgard, G. Lucotte, P. Despres, C. Billard, M. A. Barthez, G. Ponsot, M. G. Bousser, and E. Tournier-Lasserve, "Recurrence of the t666m calcium channel cacna1a gene mutation in familial hemiplegic migraine with progressive cerebellar ataxia," *Am J Hum Genet*, vol. 64, no. 1, pp. 89–98, 1999.

[131] K. L. Friend, D. Crimmins, T. G. Phan, C. M. Sue, A. Colley, V. S. Fung, J. G. Morris, G. R. Sutherland, and R. I. Richards, "Detection of a novel missense mutation and second recurrent mutation in the cacna1a gene in individuals with ea-2 and fhm," *Hum Genet*, vol. 105, no. 3, pp. 261–5, 1999.

[132] E. E. Kors, G. M. Terwindt, F. L. Vermeulen, R. B. Fitzsimons, P. E. Jardine, P. Heywood, S. Love, A. M. van den Maagdenberg, J. Haan, R. R. Frants, and M. D. Ferrari, "Delayed cerebral edema and fatal coma after minor head trauma: role of the cacna1a calcium channel subunit gene and relationship with familial hemiplegic migraine," *Ann Neurol*, vol. 49, no. 6, pp. 753–60, 2001.

[133] I. Alonso, J. Barros, A. Tuna, J. Coelho, J. Sequeiros, I. Silveira, and P. Coutinho, "Phenotypes of spinocerebellar ataxia type 6 and familial hemi-

plegic migraine caused by a unique cacna1a missense mutation in patients from a large family," *Arch Neurol*, vol. 60, no. 4, pp. 610–4, 2003.

[134] E. E. Kors, J. Haan, N. J. Giffin, L. Pazdera, C. Schnittger, G. G. Lennox, G. M. Terwindt, F. L. Vermeulen, A. M. Van den Maagdenberg, R. R. Frants, and M. D. Ferrari, "Expanding the phenotypic spectrum of the cacna1a gene t666m mutation: a description of 5 families with familial hemiplegic migraine," *Arch Neurol*, vol. 60, no. 5, pp. 684–8, 2003.

[135] T. Takahashi, S. Igarashi, T. Kimura, I. Hozumi, I. Kawachi, O. Onodera, H. Takano, M. Saito, and S. Tsuji, "Japanese cases of familial hemiplegic migraine with cerebellar ataxia carrying a t666m mutation in the cacna1a gene," *J Neurol Neurosurg Psychiatry*, vol. 72, no. 5, pp. 676–7, 2002.

[136] T. Wada, N. Kobayashi, Y. Takahashi, T. Aoki, T. Watanabe, and S. Saitoh, "Wide clinical variability in a family with a cacna1a t666m mutation: hemiplegic migraine, coma, and progressive ataxia," *Pediatr Neurol*, vol. 26, no. 1, pp. 47–50, 2002.

[137] K. R. Vanmolkot, H. Stroink, J. B. Koenderink, E. E. Kors, J. J. van den Heuvel, E. H. van den Boogerd, A. H. Stam, J. Haan, B. B. De Vries, G. M. Terwindt, R. R. Frants, M. D. Ferrari, and A. M. van den Maagdenberg, "Severe episodic neurological deficits and permanent mental retardation in a child with a novel fhm2 atp1a2 mutation," *Ann Neurol*, vol. 59, no. 2, pp. 310–4, 2006.

[138] S. Marti, R. W. Baloh, J. C. Jen, D. Straumann, and H. H. Jung, "Progressive cerebellar ataxia with variable episodic symptoms–phenotypic diversity of r1668w cacna1a mutation," *Eur Neurol*, vol. 60, no. 1, pp. 16–20, 2008.

[139] J. C. Jen, G. W. Kim, K. A. Dudding, and R. W. Baloh, "No mutations in cacna1a and atp1a2 in probands with common types of migraine," *Arch Neurol*, vol. 61, no. 6, pp. 926–8, 2004.

[140] M. Dichgans, J. Herzog, T. Freilinger, M. Wilke, and D. P. Auer, "1h-mrs alterations in the cerebellum of patients with familial hemiplegic migraine type 1," *Neurology*, vol. 64, no. 4, pp. 608–13, 2005.

[141] I. Yabe, M. Kitagawa, Y. Suzuki, K. Fujiwara, T. Wada, T. Tsubuku, N. Takeichi, K. Sakushima, H. Soma, S. Tsuji, M. Niino, S. Saitoh, and H. Sasaki, "Downbeat positioning nystagmus is a common clinical feature despite variable phenotypes in an fhm1 family," *J Neurol*, vol. 255, no. 10, pp. 1541–4, 2008.

[142] T. Freilinger, M. Bohe, B. Wegener, B. Muller-Myhsok, M. Dichgans, and H. Knoblauch, "Expansion of the phenotypic spectrum of the cacna1a t666m mutation: a family with familial hemiplegic migraine type 1, cerebellar atrophy and mental retardation," *Cephalalgia*, vol. 28, no. 4, pp. 403–7, 2008.

[143] I. Alonso, J. Barros, A. Tuna, A. Seixas, P. Coutinho, J. Sequeiros, and I. Silveira, "A novel r1347q mutation in the predicted voltage sensor segment of the p/q-type calcium-channel alpha-subunit in a family with progressive cerebellar ataxia and hemiplegic migraine," *Clin Genet*, vol. 65, no. 1, pp. 70–2, 2004.

[144] A. H. Stam, K. R. Vanmolkot, H. P. Kremer, J. Gartner, J. Brown, E. Leshinsky-Silver, R. Gilad, E. E. Kors, W. S. Frankhuizen, H. B. Ginjaar, J. Haan, R. R. Frants, M. D. Ferrari, A. M. van den Maagdenberg, and G. M. Terwindt, "Cacna1a r1347q: a frequent recurrent mutation in hemiplegic migraine," *Clin Genet*, vol. 74, no. 5, pp. 481–5, 2008.

[145] E. E. Kors, A. Melberg, K. R. Vanmolkot, E. Kumlien, J. Haan, R. Raininko, R. Flink, H. B. Ginjaar, R. R. Frants, M. D. Ferrari, and A. M. van den Maagdenberg, "Childhood epilepsy, familial hemiplegic migraine, cerebellar ataxia, and a new cacna1a mutation," *Neurology*, vol. 63, no. 6, pp. 1136–7, 2004.

[146] T. Freilinger, N. Ackl, A. Ebert, C. Schmidt, B. Rautenstrauss, M. Dichgans, and A. Danek, "A novel mutation in cacna1a associated with hemiplegic migraine, cerebellar dysfunction and late-onset cognitive decline," *J Neurol Sci*, vol. 300, no. 1-2, pp. 160–3, 2011.

[147] K. R. Vanmolkot, E. E. Kors, J. J. Hottenga, G. M. Terwindt, J. Haan, W. A. Hoefnagels, D. F. Black, L. A. Sandkuijl, R. R. Frants, M. D. Ferrari, and A. M. van den Maagdenberg, "Novel mutations in the na+, k+-atpase pump gene atp1a2 associated with familial hemiplegic migraine and benign familial infantile convulsions," *Ann Neurol*, vol. 54, no. 3, pp. 360–6, 2003.

[148] K. J. Swoboda, E. Kanavakis, A. Xaidara, J. E. Johnson, M. F. Leppert, M. B. Schlesinger-Massart, L. J. Ptacek, K. Silver, and S. Youroukos, "Alternating hemiplegia of childhood or familial hemiplegic migraine? a novel atp1a2 mutation," *Ann Neurol*, vol. 55, no. 6, pp. 884–7, 2004.

[149] J. P. Dreier, K. Jurkat-Rott, G. C. Petzold, O. Tomkins, R. Klingebiel, U. A. Kopp, F. Lehmann-Horn, A. Friedman, and M. Dichgans, "Opening of the blood-brain barrier preceding cortical edema in a severe attack of fhm type ii," *Neurology*, vol. 64, no. 12, pp. 2145–7, 2005.

[150] J. C. Jen, A. Klein, E. Boltshauser, M. S. Cartwright, E. S. Roach, H. Mamsa, and R. W. Baloh, "Prolonged hemiplegic episodes in children due to mutations in atp1a2," *J Neurol Neurosurg Psychiatry*, vol. 78, no. 5, pp. 523–6, 2007.

[151] L. Deprez, S. Weckhuysen, K. Peeters, T. Deconinck, K. G. Claeys, L. R. Claes, A. Suls, T. Van Dyck, A. Palmini, G. Matthijs, W. Van Paesschen, and P. De Jonghe, "Epilepsy as part of the phenotype associated with atp1a2 mutations," *Epilepsia*, vol. 49, no. 3, pp. 500–8, 2008.

[152] D. M. Fernandez, C. K. Hand, B. J. Sweeney, and N. A. Parfrey, "A novel atp1a2 gene mutation in an irish familial hemiplegic migraine kindred," *Headache*, vol. 48, no. 1, pp. 101–8, 2008.

[153] B. de Vries, A. H. Stam, M. Kirkpatrick, K. R. Vanmolkot, J. B. Koenderink, J. J. van den Heuvel, B. Stunnenberg, D. Goudie, J. Shetty, V. Jain, J. van Vark, G. M. Terwindt, R. R. Frants, J. Haan, A. M. van den Maagdenberg, and M. D. Ferrari, "Familial hemiplegic migraine is associated with febrile seizures in an fhm2 family with a novel de novo atp1a2 mutation," *Epilepsia*, vol. 50, no. 11, pp. 2503–4, 2009.

[154] F. Riant, A. Ducros, C. Ploton, C. Barbance, C. Depienne, and E. Tournier-Lasserve, "De novo mutations in atp1a2 and cacna1a are frequent in early-

onset sporadic hemiplegic migraine," *Neurology*, vol. 75, no. 11, pp. 967–72, 2010.

[155] K. R. Vanmolkot, E. Babini, B. de Vries, A. H. Stam, T. Freilinger, G. M. Terwindt, L. Norris, J. Haan, R. R. Frants, N. M. Ramadan, M. D. Ferrari, M. Pusch, A. M. van den Maagdenberg, and M. Dichgans, "The novel p.l1649q mutation in the scn1a epilepsy gene is associated with familial hemiplegic migraine: genetic and functional studies. mutation in brief 957. online," *Hum Mutat*, vol. 28, no. 5, p. 522, 2007.

[156] M. J. Castro, A. H. Stam, C. Lemos, B. de Vries, K. R. Vanmolkot, J. Barros, G. M. Terwindt, R. R. Frants, J. Sequeiros, M. D. Ferrari, J. M. Pereira-Monteiro, and A. M. van den Maagdenberg, "First mutation in the voltage-gated nav1.1 subunit gene scn1a with co-occurring familial hemiplegic migraine and epilepsy," *Cephalalgia*, vol. 29, no. 3, pp. 308–13, 2009.

[157] K. Vahedi, C. Depienne, D. Le Fort, F. Riant, P. Chaine, O. Trouillard, A. Gaudric, M. A. Morris, E. Leguern, E. Tournier-Lasserve, and M. G. Bousser, "Elicited repetitive daily blindness: a new phenotype associated with hemiplegic migraine and scn1a mutations," *Neurology*, vol. 72, no. 13, pp. 1178–83, 2009.

[158] R. W. Labrum, S. Rajakulendran, T. D. Graves, L. H. Eunson, R. Bevan, M. G. Sweeney, S. R. Hammans, N. Tubridy, T. Britton, L. J. Carr, J. R. Ostergaard, C. R. Kennedy, A. Al-Memar, D. M. Kullmann, S. Schorge, K. Temple, M. B. Davis, and M. G. Hanna, "Large scale calcium channel gene

rearrangements in episodic ataxia and hemiplegic migraine: implications for diagnostic testing," *J Med Genet*, vol. 46, no. 11, pp. 786–91, 2009.

[159] L. Veneziano, S. Guida, E. Mantuano, P. Bernard, P. Tarantino, L. Boccone, F. M. Hisama, P. Carrera, C. Jodice, and M. Frontali, "Newly characterised 5' and 3' regions of cacna1a gene harbour mutations associated with familial hemiplegic migraine and episodic ataxia," *J Neurol Sci*, vol. 276, no. 1-2, pp. 31–7, 2009.

[160] P. Bradshaw and M. Parsons, "Hemiplegic migraine, a clinical study," *Q J Med*, vol. 34, pp. 65–85, 1965.

[161] K. Jurkat-Rott, T. Freilinger, J. P. Dreier, J. Herzog, H. Gobel, G. C. Petzold, P. Montagna, T. Gasser, F. Lehmann-Horn, and M. Dichgans, "Variability of familial hemiplegic migraine with novel a1a2 na+/k+-atpase variants," *Neurology*, vol. 62, no. 10, pp. 1857–61, 2004.

[162] V. Anttila, M. Kallela, G. Oswell, M. Kaunisto, D. Nyholt, E. Hamalainen, H. Havanka, M. Wessman, A. Palotie, and a. et, "Trait components provide tools to dissect the genetic susceptibility of migraine," *Am J Hum Genet*, vol. 79, no. 1, pp. 85–99, 2006.

[163] V. Anttila, D. R. Nyholt, M. Kallela, V. Artto, S. Vepsalainen, E. Jakkula, A. Wennerstrom, P. Tikka-Kleemola, M. A. Kaunisto, E. Hamalainen, E. Widen, J. Terwilliger, K. Merikangas, G. W. Montgomery, N. G. Martin, M. Daly, J. Kaprio, L. Peltonen, M. Farkkila, M. Wessman, and A. Palotie, "Consistently replicating locus linked to migraine on 10q22-q23," *Am J Hum Genet*, vol. 82, no. 5, pp. 1051–63, 2008.

[164] A. Bjornsson, G. Gudmundsson, E. Gudfinnsson, M. Hrafnsdottir, J. Benedikz, S. Skuladottir, K. Kristjansson, M. L. Frigge, A. Kong, K. Stefansson, and J. R. Gulcher, "Localization of a gene for migraine without aura to chromosome 4q21," *Am J Hum Genet*, vol. 73, no. 5, pp. 986–93, 2003.

[165] A. Carlsson, L. Forsgren, P. O. Nylander, U. Hellman, K. Forsman-Semb, G. Holmgren, D. Holmberg, and M. Holmberg, "Identification of a susceptibility locus for migraine with and without aura on 6p12.2-p21.1," *Neurology*, vol. 59, no. 11, pp. 1804–7, 2002.

[166] R. A. Lea, A. G. Shepherd, R. P. Curtain, D. R. Nyholt, S. Quinlan, P. J. Brimage, and L. R. Griffiths, "A typical migraine susceptibility region localizes to chromosome 1q31," *Neurogenetics*, vol. 4, no. 1, pp. 17–22, 2002.

[167] R. A. Lea, D. R. Nyholt, R. P. Curtain, M. Ovcaric, R. Sciascia, C. Bellis, J. Macmillan, S. Quinlan, R. A. Gibson, L. C. McCarthy, J. H. Riley, Y. J. Smithies, S. Kinrade, and L. R. Griffiths, "A genome-wide scan provides evidence for loci influencing a severe heritable form of common migraine," *Neurogenetics*, vol. 6, no. 2, pp. 67–72, 2005.

[168] D. R. Nyholt, J. L. Dawkins, P. J. Brimage, P. J. Goadsby, G. A. Nicholson, and L. R. Griffiths, "Evidence for an x-linked genetic component in familial typical migraine," *Hum Mol Genet*, vol. 7, no. 3, pp. 459–63, 1998.

[169] D. R. Nyholt, R. A. Lea, P. J. Goadsby, P. J. Brimage, and L. R. Griffiths, "Familial typical migraine: linkage to chromosome 19p13 and evidence for genetic heterogeneity," *Neurology*, vol. 50, no. 5, pp. 1428–32, 1998.

[170] D. R. Nyholt, K. I. Morley, M. A. Ferreira, S. E. Medland, D. I. Boomsma, A. C. Heath, K. R. Merikangas, G. W. Montgomery, and N. G. Martin, "Genomewide significant linkage to migrainous headache on chromosome 5q21," *Am J Hum Genet*, vol. 77, no. 3, pp. 500–12, 2005.

[171] M. B. Russell, L. Iselius, and J. Olesen, "Inheritance of migraine investigated by complex segregation analysis," *Hum Genet*, vol. 96, no. 6, pp. 726–30, 1995.

[172] M. Wessman, M. Kallela, M. A. Kaunisto, P. Marttila, E. Sobel, J. Hartiala, G. Oswell, S. M. Leal, J. C. Papp, E. Hamalainen, P. Broas, G. Joslyn, I. Hovatta, T. Hiekkalinna, J. Kaprio, J. Ott, R. M. Cantor, J. A. Zwart, M. Ilmavirta, H. Havanka, M. Farkkila, L. Peltonen, and A. Palotie, "A susceptibility locus for migraine with aura, on chromosome 4q24," *Am J Hum Genet*, vol. 70, no. 3, pp. 652–62, 2002.

[173] N. J. Colson, R. A. Lea, S. Quinlan, J. MacMillan, and L. R. Griffiths, "The estrogen receptor 1 g594a polymorphism is associated with migraine susceptibility in two independent case/control groups," *Neurogenetics*, vol. 5, no. 2, pp. 129–33, 2004.

[174] N. J. Colson, R. A. Lea, S. Quinlan, and L. R. Griffiths, "No role for estrogen receptor 1 gene intron 1 pvu ii and exon 4 c325g polymorphisms in migraine susceptibility," *BMC Med Genet*, vol. 7, p. 12, 2006.

[175] R. Curtain, R. A. Lea, S. Quinlan, C. Bellis, L. Tajouri, R. Hughes, J. Macmillan, and L. R. Griffiths, "Investigation of the low-density lipopro-

tein receptor gene and cholesterol as a risk factor for migraine," *J Neurol Sci*, vol. 227, no. 1, pp. 95–100, 2004.

[176] M. A. Kaunisto, M. Kallela, E. Hamalainen, R. Kilpikari, H. Havanka, H. Harno, M. Nissila, E. Sako, M. Ilmavirta, J. Liukkonen, H. Teirmaa, O. Tornwall, M. Jussila, J. Terwilliger, M. Farkkila, J. Kaprio, A. Palotie, and M. Wessman, "Testing of variants of the mthfr and esr1 genes in 1798 finnish individuals fails to confirm the association with migraine with aura," *Cephalalgia*, vol. 26, no. 12, pp. 1462–72, 2006.

[177] R. A. Lea, M. Ovcaric, J. Sundholm, J. MacMillan, and L. R. Griffiths, "The methylenetetrahydrofolate reductase gene variant c677t influences susceptibility to migraine with aura," *BMC Med*, vol. 2, p. 3, 2004.

[178] L. C. McCarthy, D. A. Hosford, J. H. Riley, M. I. Bird, N. J. White, D. R. Hewett, S. J. Peroutka, L. R. Griffiths, P. R. Boyd, R. A. Lea, S. M. Bhatti, L. K. Hosking, C. M. Hood, K. W. Jones, A. R. Handley, R. Rallan, K. F. Lewis, A. J. Yeo, P. M. Williams, R. C. Priest, P. Khan, C. Donnelly, S. M. Lumsden, J. O'Sullivan, C. G. See, D. H. Smart, S. Shaw-Hawkins, J. Patel, T. C. Langrish, W. Feniuk, R. G. Knowles, M. Thomas, V. Libri, D. S. Montgomery, P. K. Manasco, C. F. Xu, C. Dykes, P. P. Humphrey, A. D. Roses, and I. J. Purvis, "Single-nucleotide polymorphism alleles in the insulin receptor gene are associated with typical migraine," *Genomics*, vol. 78, no. 3, pp. 135–49, 2001.

[179] M. Mochi, S. Cevoli, P. Cortelli, G. Pierangeli, C. Scapoli, S. Soriani, and P. Montagna, "Investigation of an ldlr gene polymorphism (19p13.2) in sus-

ceptibility to migraine without aura," *J Neurol Sci*, vol. 213, no. 1-2, pp. 7–10, 2003.

[180] C. Netzer, J. Freudenberg, A. Heinze, K. Heinze-Kuhn, I. Goebel, L. C. McCarthy, A. D. Roses, H. Gobel, U. Todt, and C. Kubisch, "Replication study of the insulin receptor gene in migraine with aura," *Genomics*, vol. 91, no. 6, pp. 503–7, 2008.

[181] A. Oterino, N. Valle, Y. Bravo, P. Munoz, P. Sanchez-Velasco, C. Ruiz-Alegria, J. Castillo, F. Leyva-Cobian, A. Vadillo, and J. Pascual, "Mthfr t677 homozygosis influences the presence of aura in migraineurs," *Cephalalgia*, vol. 24, no. 6, pp. 491–4, 2004.

[182] A. Oterino, J. Pascual, C. Ruiz de Alegria, N. Valle, J. Castillo, Y. Bravo, F. Gonzalez, P. Sanchez-Velasco, A. Cayon, F. Leyva-Cobian, A. Alonso-Arranz, and P. Munoz, "Association of migraine and esr1 g325c polymorphism," *Neuroreport*, vol. 17, no. 1, pp. 61–4, 2006.

[183] E. Rubino, M. Ferrero, I. Rainero, E. Binello, G. Vaula, and L. Pinessi, "Association of the c677t polymorphism in the mthfr gene with migraine: a meta-analysis," *Cephalalgia*, vol. 29, no. 8, pp. 818–25, 2009.

[184] A. I. Scher, G. M. Terwindt, W. M. Verschuren, M. C. Kruit, H. J. Blom, H. Kowa, R. R. Frants, A. M. van den Maagdenberg, M. van Buchem, M. D. Ferrari, and L. J. Launer, "Migraine and mthfr c677t genotype in a population-based sample," *Ann Neurol*, vol. 59, no. 2, pp. 372–5, 2006.

[185] Z. Cader, S. Noble-Topham, D. Dyment, S. Cherny, J. Brown, G. Rice, and G. Ebers, "Significant linkage to migraine with aura on chromosome 11q24," *Hum Mol Genet*, vol. 12, no. 19, pp. 2511–7, 2003.

[186] D. Soragna, A. Vettori, G. Carraro, E. Marchioni, G. Vazza, S. Bellini, R. Tupler, F. Savoldi, and M. L. Mostacciuolo, "A locus for migraine without aura maps on chromosome 14q21.2-q22.3," *Am J Hum Genet*, vol. 72, no. 1, pp. 161–7, 2003.

[187] N. Risch and K. Merikangas, "The future of genetic studies of complex human diseases," *Science*, vol. 273, no. 5281, pp. 1516–7, 1996.

[188] Colson, Lea, Quinlan., and Griffiths, "No role for estrogen receptor 1 gene intron 1 pvu ii and exon 4 c325g polymorphisms in migraine susceptibility," *BMC Med Genet.*, vol. 1471, no. 12, pp. 7–12, 2006.

[189] N. Colson, R. Lea, S. Quinlan, J. MacMillan, and L. Griffiths, "Investigation of hormone receptor genes in migraine," *Neurogenetics*, vol. 6, no. 1, pp. 17–23, 2005.

[190] A. Scher, G. Terwindt, W. Verschuren, M. Kruit, H. Blom, H. Kowa, R. Frants, A. van den Maagdenberg, M. van Buchem, M. Ferrari, and L. Launer, "Migraine and mthfr c677t genotype in a population-based sample.," *Ann Neurolog*, vol. 59, no. 2, pp. 372–375, 2006.

[191] F. Fernandez, S. Quinlan, J. MacMillan, R. Lea, and L. Griffiths, "Association between migraine and a functional polymorphism at the dopamine beta-hydroxylase locus.," *Neurogenetics*, vol. 10, no. 3, pp. 199–208, 2009.

[192] U. Todt, C. Netzer, M. Toliat, and C. Kubisch, "New genetic evidence for involvement of the dopamine system in migraine with aura.," *Hum Gen*, vol. 125, no. 3, pp. 265–279, 2009.

[193] Corominas, Ribases, and Macaya, "Two-stage case-control association study of dopamine-related genes and migraine," *BMC Med Genet.*, p. 10, 2009.

[194] J. N. Hirschhorn, "Genomewide association studies–illuminating biologic pathways," *N Engl J Med*, vol. 360, no. 17, 2009.

[195] E. R. Mardis, "The impact of next-generation sequencing technology on genetics," *Trends Genet*, vol. 24, no. 3, pp. 133–41, 2008.

[196] T. J. Albert, M. N. Molla, D. M. Muzny, L. Nazareth, D. Wheeler, X. Song, T. A. Richmond, C. M. Middle, M. J. Rodesch, C. J. Packard, G. M. Weinstock, and R. A. Gibbs, "Direct selection of human genomic loci by microarray hybridization," *Nat Methods*, vol. 4, no. 11, pp. 903–5, 2007.

[197] A. Gnirke, A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum, "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing," *Nat Biotechnol*, vol. 27, no. 2, pp. 182–9, 2009.

[198] E. Hodges, Z. Xuan, V. Balija, M. Kramer, M. N. Molla, S. W. Smith, C. M. Middle, M. J. Rodesch, T. J. Albert, G. J. Hannon, and W. R. McCombie, "Genome-wide in situ exon capture for selective resequencing," *Nat Genet*, vol. 39, no. 12, pp. 1522–7, 2007.

[199] D. T. Okou, K. M. Steinberg, C. Middle, D. J. Cutler, T. J. Albert, and M. E. Zwick, "Microarray-based genomic selection for high-throughput resequencing," *Nat Methods*, vol. 4, no. 11, pp. 907–9, 2007.

[200] J. Harrow, F. Denoeud, A. Frankish, A. Reymond, C. K. Chen, J. Chrast, J. Lagarde, J. G. Gilbert, R. Storey, D. Swarbreck, C. Rossier, C. Ucla, T. Hubbard, S. E. Antonarakis, and R. Guigo, "Gencode: producing a reference annotation for encode," *Genome Biol*, vol. 7 Suppl 1, pp. S4 1–9, 2006.

[201] M. Kallela, M. Wessman, and M. Farkkila, "Validation of a migraine-specific questionnaire for use in family studies," *Eur J Neurol*, vol. 8, no. 1, pp. 61–6, 2001.

[202] H. Wichmann, C. Gieger, T. Illig, and M. S. Group, "Kora-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes," *Gesundheitswesen*, vol. 67, no. Suppl 1, pp. S26–30, 2005.

[203] A. Schmermund, S. Mhlenkamp, A. Stang, D. Grnemeyer, R. Seibel, H. Hirche, K. Mann, W. Siffert, K. Lauterbach, J. Siegrist, K. Jckel, and R. Erbel, "Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the heinz nixdorf recall study. risk factors, evaluation of coronary calcium and lifestyle," *Am Heart J*, vol. 144, no. 2, pp. 212–8, 2002.

[204] A. Hofman, M. Breteler, C. van Duijn, G. Krestin, H. Pols, B. Stricker, H. Tiemeier, A. Uitterlinden, J. Vingerling, and J. Witteman, "The rotter-

dam study: objectives and design update," *Eur J Epidemiol*, vol. 22, no. 11, pp. 819–29, 2007.

[205] D. Barker, C. Osmond, T. Forsn, E. Kajantie, and J. Eriksson, "Trajectories of growth among children who have coronary events as adults," *N Engl J Med*, vol. 353, no. 17, pp. 1802–9, 2005.

[206] M. Krawczak, S. Nikolaus, H. von Eberstein, P. Croucher, N. El Mokhtari, and S. Schreiber, "Popgen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships," *Community Genet*, vol. 9, no. 1, pp. 55–61, 2006.

[207] P. Muglia, F. Tozzi, N. Galwey, C. Francks, R. Upmanyu, X. Kong, A. Antoniades, E. Domenici, J. Perry, S. Rothen, C. Vandeleur, V. Mooser, G. Waeber, P. Vollenweider, M. Preisig, S. Lucae, B. Mller-Myhsok, F. Holsboer, L. Middleton, and A. Roses, "Genome-wide association study of recurrent major depressive disorder in two european case-control cohorts," *Mol Psychiatry*, 2008.

[208] Y. Teo, M. Inouye, K. Small, R. Gwilliam, P. Deloukas, D. Kwiatkowski, and T. Clark, "A genotype calling algorithm for the illumina beadarray platform," *Bioinformatics.*, vol. 23, no. 20, pp. 2741–6, 2007.

[209] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, J. Maller, P. Sklar, P. deBakker, M. Daly, and P. Sham, "Plink: a tool set for whole-genome association and population-based linkage analyses," *Am J Hum Genet*, vol. 81, no. 3, pp. 559–75, 2007.

[210] D. Nyholt, "A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other.," *Am J Hum Genet.*, vol. 74, no. 4, pp. 765–769, 2004.

[211] B. N. Howie, P. Donnelly, and J. Marchini, "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies," *PLoS Genet*, vol. 5, no. 6, p. e1000529, 2009.

[212] A. Dimas, S. Deutsch, B. Stranger, S. Montgomery, C. Borel, H. Attar-Cohen, C. Ingle, C. Beazley, M. Gutierrez Arcelus, M. Sekowska, M. Gagnebin, J. Nisbett, P. Deloukas, E. Dermitzakis, and S. Antonarakis, "Common regulatory variation impacts gene expression in a cell type-dependent manner," *Science*, vol. 325, no. 5945, pp. 1246–50, 2009.

[213] J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly, "A new multipoint method for genome-wide association studies by imputation of genotypes," *Nat Genet*, vol. 39, no. 7, pp. 906–13, 2007.

[214] R. Magi and A. P. Morris, "Gwama: software for genome-wide association meta-analysis," *BMC Bioinformatics*, vol. 11, p. 288, 2010.

[215] P. Flicek, B. L. Aken, B. Ballester, K. Beal, E. Bragin, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, J. Fernandez-Banet, L. Gordon, S. Graf, S. Haider, M. Hammond, K. Howe, A. Jenkinson, N. Johnson, A. Kahari, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, G. Koscielny, E. Kulesha, D. Lawson, I. Longden, T. Massingham, W. McLaren, K. Megy, B. Overduin, B. Pritchard, D. Rios, M. Ruffier, M. Schuster, G. Slater, D. Smedley, G. Spudich, Y. A. Tang, S. Trevanion, A. Vilella, J. Vogel,

S. White, S. P. Wilder, A. Zadissa, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J. Hubbard, A. Parker, G. Proctor, J. Smith, and S. M. Searle, "Ensembl's 10th year," *Nucleic Acids Res*, vol. 38, no. Database issue, pp. D557–62, 2010.

[216] L. G. Wilming, J. G. Gilbert, K. Howe, S. Trevanion, T. Hubbard, and J. L. Harrow, "The vertebrate genome annotation (vega) database," *Nucleic Acids Res*, vol. 36, no. Database issue, pp. D753–60, 2008.

[217] H. Li, J. Ruan, and R. Durbin, "Mapping short dna sequencing reads and calling variants using mapping quality scores," *Genome Res*, vol. 18, no. 11, pp. 1851–8, 2008.

[218] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The sequence alignment/map format and samtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–9, 2009.

[219] A. R. Quinlan and I. M. Hall, "Bedtools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841–2, 2010.

[220] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, "Haploview: analysis and visualization of ld and haplotype maps," *Bioinformatics*, vol. 21, no. 2, pp. 263–5, 2005.

[221] S. B. Gabriel, S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, and D. Altshuler, "The structure of haplotype blocks in the human genome," *Science*, vol. 296, no. 5576, pp. 2225–9, 2002.

[222] M. Morley, C. M. Molony, T. M. Weber, J. L. Devlin, K. G. Ewens, R. S. Spielman, and V. G. Cheung, "Genetic analysis of genome-wide variation in human gene expression," *Nature*, vol. 430, no. 7001, pp. 743–7, 2004.

[223] E. E. Schadt, S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, P. S. Linsley, M. Mao, R. B. Stoughton, and S. H. Friend, "Genetics of gene expression surveyed in maize, mouse and man," *Nature*, vol. 422, no. 6929, pp. 297–302, 2003.

[224] Y. Nishimura, C. Martin, A. Vazquez-Lopez, S. Spence, A. Alvarez-Retuerto, M. Sigman, C. Steindler, S. Pellegrini, N. Schanen, S. Warren, and D. Geschwind, "Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways.," *Hum Mol Genet.*, vol. 16, no. 14, pp. 1682–1698, 2007.

[225] R. Gingras, C. Richard, M. El-Alfy, C. R. Morales, M. Potier, and A. V. Pshezhetsky, "Purification, cdna cloning, and expression of a new human blood plasma glutamate carboxypeptidase homologous to n-acetyl-aspartyl-alpha-glutamate carboxypeptidase/prostate-specific membrane antigen," *J Biol Chem*, vol. 274, no. 17, pp. 11742–50, 1999.

[226] Z. Alam, N. Coombes, R. H. Waring, A. C. Williams, and G. B. Steventon, "Plasma levels of neuroexcitatory amino acids in patients with migraine or tension headache," *J Neurol Sci*, vol. 156, no. 1, pp. 102–6, 1998.

[227] M. D. Ferrari, J. Odink, K. D. Bos, M. J. Malessy, and G. W. Bruyn, "Neuroexcitatory plasma amino acids are elevated in migraine," *Neurology*, vol. 40, no. 10, pp. 1582–6, 1990.

[228] A. Ferrari, L. Spaccapelo, D. Pinetti, R. Tacchi, and A. Bertolini, "Effective prophylactic treatments of migraine lower plasma glutamate levels," *Cephalalgia*, vol. 29, no. 4, pp. 423–9, 2009.

[229] M. Vaccaro, C. Riva, L. Tremolizzo, M. Longoni, A. Aliprandi, E. Agostoni, A. Rigamonti, M. Leone, G. Bussone, and C. Ferrarese, "Platelet glutamate uptake and release in migraine with and without aura," *Cephalalgia*, vol. 27, no. 1, pp. 35–40, 2007.

[230] H. H. Schaumburg, R. Byck, R. Gerstl, and J. H. Mashman, "Monosodium l-glutamate: its pharmacology and role in the chinese restaurant syndrome," *Science*, vol. 163, no. 869, pp. 826–8, 1969.

[231] D. C. Kang, Z. Z. Su, D. Sarkar, L. Emdad, D. J. Volsky, and P. B. Fisher, "Cloning and characterization of hiv-1-inducible astrocyte elevated gene-1, aeg-1," *Gene*, vol. 353, no. 1, pp. 8–15, 2005.

[232] Z. Z. Su, D. C. Kang, Y. Chen, O. Pekarskaya, W. Chao, D. J. Volsky, and P. B. Fisher, "Identification and cloning of human astrocyte genes displaying elevated expression after infection with hiv-1 or exposure to hiv-1 envelope glycoprotein by rapid subtraction hybridization, rash," *Oncogene*, vol. 21, no. 22, pp. 3592–602, 2002.

[233] Z. Z. Su, Y. Chen, D. C. Kang, W. Chao, M. Simm, D. J. Volsky, and P. B. Fisher, "Customized rapid subtraction hybridization (rash) gene microarrays identify overlapping expression changes in human fetal astrocytes resulting from human immunodeficiency virus-1 infection or tumor necrosis factor-alpha treatment," *Gene*, vol. 306, pp. 67–78, 2003.

[234] F. A. Chaudhry, K. P. Lehre, M. van Lookeren Campagne, O. P. Ottersen, N. C. Danbolt, and J. Storm-Mathisen, "Glutamate transporters in glial plasma membranes: highly differentiated localizations revealed by quantitative ultrastructural immunocytochemistry," *Neuron*, vol. 15, no. 3, pp. 711–20, 1995.

[235] G. Pines, N. C. Danbolt, M. Bjoras, Y. Zhang, A. Bendahan, L. Eide, H. Koepsell, J. Storm-Mathisen, E. Seeberg, and B. I. Kanner, "Cloning and expression of a rat brain l-glutamate transporter," *Nature*, vol. 360, no. 6403, pp. 464–7, 1992.

[236] J. D. Rothstein, L. Martin, A. I. Levey, M. Dykes-Hoberg, L. Jin, D. Wu, N. Nash, and R. W. Kuncl, "Localization of neuronal and glial glutamate transporters," *Neuron*, vol. 13, no. 3, pp. 713–25, 1994.

[237] M. A. Moskowitz, K. Nozaki, and R. P. Kraig, "Neocortical spreading depression provokes the expression of c-fos protein-like immunoreactivity within trigeminal nucleus caudalis via trigeminovascular mechanisms," *J Neurosci*, vol. 13, no. 3, pp. 1167–77, 1993.

[238] R. P. Woods, M. Iacoboni, and J. C. Mazziotta, "Brief report: bilateral spreading cerebral hypoperfusion during spontaneous migraine headache," *N Engl J Med*, vol. 331, no. 25, pp. 1689–92, 1994.

[239] B. de Vries, H. Mamsa, A. H. Stam, J. Wan, S. L. Bakker, K. R. Vanmolkot, J. Haan, G. M. Terwindt, E. M. Boon, B. D. Howard, R. R. Frants, R. W. Baloh, M. D. Ferrari, J. C. Jen, and A. M. van den Maagdenberg, "Episodic

ataxia associated with eaat1 mutation c186s affecting glutamate reuptake," *Arch Neurol*, vol. 66, no. 1, pp. 97–101, 2009.

[240] M. Peeters, M. J. Gunthorpe, P. J. Strijbos, P. Goldsmith, N. Upton, and M. F. James, "Effects of pan- and subtype-selective n-methyl-d-aspartate receptor antagonists on cortical spreading depression in the rat: therapeutic potential for migraine," *J Pharmacol Exp Ther*, vol. 321, no. 2, pp. 564–72, 2007.

[241] M. Bigal, A. Rapoport, F. Sheftell, D. Tepper, and S. Tepper, "Memantine in the preventive treatment of refractory migraine," *Headache*, vol. 48, no. 9, pp. 1337–42, 2008.

[242] A. Charles, C. Flippen, M. Romero Reyes, and K. C. Brennan, "Memantine for prevention of migraine: a retrospective study of 60 cases," *J Headache Pain*, vol. 8, no. 4, pp. 248–50, 2007.

[243] V. Anttila, H. Stefansson, M. Kallela, U. Todt, G. M. Terwindt, M. S. Calafato, D. R. Nyholt, A. S. Dimas, T. Freilinger, B. Muller-Myhsok, V. Artto, M. Inouye, K. Alakurtti, M. A. Kaunisto, E. Hamalainen, B. de Vries, A. H. Stam, C. M. Weller, A. Heinze, K. Heinze-Kuhn, I. Goebel, G. Borck, H. Gobel, S. Steinberg, C. Wolf, A. Bjornsson, G. Gudmundsson, M. Kirchmann, A. Hauge, T. Werge, J. Schoenen, J. G. Eriksson, K. Hagen, L. Stovner, H. E. Wichmann, T. Meitinger, M. Alexander, S. Moebus, S. Schreiber, Y. S. Aulchenko, M. M. Breteler, A. G. Uitterlinden, A. Hofman, C. M. van Duijn, P. Tikka-Kleemola, S. Vepsalainen, S. Lucae, F. Tozzi, P. Muglia, J. Barrett, J. Kaprio, M. Farkkila, L. Peltonen, K. Ste-

fansson, J. A. Zwart, M. D. Ferrari, J. Olesen, M. Daly, M. Wessman, A. M. van den Maagdenberg, M. Dichgans, C. Kubisch, E. T. Dermitzakis, R. R. Frants, and A. Palotie, "Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1," *Nat Genet*, vol. 42, no. 10, pp. 869–73, 2010.

[244] D. I. Chasman, M. Schurks, V. Anttila, B. de Vries, U. Schminke, L. J. Launer, G. M. Terwindt, A. M. van den Maagdenberg, K. Fendrich, H. Volzke, F. Ernst, L. R. Griffiths, J. E. Buring, M. Kallela, T. Freilinger, C. Kubisch, P. M. Ridker, A. Palotie, M. D. Ferrari, W. Hoffmann, R. Y. Zee, and T. Kurth, "Genome-wide association study reveals three susceptibility loci for common migraine in the general population," *Nat Genet*, vol. 43, no. 7, pp. 695–8, 2011.

[245] J. Marchini and B. Howie, "Genotype imputation for genome-wide association studies," *Nat Rev Genet*, vol. 11, no. 7, pp. 499–511, 2010.

[246] W. Knowlton, R. Daniels, R. Palkar, D. McCoy, and M. DD., "Pharmacological blockade of trpm8 ion channels alters cold and cold pain responses in mice," *Plos*, vol. 6, no. 9, p. e25894, 2011.

[247] P. Wrigley, H. Jeong, and C. Vaughan, "Primary afferents with trpm8 and trpa1 profiles target distinct subpopulations of rat superficial dorsal horn neurones.," *Br J Pharmacol.*, vol. 157, no. 3, pp. 371–80, 2009.

[248] Y. Liu and N. Qin, "Trpm8 in health and disease: cold sensing and beyond," *Adv Exp Med Biol.*, vol. 704, pp. 185–208, 2011.

[249] Biondi, "Is migraine a neuropathic pain syndrome?.," *Curr Pain Headache Rep.*, vol. 10, no. 3, pp. 167–78, 2006.

[250] L. Broad, A. Mogg, R. Beattie, A. Ogden, M. Blanco, and D. Bleakman, "Trp channels as emerging targets for pain therapeutics.," *Expert Opin Ther Targets.*, vol. 13, no. 1, pp. 69–81, 2009.

[251] M. Franchini and M. Montagnana, "Low-density lipoprotein receptor-related protein 1: new functions for an old molecule.," *Clin Chem Lab Med.*., vol. 49, no. 6, pp. 967–70, 2011.

[252] S. Ranganathan, C. Cao, J. Catania, M. Migliorini, L. Zhang, and D. Strickland, "Molecular basis for the interaction of low density lipoprotein receptor-related protein 1 (lrp1) with integrin alphambeta2: identification of binding sites within alphambeta2 for lrp1.," *J Biol Chem.*, vol. 286, no. 35, p. 30535, 2011.

[253] A. Lillis, L. Van Duyn, J. Murphy-Ullrich, and D. Strickland, "Ldl receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies..," *Physiol Rev.*, vol. 88, no. 3, pp. 887–918, 2008.

[254] P. Boucher, M. Gotthardt, W. Li, R. Anderson, and J. Herz, "Lrp: role in vascular wall integrity and protection from atherosclerosis.," *Science*, vol. 300, no. 5617, pp. 329–32, 2003.

[255] Q. Liu, J. Trotter, J. Zhang, M. Peters, H. Cheng, J. Bao, X. Han, E. Weeber, and G. Bu, "Neuronal lrp1 knockout in adult mice leads to impaired brain lipid metabolism and progressive, age-dependent synapse loss and neurodegeneration.," *J Neurosci*, vol. 30, no. 50, pp. 17068–78, 2010.

[256] M. Vikelis and D. Mitsikostas, "The role of glutamate and its receptors in migraine.," *CNS Neurol Disord Drug Targets*, vol. 6, no. 4, pp. 251–7, 2007.

[257] A. Kanai, K. Suzuki, K. Tanimoto, J. Mizushima-Sugano, Y. Suzuki, and S. Sugano, "Characterization of stat6 target genes in human b cells and lung epithelial cells.," *DNA Res*, vol. 18, no. 5, pp. 379–92, 2011.

[258] M. Bennett, R. Pollitt, S. Goodman, D. Hale, and J. Vamecq, "Atypical riboflavin-responsive glutaric aciduria, and deficient peroxisomal glutaryl-coa oxidase activity: a new peroxisomal disorder," *J Inherit Metab Dis*, vol. 14, no. 2, pp. 165–73, 1991.

[259] E. Sherman, K. Strauss, S. Tortorelli, M. Bennett, I. Knerr, D. Morton, and E. Puffenberger, "Genetic mapping of glutaric aciduria, type 3, to chromosome 7 and identification of mutations in c7orf10," *Am J Hum Genet*, vol. 83, no. 5, pp. 604–609, 2008.

[260] Y. Even, S. Durieux, M. Escande, J. Lozano, G. Peaucellier, D. Weil, and A. Genevir, "Cdc2l5, a cdk-like kinase with rs domain, interacts with the asf/sf2-associated protein p32 and affects splicing in vivo.," *J Cell Biochem.*, vol. 99, no. 3, pp. 890–904, 2006.

[261] S. B. Ng, A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad, and J. Shendure, "Exome sequencing identifies mll2 mutations as a cause of kabuki syndrome," *Nat Genet*, vol. 42, no. 9, pp. 790–3, 2010.

[262] M. Choi, U. I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, and R. P. Lifton, "Genetic diagnosis by whole exome capture and massively parallel dna sequencing," *Proc Natl Acad Sci U S A*, vol. 106, no. 45, pp. 19096–101, 2009.

[263] S. B. Ng, E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D. A. Nickerson, and J. Shendure, "Targeted capture and massively parallel sequencing of 12 human exomes," *Nature*, vol. 461, no. 7261, pp. 272–6, 2009.

[264] K. D. Pruitt, J. Harrow, R. A. Harte, C. Wallin, M. Diekhans, D. R. Maglott, S. Searle, C. M. Farrell, J. E. Loveland, B. J. Ruef, E. Hart, M. M. Suner, M. J. Landrum, B. Aken, S. Ayling, R. Baertsch, J. Fernandez-Banet, J. L. Cherry, V. Curwen, M. Dicuccio, M. Kellis, J. Lee, M. F. Lin, M. Schuster, A. Shkeda, C. Amid, G. Brown, O. Dukhanina, A. Frankish, J. Hart, B. L. Maidak, J. Mudge, M. R. Murphy, T. Murphy, J. Rajan, B. Rajput, L. D. Riddick, C. Snow, C. Steward, D. Webb, J. A. Weber, L. Wilming, W. Wu, E. Birney, D. Haussler, T. Hubbard, J. Ostell, R. Durbin, and D. Lipman, "The consensus coding sequence (ccds) project: Identifying a common protein-coding gene set for the human and mouse genomes," *Genome Res*, vol. 19, no. 7, pp. 1316–23, 2009.

[265] A. J. Coffey, F. Kokocinski, M. S. Calafato, C. E. Scott, P. Palta, E. Drury, C. J. Joyce, E. M. Leproust, J. Harrow, S. Hunt, A. E. Lehesjoki, D. J. Turner, T. J. Hubbard, and A. Palotie, "The gencode exome: sequencing

the complete human exome," *Eur J Hum Genet*, vol. 19, no. 7, pp. 827–31, 2011.

[266] C. A. Hubner and T. J. Jentsch, "Ion channel diseases," *Hum Mol Genet*, vol. 11, no. 20, pp. 2435–45, 2002.

[267] P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. Stratton, "A census of human cancer genes," *Nat Rev Canc*, vol. 4, no. 3, pp. 177–83, 2004.

[268] E. Kalay, G. Yigit, Y. Aslan, K. Brown, E. Pohl, L. Bicknell, H. Kayserili, Y. Li, a. et, and B. Wollnik, "Cep152 is a genome maintenance protein disrupted in seckel syndrome.," *Nat Gen*, vol. 43, no. 1, pp. 23–26, 2011.

[269] C. A. Albers, A. Cvejic, R. Favier, E. E. Bouwmans, M. C. Alessi, P. Bertone, G. Jordan, R. N. Kettleborough, G. Kiddle, M. Kostadima, R. J. Read, B. Sipos, S. Sivapalaratnam, P. A. Smethurst, J. Stephens, K. Voss, A. Nurden, A. Rendon, P. Nurden, and W. H. Ouwehand, "Exome sequencing identifies nbeal2 as the causative gene for gray platelet syndrome," *Nat Genet*, 2011.

[270] M. Krawczak, J. Reiss, and D. N. Cooper, "The mutational spectrum of single base-pair substitutions in mrna splice junctions of human genes: causes and consequences," *Hum Genet*, vol. 90, no. 1-2, pp. 41–54, 1992.

[271] G. S. Wang and T. A. Cooper, "Splicing in disease: disruption of the splicing code and the decoding machinery," *Nat Rev Genet*, vol. 8, no. 10, pp. 749–61, 2007.

[272] S. G. Amara and A. C. Fontana, "Excitatory amino acid transporters: keeping up with glutamate," *Neurochem Int*, vol. 41, no. 5, pp. 313–8, 2002.

[273] J. Johnson, J. Mandrioli, M. Benatar, Y. Abramzon, V. Van Deerlin, J. Trojanowski, J. Gibbs, M. Brunetti, S. Gronka, J. Wuu, D. J., L. McCluskey, M. Martinez-Lage, D. Falcone, D. Hernandez, S. Arepalli, S. Chong, J. Schymick, J. Rothstein, F. Landi, Y. Wang, A. Calvo, G. Mora, M. Sabatelli, M. Monsurr, S. Battistini, F. Salvi, R. Spataro, P. Sola, G. Borghero, I. Consortium, G. Galassi, S. Scholz, J. Taylor, G. Restagno, A. Chi, and B. Traynor, "Exome sequencing reveals vcp mutations as a cause of familial als.," *Neuron*, vol. 68, no. 5, p. 857, 2010.