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

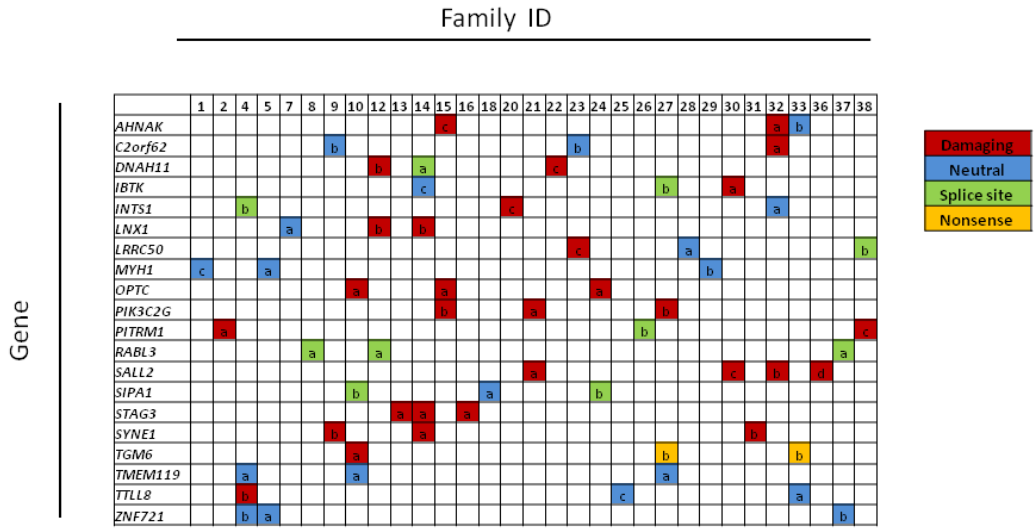

Figure 6.1: **Genes with shared rare functional variants in three or more families.** Letters (a, b, c, d) represent variants. Colors represent functional consequence of the variant as specified in the legend on the left side of the figure. Damaging (red) and neutral (blue) represent the predicted deleteriousness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

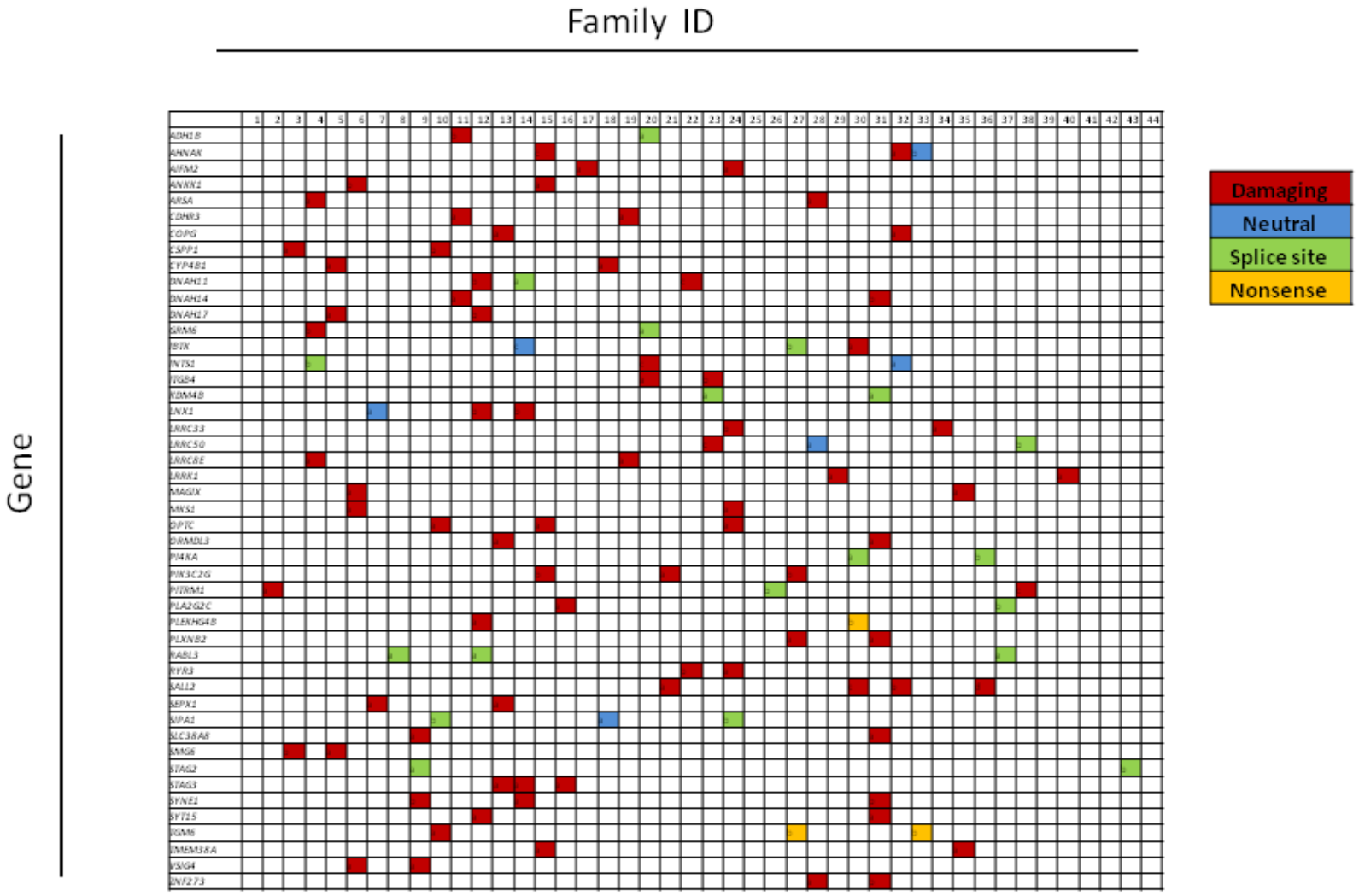

Figure 6.2: **Genes with possibly damaging missense, splice site or nonsense variants in two or more families.** Letters (a, b, c, d) represent variants. Colors represent functional consequence of the variant as specified in the legend on the left side of the figure. Damaging (red) and neutral (blue)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

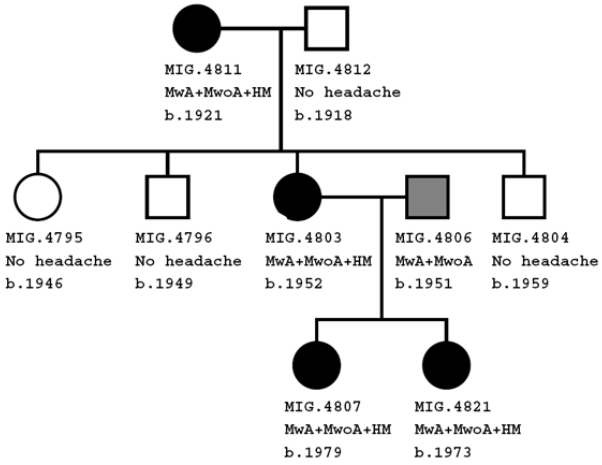

Figure 6.3: **Pedigree of family 1.** Individuals are represented as males (squares), females (circles) and affected (blackened symbols). In family 1, MIG 4821 (proband), MIG 4807 (sister), MIG 4803 (mother) and MIG 4811 (grandmother) suffered from hemiplegic migraine. Exome sequencing was done for MIG 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

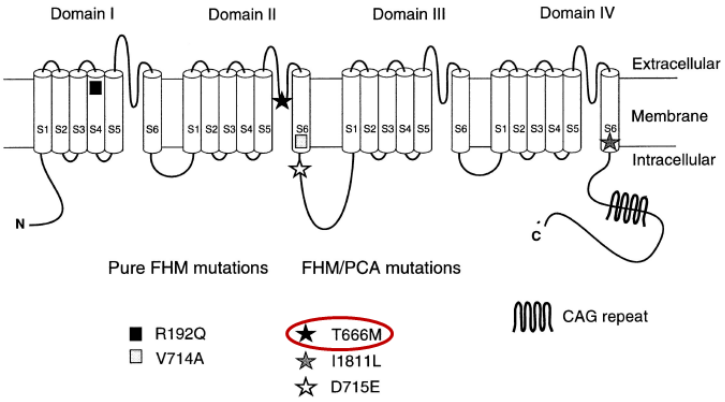

Figure 6.4: *CACNA1A* **missense mutations causing hemiplegic migraine (HM).** A schematic structure of the *CACNA1A* is represented, and positions of some HM causing mutations are marked. A red circle has been drawn around the mutation detected in the 2 cases of family 1. PCA, progressive cerebellar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

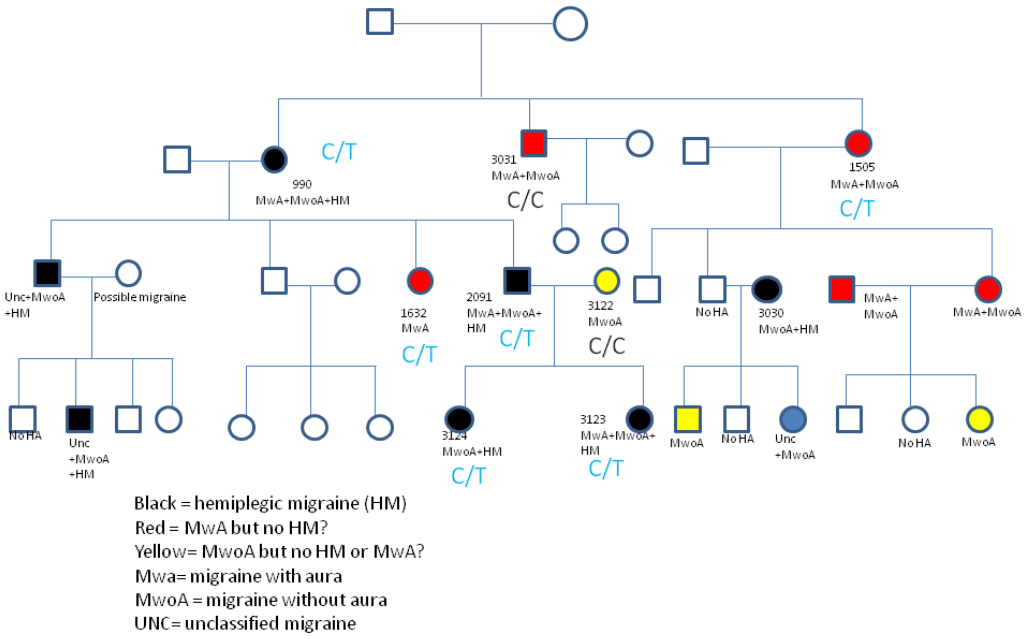

Figure 6.5: **Pedigree of family 2.** Individuals are represented as males (squares), females (circles), FHM cases (blackened symbols), migraine with aura cases (red symbols), migraine without aura cases (yellow symbols) and unclassified (blue symbols). In family 2, patient 2091 (proband), patient 3124 (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-

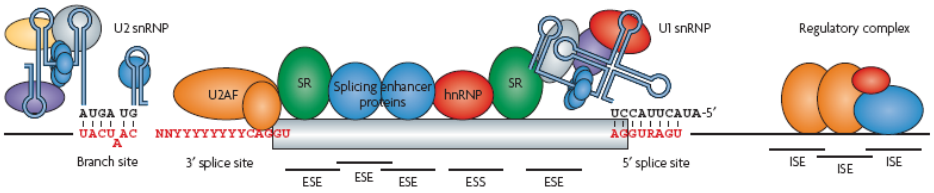

Figure 6.6: **The splicing code** Red indicates consensus splice site sequences at the intronexon boundaries. ESE, exon splicing enhancer. ISE, intron splicing enhancer. Adapted from Wang et al. (2007) [271]

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.

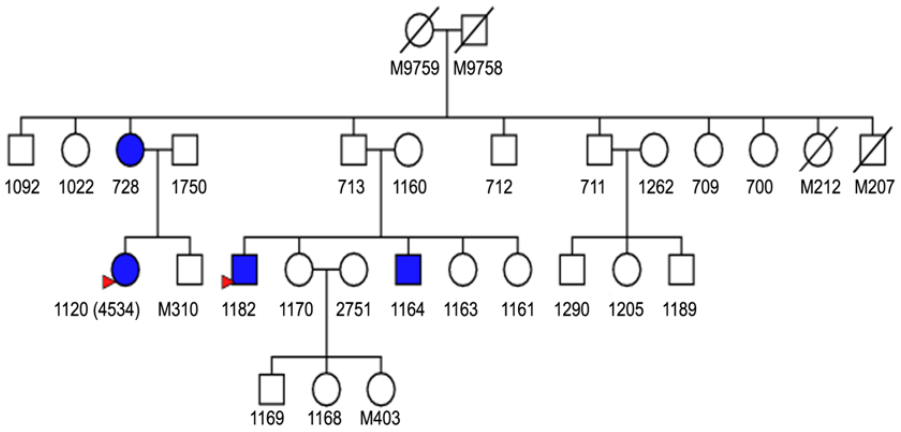

Figure 6.7: **Pedigree of family 3.** Individuals are represented as males (squares), females (circles) and FHM affected (blue symbols). In family 3, patient 1120 (proband), patient 728 (mother), patient 1182 (cousin) and patient 1164 (cousin) suffered from hemiplegic migraine. Exome sequencing was done 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,