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.