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

List of Figures

List of Tables

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