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.