Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine

Padhraig Gormley*[,1,2,3,4], Verneri Anttila*[,2,3,5], Bendik S Winsvold[6,7,8], Priit Palta[9], Tonu Esko[2,10,11], Tune H. Pers[2,11,12,13], Kai-How Farh[2,5,14], Ester Cuenca-Leon[1,2,3,15], Mikko Muona[9,16,17,18], Nicholas A Furlotte[19], Tobias Kurth[20,21], Andres Ingason[22], George McMahon[23], Lannie Ligthart[24], Gisela M Terwindt[25], Mikko Kallela[26], Tobias M Freilinger[27,28], Caroline Ran[29], Scott G Gordon[30], Anine H Stam[25], Stacy Steinberg[22], Guntram Borck[31], Markku Koiranen[32], Lydia Quaye[33], Hieab HH Adams[34,35], Terho Lehtimäki[36], Antti-Pekka Sarin[9], Juho Wedenoja[37], David A Hinds[19], Julie E Buring[21,38], Markus Schürks[39], Paul M Ridker[21,38], Maria Gudlaug Hrafnsdottir[40], Hreinn Stefansson[22], Susan M Ring[23], Jouke-Jan Hottenga[24], Brenda WJH Penninx[41], Markus Färkkilä[26], Ville Artto[26], Mari Kaunisto[9], Salli Vepsäläinen[26], Rainer Malik[27], Andrew C Heath[42], Pamela A F Madden[42], Nicholas G Martin[30], Grant W Montgomery[30], Mitja Kurki[1,2,3], Mart Kals[10], Reedik Mägi[10], Kalle Pärn[10], Eija Hämäläinen[9], Hailiang Huang[2,3,5], Andrea E Byrnes[2,3,5], Lude Franke[43], Jie Huang[4], Evie Stergiakouli[23], Phil H Lee[1,2,3], Cynthia Sandor[44], Caleb Webber[44], Zameel Cader[45,46], Bertram Muller-Myhsok[47], Stefan Schreiber[48], Thomas Meitinger[49], Johan G Eriksson[50,51], Veikko Salomaa[51], Kauko Heikkilä[52], Elizabeth Loehrer[34,53], Andre G Uitterlinden[54], Albert Hofman[34], Cornelia M van Duijn[34], Lynn Cherkas[33], Linda M. Pedersen[6], Audun Stubhaug[55,56], Christopher S Nielsen[55,57], Minna Männikkö[32], Evelin Mihailov[10], Lili Milani[10], Hartmut Göbel[58], Ann-Louise Esserlind[59], Anne Francke Christensen[59], Thomas Folkmann Hansen[60], Thomas Werge[61,62,63], International Headache Genetics Consortium[64], Jaakko Kaprio[9,65,66], Arpo J Aromaa[51], Olli Raitakari[67,68], M Arfan Ikram[34,35,68], Tim Spector[33], Marjo-Riitta Järvelin[32,70,71,72], Andres Metspalu[10], Christian Kubisch[73], David P Strachan[74], Michel D Ferrari[25], Andrea C Belin[29], Martin Dichgans[27,75], Maija Wessman[9,16], Arn MJM van den Maagdenberg[25,76], John-Anker Zwart[6,7,8], Dorret I Boomsma[24], George Davey Smith[23], Kari Stefansson[22,77], Nicholas Eriksson[19], Mark J Daly[2,3,5], Benjamin M Neale[§,2,3,5], Jes Olesen[§,59], Daniel I Chasman[§,21,38], Dale R Nyholt[§,78], and Aarno Palotie[§,1,2,3,4,5,9,79].

[1]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, USA. [2]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA. [4]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. [5]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, USA. [6]FORMI, Oslo University Hospital, P.O. 4956 Nydalen, 0424 Oslo, Norway. [7]Department of Neurology, Oslo University Hospital, P.O. 4956 Nydalen, 0424 Oslo, Norway. [8]Institute of Clinical Medicine, University of Oslo, P.O. 1171 Blindern, 0318 Oslo, Norway. [9]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [10]Estonian Genome Center, University of Tartu, Tartu, Estonia. [11]Division of Endocrinology, Boston Children's Hospital, Boston, USA. [12]Statens Serum Institut, Dept of Epidemiology Research, Copenhagen, Denmark. [13]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. [14]Illumina, 5200 Illumina Way, San Diego, USA. [15]Vall d'Hebron Research Institute, Pediatric Neurology, Barcelona, Spain. [16]Folkhälsan Institute of Genetics, Helsinki, Finland, FI-00290. [17]Neuroscience Center, University of Helsinki, Helsinki, Finland, FI-00014. [18]Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland, FI-00014. [19]23andMe, Inc., 899 W. Evelyn Avenue, Mountain View, CA, USA. [20]Inserm Research Center for Epidemiology and Biostatistics (U897), University of Bordeaux, 33076 Bordeaux, France. [21]Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA 02215. [22]deCODE Genetics, 101 Reykjavik, Iceland. [23]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [24]VU University Amsterdam, Department of Biological Psychology, Amsterdam, the Netherlands, 1081 BT. [25]Leiden University Medical Centre, Department

of Neurology, Leiden, The Netherlands, PO Box 9600, 2300 RC. [26]Department of Neurology, Helsinki University Central Hospital, Haartmaninkatu 4, 00290 Helsinki, Finland. [27]Institute for Stroke and Dementia Research, Klinikum der Universtität München, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 17, 81377 Munich Germany. [28]Department of Neurology and Epileptology, Hertie Institute for Clincal Brain Research, University of Tuebingen. [29]Karolinska Institutet, Department of Neuroscience, 171 77 Stockholm, Sweden. [30]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, QLD 4006, Australia. [31]Ulm University, Institute of Human Genetics, 89081 Ulm, Germany. [32]University of Oulu, Center for Life Course Epidemiology and Systems Medicine, Oulu, Finland, Box 5000, Fin-90014 University of Oulu. [33]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [34]Dept of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands, 3015 CN. [35]Dept of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands, 3015 CN. [36]Department of Clinical Chemistry, Fimlab Laboratories, and School of Medicine, University of Tampere, Tampere, Finland, 33520. [37]Department of Public Health, University of Helsinki, Helsinki, Finland. [38]Harvard Medical School, Boston MA 02115. [39]University Duisburg Essen, Essen, Germany. [40]Landspitali University Hospital, 101 Reykjavik, Iceland. [41]VU University Medical Centre, Department of Psychiatry, Amsterdam, the Netherlands, 1081 HL. [42]Department of Psychiatry, Washington University School of Medicine, 660 South Euclid, CB 8134, St. Louis, MO 63110, USA. [43]University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, 9700RB. [44]MRC Functional Genomics Unit, Department of Physiology, Anatomy & Genetics, Oxford University, UK. [45]Nuffield Department of Clinical Neuroscience, University of Oxford, UK. [46]Oxford Headache Centre, John Radcliffe Hospital, Oxford, UK. [47]Max-Planck-Institute of Psychiatry, Munich, Germany. [48]Christian Albrechts University, Kiel, Germany. [49]Institute of Human Genetics, Helmholtz Center Munich, Neuherberg, Germany. [50]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki Finland. [51]National Institute for Health and Welfare, Helsinki, Finland. [52]Institute of Clinical Medicine, University of Helsinki, Helsinki, Finland. [53]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, USA 02115. [54]Dept of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands, 3015 CN. [55]Dept of Pain Management and Research, Oslo University Hospital, Oslo, 0424 Oslo, Norway. [56]Medical Faculty, University of Oslo, Oslo, 0318 Oslo, Norway. [57]Division of Mental Health, Norwegian Institute of Public Health,P.O. Box 4404 Nydalen, Oslo, Norway, NO-0403. [58]Kiel Pain and Headache Center, 24149 Kiel, Germany. [59]Danish Headache Center, Department of Neurology, Rigshospitalet, Glostrup Hospital, University of Copenhagen, Denmark. [60]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, University of Copenhagen, Roskilde, Denmark. [61]Institute Of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, DK-2100 Copenhagen, Denmark. [62]Institute of Clinical Sciences, Faculty of Medicine and Health Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark. [63]iPSYCH - The Lundbeck Foundation's Initiative for Integrative Psychiatric Research, DK-2100 Copenhagen, Denmark. [64]A list of members and affiliations appears in the Supplementary Note. [65]Department of Public Health, University of Helsinki, Helsinki, Finland. [66]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [67]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, 20521. [68]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, 20521. [69]Dept of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands, 3015 CN. [70]Imperial College London, Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPE) Centre for Environment and Health, School of Public Health, UK, W2 1PG. [71]University of Oulu, Biocenter Oulu, Finland, Box 5000, Fin-90014 University of Oulu. [72]Oulu University Hospital, Unit of Primary Care, Oulu, Finland, Box 10, Fin-90029 OYS. [73]University Medical Center Hamburg Eppendorf, Institute of Human Genetics, 20246 Hamburg, Germany. [74]Population Health Research Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK. [75]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [76]Leiden University Medical Centre, Department of Human Genetics, Leiden, The Netherlands, PO Box 9600, 2300 RC. [77]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. [78]Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Ave, Kelvin Grove, QLD 4059, Australia. [79]Department of Neurology, Massachusetts General Hospital, Boston, USA.

* These authors contributed equally to this work.

§ These authors jointly supervised this work.

Correspondence should be addressed to Aarno Palotie (aarno.palotie@helsinki.fi).

Migraine is a debilitating neurological disorder affecting around 1 in 7 people worldwide, but its molecular mechanisms remain poorly understood. Some debate exists over whether migraine is a disease of vascular dysfunction or a result of neuronal dysfunction with secondary vascular changes. Genome-wide association (GWA) studies have thus far identified 13 independent loci associated with migraine. To identify new susceptibility loci, we performed the largest genetic study of migraine to date, comprising 59,674 cases and 316,078 controls from 22 GWA studies. We identified 44 independent single nucleotide polymorphisms (SNPs) significantly associated with migraine risk ($P < 5 \times 10^{-8}$) that map to 38 distinct genomic loci, including 28 loci not previously reported and the first locus identified on chromosome X. In subsequent computational analyses, the identified loci showed enrichment for genes expressed in vascular and smooth muscle tissues, consistent with a predominant theory of migraine that highlights vascular etiologies.

Migraine is ranked as the third most common disease worldwide, with a lifetime prevalence of 15-20%, affecting up to one billion people across the globe[1,2]. It ranks as the 7th most disabling of all diseases worldwide (or 1st most disabling neurological disease) in terms of years of life lost to disability[1] and is the 3rd most costly neurological disorder after dementia and stroke[3]. There is debate about whether migraine is a disease of vascular dysfunction, or a result of neuronal dysfunction with vascular changes representing downstream effects not themselves causative of migraine[4,5]. However, genetic evidence favoring one theory versus the other is lacking. At the phenotypic level, migraine is defined by diagnostic criteria from the International Headache Society[6]. There are two prevalent sub-forms: migraine without aura is characterized by recurrent attacks of moderate or severe headache associated with nausea or hypersensitivity to light and sound. Migraine with aura is characterized by transient visual and/or sensory and/or speech symptoms usually followed by a headache phase similar to migraine without aura.

Family and twin studies estimate a heritability of 42% (95% confidence interval [CI] = 36-47%) for migraine[7], pointing to a genetic component of the disease. Despite this, genetic association studies have revealed relatively little about the molecular mechanisms that contribute to pathophysiology. Understanding has been limited partly because, to date, only 13 genome-wide significant risk loci have been identified for the prevalent forms of migraine[8–11]. In familial hemiplegic migraine (FHM), a rare Mendelian form of the disease, three ion transport-related genes (*CACNA1A*, *ATP1A2* and *SCN1A*) have been implicated[12–14]. These findings suggest that

mechanisms that regulate neuronal ion homeostasis might also be involved in migraine more generally, however, no genes related to ion transport have yet been identified for these more prevalent forms of migraine[15].

We performed a meta-analysis of 22 genome-wide association (GWA) studies, consisting of 59,674 cases and 316,078 controls collected from six tertiary headache clinics and 27 population-based cohorts through our worldwide collaboration in the International Headache Genetics Consortium (IHGC). This combined dataset contained over 35,000 new migraine cases not included in previously published GWA studies. Here we present the findings of this new meta-analysis, including 38 genomic loci, harboring 44 independent association signals identified at levels of genome-wide significance, which support current theories of migraine pathophysiology and also offer new insights into the disease.

Results

Significant associations at 38 independent genomic loci

The primary meta-analysis was performed on all migraine samples available through the IHGC, regardless of ascertainment. These case samples included both individuals diagnosed with migraine by a doctor as well as individuals with self-reported migraine via questionnaires. Study design and sample ascertainment for each individual study is outlined in the **Supplementary Note** (and summarized in **Supplementary Table 1**). The final combined sample consisted of 59,674 cases and 316,078 controls in 22 non-overlapping case-control samples (**Table 1**). All samples were of European ancestry. Before including the largest study from 23andMe, we confirmed that it did not contribute any additional heterogeneity compared to the other population and clinic-based studies **(Supplementary Table 2)**.

The 22 individual GWA studies completed standard quality control protocols (**Online Methods**) summarized in **Supplementary Table 3.** Missing genotypes were then imputed into each sample using a common 1000 Genomes Project reference panel[16]. Association analyses were performed within each study using logistic regression on the imputed marker dosages while adjusting for sex and other covariates where necessary (**Online Methods** and **Supplementary Table 4**). The association results were combined using an inverse-variance weighted fixed-effects meta-analysis. Markers were filtered for imputation quality and other metrics (**Online Methods**) leaving 8,094,889 variants for consideration in our primary analysis.

Among these variants in the primary analysis, we identified 44 genome-wide significant SNP associations ($P < 5 \times 10^{-8}$) that are independent ($r^2 < 0.1$) with regards to linkage disequilibrium (LD). We validated the 44 SNPs by comparing genotypes in a subset of the sample to those obtained from whole-genome sequencing (**Supplementary Table 5**). To help identify candidate risk genes from these, we defined an associated locus as the genomic region bounded by all markers in LD ($r^2 > 0.6$ in 1000 Genomes, Phase I, EUR individuals) with each of the 44 index SNPs and in addition, all such regions in close proximity (< 250 kb) were merged. From these defined regions we implicate 38 distinct genomic loci in total for the prevalent forms of migraine, 28 of which have not previously been reported (**Figure 1**).

These 38 loci replicate 10 of the 13 previously reported genome-wide associations to migraine (**Table 2**). Six of the 38 loci contain a secondary genome-wide significant SNP ($P < 5 \times 10^{-8}$) not in LD ($r^2 < 0.1$) with the top SNP in the locus (**Table 2**). Five of these secondary signals were found in known loci (at *LRP1*, *PRDM16*, *FHL5*, *TRPM8*, and *TSPAN2*), while the sixth was found within one of the 28 new loci (*PLCE1*). Therefore, out of the 44 LD-independent SNPs reported here, 34 are new associations to migraine. Three previously reported loci that were associated to subtypes of migraine (rs1835740 near *MTDH* to migraine with aura, rs10915437 near *AJAP1* to migraine clinical-samples, and rs10504861 near *MMP16* to migraine without aura)[8,11] show only nominal significance in the current meta-analysis ($P = 5 \times 10^{-3}$ for rs1835740, $P = 4.4 \times 10^{-5}$ for rs10915437, and $P = 4.9 \times 10^{-5}$ for rs10504861, **Supplementary Table 6**), however, these loci have since been shown to be associated to specific phenotypic features of migraine[17] and therefore may require a more phenotypically homogeneous sample to be accurately assessed for association. Four out of 44 SNPs (at *TRPM8*, *ZCCHC14*, *MRVI1*, and *CCM2L*) exhibited moderate heterogeneity across the individual GWA studies (Cochran's Q test *p-value* < 0.05, **Supplementary Table 7**) therefore at these markers we applied a random effects model[18].

Characterization of the associated loci

In total, 32 of 38 (84%) loci overlap with transcripts from protein-coding genes, and 17 (45%) of these regions contain just a single gene (see **Supplementary Figure 1** for regional plots of the 38 genomic loci and **Supplementary Table 8** for extended information on each locus). Among the 38 loci, only two contain ion channel genes (*KCNK5*[19] and *TRPM8*[20]). Hence, despite previous hypotheses of migraine as a potential channelopathy[5,21], the loci identified to date do

not support common variants in ion channel genes as strong susceptibility components in prevalent forms of migraine. However, three other loci do contain genes involved more generally in ion homeostasis (*SLC24A3*[22], *ITPK1*[23], and GJA1[24], **Supplementary Table 9**).

Several of the genes have previous associations to vascular disease (*PHACTR1*,[25,26] *TGFBR2*,[27] *LRP1*,[28] *PRDM16*,[29] *RNF213*,[30] *JAG1*,[31] *HEY2*,[32] *GJA1*[33], *ARMS2*[34]), or are involved in smooth muscle contractility and regulation of vascular tone (*MRVI1*,[35] *GJA1,*[36] *SLC24A3*,[37] *NRP1*[38]). Three of the 44 migraine index SNPs have previously reported associations in the National Human Genome Research Institute (NHGRI) GWAS catalog at exactly the same SNP (rs9349379 at *PHACTR1* with coronary heart disease[39–41], coronary artery calcification[42], and cervical artery dissection; rs11624776 at *ITPK1* with thyroid hormone levels[43]; and rs11172113 at *LRP1* with pulmonary function; **Supplementary Table 10**). Six of the loci harbor genes that are involved in nitric oxide signaling and oxidative stress (*REST*[44], *GJA1*[45], *YAP1*[46], *PRDM16*[47], *LRP1*[48], and *MRVI1*[49]).

From each locus we chose the nearest gene to the index SNP to assess gene expression activity in tissues from the GTEx consortium (**Supplementary Figure 2**). While we found that most of the putative migraine loci genes were expressed in many different tissue types, we could detect tissue specificity in certain instances whereby some genes showed significantly higher expression in a particular tissue group relative to the others. For instance four genes were more actively expressed in brain (*GPR149*, *CFDP1*, *DOCK4*, and *MPPED2*) compared to other tissues, whereas eight genes were specifically active in vascular tissues (*PRDM16*, *MEF2D*, *FHL5*, *C7orf10*, *YAP1*, *LRP1*, *ZCCHC14*, and *JAG1*). Many of the other putative migraine loci genes were actively expressed in more than one tissue group.

Genomic inflation and LD-score regression analysis

To assess whether the 38 loci harbor true associations with migraine rather than reflecting systematic differences between cases and controls (such as population stratification) we analyzed the genome-wide inflation of test statistics in our primary meta-analysis. As expected for a complex polygenic trait, the distribution of test statistics deviates from the null (genomic inflation factor λ_{GC} = 1.24, **Supplementary Figure 3**) which is in line with other large GWA study meta-analyses[50–53]. Since much of the inflation in a polygenic trait arises from LD between the causal SNPs and many other neighboring SNPs in the local region, we LD-pruned the meta-analysis results to create a set of LD-independent markers (i.e. in PLINK[54] with a 250-kb sliding

window and $r^2 > 0.2$). The resulting genomic inflation was reduced ($\lambda_{GC} = 1.15$, **Supplementary Figure 4**) and likely reflects the inflation remaining due to the polygenic signal at many independent loci, including those not yet significantly associated.

To confirm that the observed inflation is primarily coming from true polygenic signal, we analyzed the meta-analysis results from all imputed markers using LD-score regression[55]. This method tests for a linear relationship between marker test statistics and LD score, defined as the sum of r^2 values between a marker and all other markers within a 1-Mb window. The primary analysis results show a linear relationship between association test statistics and LD-score (**Supplementary Figure 5**) and estimate that the majority (88.2%) of the inflation in test statistics can be ascribed to true polygenic signal rather than population stratification or other confounders. These results are consistent with the theory of polygenic disease architecture shown previously by both simulation and real data for GWAS samples of similar size[56].

Migraine subtype analyses

To elucidate pathophysiological mechanisms underpinning the migraine aura, we performed a secondary analysis by creating two subsets that included only samples with the subtypes; migraine with aura and migraine without aura. These subsets only included those studies where sufficient information was available to assign a diagnosis of either subtype according to classification criteria standardized by the International Headache Society (IHS)[6]. For the population-based study samples this involved questionnaires, whereas for the clinic-based study samples the diagnosis was assigned on the basis of a structured interview by telephone or in person. A stricter diagnosis is required for these migraine subtypes as the migraine aura specifically is challenging to distinguish from other neurological features that can present as symptoms from unrelated conditions.

As a result, the migraine subtype analyses consisted of considerably smaller sample sizes compared to the main analysis (6,332 cases *vs.* 144,883 controls for migraine with aura and 8,348 cases *vs.* 139,622 controls for migraine without aura, see **Table 1**). As with the primary migraine analysis, the test statistics for migraine with aura or migraine without aura were consistent with underlying polygenic architecture rather than other potential sources of inflation (**Supplementary Figure 6 and 7**). For the migraine without aura subset analysis we found seven independent genomic loci (near *TSPAN2*, *TRPM8*, *PHACTR1*, *FHL5*, *ASTN2*, near *FGF6*, and *LRP1*) to be significantly associated (**Supplementary Table 11** and **Supplementary**

Figure 8). All seven of these loci were already identified in the primary analysis of 'all migraine' types, possibly reflecting the fact that migraine without aura is the most common form of migraine (around 2 in 3 cases) and likely drives the association signals in the primary analysis. Notably, no loci were associated to migraine with aura in the other subset analysis (**Supplementary Figure 9**).

To investigate whether excess heterogeneity could be contributing to the lack of associations in migraine with aura, we performed a heterogeneity analysis between the two subgroups. First we created two subsets of the migraine with aura and migraine without aura datasets from which none of the case or control individuals were overlapping (**Supplementary Table 12**). Then we selected the 44 LD-independent SNPs associated from the primary analysis and used a random-effects model to combine the migraine with aura and migraine without aura samples in a meta-analysis that allows for heterogeneity between the two migraine groups[57]. We found little heterogeneity with only seven of the 44 SNPs (at *REST*, *MPPED2*, *PHACTR1*, *ASTN2*, *MEF2D*, *PLCE1*, and *MED14*) exhibiting some signs of heterogeneity across subtype groups (**Supplementary Table 13**).

Credible sets of markers within each locus

For each of the 38 migraine-associated loci, we defined a credible set of markers that could plausibly be considered as causal using a Bayesian-likelihood based approach[58]. This method incorporates evidence from association test statistics and the LD structure between SNPs in a locus (**Online Methods**). A list of the credible set SNPs obtained for each locus is provided in **Supplementary Table 14**. We found three instances (in *RNF213*, *PLCE1*, and *MRVI1*) where the association signal could be credibly attributed to exonic missense polymorphisms (**Supplementary Table 15**). However, most of the credible markers at each locus were either intronic or intergenic, which is consistent with the theory that most variants detected by GWA studies involve regulatory effects on gene expression rather than disrupting protein structure[59,60].

Overlap with eQTLs in specific tissues

To try to identify specific migraine loci that might influence gene expression, we used previously published datasets that catalog expression quantitative trait loci (eQTLs) in either of two microarray-based studies from peripheral venous blood (N_1 = 3,754) or from human brain cortex tissue (N_2 = 550). Additionally, we used a third study based on RNAseq data from a collection of

42 tissues and three cell lines (N_3 = 1,641) from the Genotype-Tissue Expression (GTEx) consortium[61]. While this data has the advantage of a diverse tissue catalog, the number of samples per tissue is relatively small (**Supplementary Table 16**) compared to the two microarray datasets, possibly resulting in reduced power to detect significant eQTLs in some tissues. Using these datasets we applied a method based on the overlap of migraine and eQTL credible sets to identify eQTLs that could explain associations at the 38 migraine loci (**Online Methods**). This approach merged the migraine credible sets defined above with credible sets from *cis*-eQTL signals within a 1-Mb window and tested if the association signals between the migraine and eQTL credible sets were correlated. After adjusting for multiple testing we found no plausible eQTL associations in the peripheral blood or brain cortex data (**Supplementary Tables 17-18 and Supplementary Figure 10**). In GTEx, however, we found evidence for overlap from eQTLs in three tissues (Lung, Tibial Artery, and Aorta) at the *HPSE2* locus and in one tissue (Thyroid) at the *HEY2* locus (**Supplementary Table 19** and **Supplementary Figure 15**).

In summary, from three datasets we implicate eQTL signals at only two loci (*HPSE2, and HEY2).* This low number (two out of 38) is consistent with previous studies which have observed that available eQTL catalogues currently lack sufficient tissue specificity and developmental diversity to provide enough power to provide meaningful biological insight[52]. No plausibly causal eQTLs were observed in expression data from brain.

Gene expression enrichment in specific tissues

To understand if the 38 migraine loci as a group are enriched for expression in certain tissue groups, we again used the GTEx pilot data[61]. This time we tested whether genes near to credibly causal SNPs at the 38 migraine loci were significantly enriched for expression in certain tissues (**Online Methods**). We found four tissues that were significantly enriched (after Bonferroni correction) for expression of the migraine genes (**Figure 2**). The two most strongly enriched tissues were part of the cardiovascular system; the *aorta* and *tibial artery*. Two other significant tissues were from the digestive system; *esophagus muscularis* and *esophageal mucosa.* We replicated these enrichment results in an independent dataset using a component of the DEPICT[62] tool that conducts a tissue-specific enrichment analysis on microarray-based gene expression data (**Supplementary Methods**). DEPICT highlighted four tissues (**Figure 3 and Supplementary Table 20**) with significant enrichment of genes within the migraine loci;

arteries ($P = 1.58 \times 10^{-5}$), the upper gastrointestinal tract ($P = 2.97 \times 10^{-3}$), myometrium ($P = 3.03 \times 10^{-3}$), and stomach ($P = 3.38 \times 10^{-3}$).

Taken together, the expression analyses implicate arterial and gastrointestinal (GI) tissues. To discover if this enrichment signature could be attributed to a more specific type of smooth muscle, we examined the expression of the nearest genes at migraine loci in a panel of 60 types of human smooth muscle tissue[63]. Overall, migraine loci genes were not significantly enriched in a particular class of smooth muscle (**Supplementary Figures 11-13**). This suggests that the enrichment of migraine disease variants in genes expressed in tissues with a smooth muscle component is not specific to blood vessels, the stomach or GI tract, but rather appears to be generalizable across vascular and visceral smooth muscle types.

Combined, these results suggest that some of the genes affected by migraine-associated variants are highly expressed in vascular tissues and their dysfunction could play a role in migraine. Furthermore, the enrichment results suggest that other tissue types (e.g. smooth muscle) could also play a role and this may become evident once more migraine loci are discovered.

Enrichment in tissue-specific enhancers

To further assess the hypothesis that migraine variants might operate via effects on gene-regulation, we investigated the degree of overlap with histone modifications. We identified candidate causal variants underlying the 38 migraine loci, and examined their enrichment within cell-type specific enhancers from 56 primary human tissues and cell types from the Roadmap Epigenomics[64] and ENCODE projects[65] (**Online Methods** and **Supplementary Table 21**). Candidate causal variants showed highest enrichment in tissues from the mid-frontal lobe and duodenum smooth muscle, but these enrichments were not significant after adjusting for multiple testing (**Figure 4**).

Gene set enrichment analyses

To implicate underlying biological pathways involved in migraine, we applied a Gene Ontology (GO) over-representation analysis of the 38 migraine loci (**Online Methods**). We found nine vascular-related biological function categories that are significantly enriched after correction for multiple testing (**Supplementary Table 22**). Interestingly, we found little statistical support from the identified loci for some molecular processes that have been previously linked to migraine,

e.g. ion homeostasis, glutamate signaling, serotonin signaling, nitric oxide signaling, and oxidative stress (**Supplementary Table 23**). However, it is possible that the lack of enrichment for these functions may be explained by recognizing that current annotations for many genes and pathways are far from comprehensive, or that larger numbers of migraine loci need to be identified before we have sensitivity to detect enrichment in these mechanisms.

For a more comprehensive pathway analysis we used DEPICT, which incorporates gene co-expression information from microarray data to implicate additional, functionally less well-characterized genes in known biological pathways, protein-protein complexes and mouse phenotypes[62] (by forming so-called 'reconstituted gene sets'). From DEPICT we identified 67 reconstituted gene sets that are significantly enriched (FDR < 5%) for genes found among the 38 migraine associated loci (**Supplementary Table 24**). Because the reconstituted gene sets had genes in common, we clustered them into 10 distinct groups of gene sets **(Figure 5 and Online Methods).** Several gene sets, including the most significantly enriched reconstituted gene set (*Abnormal Vascular Wound Healing*; $P = 1.86 \times 10^{-6}$), were grouped into clusters related to cell-cell interactions (*ITGB1 PPI*, *Adherens Junction, Integrin Complex*). Several of the other gene set clusters were also related to vascular-biology (**Figure 5 and Supplementary Table 24**).

Discussion

In what is the largest genetic study of migraine to date, we identified 38 distinct genomic loci harboring 44 independent susceptibility markers for the prevalent forms of migraine. We provide evidence that migraine-associated genes are involved both in arterial and smooth muscle function. Two separate analyses, the DEPICT and the GTEx gene-expression enrichment analyses, point to vascular and smooth muscle tissues being involved in common variant susceptibility to migraine. The vascular finding is consistent with known co-morbidities and previously reported shared polygenic risk between migraine, stroke and cardiovascular diseases[66,67]. Furthermore, a recent GWA study of Cervical Artery Dissection (CeAD) identified a genome-wide significant association at exactly the same index SNP (rs9349379) as is associated to migraine in the *PHACTR1* locus, suggesting the possibility of partially shared genetic components between migraine and CeAD[26]. These results suggest that vascular dysfunction and possibly also other smooth muscle dysfunction likely play roles in migraine pathogenesis.

The support for vascular and smooth muscle enrichment of the loci is strong, with multiple lines of evidence from independent methods and independent datasets. However, it remains likely that neurogenic mechanisms are also involved in migraine. For example, several lines of evidence from previous studies have pointed to such mechanisms[5,68–71]. We found some support for this when looking at gene expression of individual genes at the 38 loci (**Supplementary Figure 2** and **Supplementary Table 25**), where many specific genes were active in brain tissues. While we did not observe statistically significant enrichment in brain across all loci, it may be that more associated loci are needed to detect this. Alternatively, it could be due to difficulties in collecting appropriate brain tissue samples with enough specificity, or other technical challenges. Additionally, there is less clarity of the biological mechanisms for a brain disease like migraine compared to some other common diseases, e.g. autoimmune or cardio-metabolic diseases where intermediate risk factors and underlying mechanisms are better understood.

Interestingly, some of the analyses highlight gastrointestinal tissues. Although migraine attacks may include gastrointestinal symptoms (e.g. nausea, vomiting, diarrhea)[72] it is likely that the signals observed here broadly represent smooth muscle signals rather than gastrointestinal specificity. Smooth muscle is a predominant tissue of the intestine, yet specific smooth muscle subtypes were not available to test this hypothesis in our primary enrichment analyses. We showed instead in a range of 60 smooth muscle subtypes, that the migraine loci are expressed in many types of smooth muscle, including vascular (**Supplementary Figure 12 and 13**). These results, while not conclusive, suggest that the enrichment of the migraine loci in smooth muscle is not specific to the stomach and GI tract.

Our results implicate cellular pathways and provide an opportunity to determine whether the genomic data supports previously presented hypotheses of pathways linked to migraine. One prevailing hypothesis stimulated by findings in familial hemiplegic migraine (FHM) has been that migraine is a channelopathy[5,21]. Among the 38 migraine loci only two harbor known ion channels (*KCNK5*[19] and *TRPM8*[20]), while three additional loci (*SLC24A3*[22], *ITPK1*[23], and *GJA1*[24]) can be linked to ion homeostasis. This further supports the findings of previous studies that in common forms of migraine, ion channel dysfunction is not the major pathophysiological mechanism[15]. However, more generally, genes involved in ion homeostasis could be a component of the genetic susceptibility. Moreover, we cannot exclude that ion channels could still be important contributors in migraine with aura, the form most closely resembling FHM, as our ability to

identify loci in this subgroup is more challenging. Another suggested hypothesis relates to oxidative stress and nitric oxide (NO) signaling[73–75]. Six genes with known links to oxidative stress and NO, within these 38 loci were identified (*REST*[44], *GJA1*[45], *YAP1*[46], *PRDM16*[47], *LRP1*[48], and *MRVI1*[49]). This is in line with previous findings[11], however, the DEPICT pathway analysis observed no association between NO-related reconstituted gene sets and migraine (*FDR* > 0.54, **Supplementary Table 23**).

Notably, in the migraine subtype analyses, it was possible to identify specific loci for migraine without aura but not for migraine with aura. However, the heterogeneity analysis (**Supplementary Tables 12-13**) demonstrated that most of the identified loci are implicated in both migraine subtypes. This suggests that no loci were identified in the migraine with aura analysis mainly due to lack of power from the reduced sample size. Additionally, as shown by the LD score analysis (**Supplementary Figures 5-7**), the amount of heritability captured by the migraine with aura dataset is considerably lower than migraine without aura, such that in order to reach comparable power, a sample size of two- to three-times larger would be required. This may reflect a higher degree of heterogeneity in the clinical capture, more complex underlying biology, or even a larger contribution from low-frequency and rare variation to migraine risk for this form of the disease.

In conclusion, the 38 genomic loci identified in this study support the notion that factors in vascular and smooth muscle tissues contribute to migraine pathophysiology and that the two major subtypes of migraine, migraine with aura and migraine without aura, have a partially shared underlying genetic susceptibility profile.

URLs

1000 Genomes Project, http://www.1000genomes.org/; BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html; DEPICT, www.broadinstitute.org/mpg/depict; Fine-mapping loci with credible sets, https://github.com/hailianghuang/FM-summary; GTEx, www.gtexportal.org; GWAMA, http://www.well.ox.ac.uk/gwama/; IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; International Headache Genetics Consortium, http://www.headachegenetics.org/; MACH, http://www.sph.umich.edu/csg/abecasis/MACH/tour/imputation.html; matSpD, http://neurogenetics.qimrberghofer.edu.au/matSpD; MINIMAC, http://genome.sph.umich.edu/wiki/Minimac; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; ProbABEL, http://www.genabel.org/packages/ProbABEL; R, https://www.r-project.org/; Roadmap Epigenomics Project, http://www.roadmapepigenomics.org/; SHAPEIT, http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.v778.html; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

Acknowledgments

We would like to thank the numerous individuals who contributed to sample collection, storage, handling, phenotyping and genotyping within each of the individual cohorts. We also thank the important contribution to research made by the study participants. We are grateful to Huiying Zhao (QIMR Berghofer Medical Research Institute) for helpful correspondence on the pathway analyses. We acknowledge the support and contribution of pilot data from the GTEx consortium. A list of study-specific acknowledgements can be found in the Supplementary Note.

Author Contributions

P.G., V.An., G.W.M., M.Ku., M.Kals., R.Mäg., K.P., E.H., E.L., A.G.U., L.C., E.M., L.M., A-L.E., A.F.C., T.F.H., A.J.A., D.I.C., and D.R.N. performed the experiments. P.G., V.An., B.S.W., P.P., T.E., T.H.P., K-H.F., M.Mu., N.A.F., A.I., G.McM., L.L., S.G.G., S.St., L.Q., H.H.H.A., D.A.H., J-J.H., R.Mal., A.E.B., E.S., C.M.v.D., E.M., D.P.S., N.E., B.M.N., D.I.C., and D.R.N. performed the statistical analyses. P.G., V.An., B.S.W., P.P., T.E., T.H.P., K-H.F., E.C-L., N.A.F., A.I., G.McM., L.L., M.Kall., T.M.F., S.G.G., S.St., M.Ko., L.Q., H.H.H.A., T.L., J.W., D.A.H., S.M.R.,

M.F., V.Ar., M.Kau., S.V., R.Mal., M.Ku., M.Kals., R.Mäg., K.P., H.H., A.E.B., J.H., E.S., C.S., C.W., Z.C., K.H., E.L., L.M.P, A-L.E., A.F.C., T.F.H., J.K., A.J.A., O.R., M.A.I., M-R.J., D.P.S., M.W., G.D.S., N.E., M.J.D., B.M.N., J.O., D.I.C., D.R.N., and A.P. participated in data analysis/interpretation. P.G., V.An., B.S.W., T.H.P., K-H.F., E.C-L., T.K., G.M.T, M.Kall., C.R., A.H.S., G.B., M.Ko., T.L., M.S., M.G.H., M.F., V.Ar., M.Kau., S.V., R.Mal., A.C.H., P.A.F.M., N.G.M., G.W.M., H.H., A.E.B., L.F., J.H., P.H.L., C.S., C.W., Z.C., B.M-M., S.Sc., T.M., J.G.E., V.S., A.G.U., C.M.v.D., A.S., C.S.N., H.G., A-L.E., A.F.C., T.F.H., T.W., A.J.A., O.R., M-R.J., C.K., M.D.F., A.C.B., M.D., M.W., J-A.Z., B.M.N., J.O., D.I.C., D.R.N., and A.P. contributed materials/analysis tools. T.E., T.K., T.L., H.S., B.W.J.H.P., A.C.H., P.A.F.M., N.G.M., G.W.M., L.F., A.H., A.S., C.S.N., M.Mä., T.W., J.K., O.R., M.A.I., T.S., M-R.J., A.M., C.K., D.P.S., M.D.F., A.M.J.M.v.d.M., J-A.Z., D.I.B., G.D.S., K.S., N.E., B.M.N., J.O., D.I.C., D.R.N., and A.P. supervised the research. T.K., G.M.T, G.B., T.L., J.E.B., M.S., P.M.R., H.S., B.W.J.H.P., A.C.H., P.A.F.M., N.G.M., G.W.M., L.F., V.S., A.H., L.C., A.S., C.S.N., H.G., J.K., A.J.A., O.R., M.A.I., M-R.J., A.M., C.K., D.P.S., M.D., A.M.J.M.v.d.M., D.I.B., G.D.S., N.E., M.J.D., B.M.N., D.I.C., D.R.N., and A.P. conceived and designed the study. P.G., V.An., B.S.W., P.P., T.E., T.H.P., E.C-L., H.H., B.M.N., J.O., D.I.C., D.R.N., and A.P. wrote the paper. All authors contributed to the final version of the manuscript.

Data access

All genome-wide significant and suggestive SNP associations ($P < 1 \times 10^{-5}$) from the meta-analysis can be obtained directly from the IHGC website (http://www.headachegenetics.org/). For access to deeper-level data please contact the data access committee (fimm-dac@helsinki.fi).

References

1. Vos, T. *et al.* Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2163–2196 (2012).

2. Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* (2015). doi:10.1016/S0140-6736(15)60692-4

3. Gustavsson, A. *et al.* Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21,** 718–779 (2011).

4. Pietrobon, D. & Striessnig, J. Neurological diseases: Neurobiology of migraine. *Nature Reviews Neuroscience* **4,** 386–398 (2003).

5. Tfelt-Hansen, P. C. & Koehler, P. J. One hundred years of migraine research: Major clinical and scientific observations from 1910 to 2010. *Headache* **51,** 752–778 (2011).

6. Society, H. C. C. of the I. H. The International Classification of Headache Disorders: 2nd edition. *Cephalalgia* **24,** 1–160 (2004).

7. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).

8. Anttila, V. *et al.* Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. *Nat. Genet.* **42,** 869–873 (2010).

9. Chasman, D. I. *et al.* Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nat Genet* **43,** 695–698 (2011).

10. Freilinger, T. *et al.* Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat. Genet.* **44,** 777–782 (2012).

11. Anttila, V. *et al.* Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.* **45,** 912–7 (2013).

12. Ophoff, R. A. *et al.* Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. *Cell* **87,** 543–552 (1996).

13. De Fusco, M. *et al.* Haploinsufficiency of ATP1A2 encoding the Na+/K+ pump alpha2 subunit associated with familial hemiplegic migraine type 2. *Nat. Genet.* **33,** 192–196 (2003).

14. Dichgans, M. *et al.* Mutation in the neuronal voltage-gated sodium channel SCN1A in familial hemiplegic migraine. *Lancet* **366,** 371–377 (2005).

15. Nyholt, D. R. *et al.* A high-density association screen of 155 ion transport genes for involvement with common migraine. *Hum. Mol. Genet.* **17,** 3318–3331 (2008).

16. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

17. Chasman, D. I. *et al.* Selectivity in Genetic Association with Sub-classified Migraine in Women. *PLoS Genet.* **10,** (2014).

18. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88,** 586–598 (2011).

19. Morton, M. J., Abohamed, A., Sivaprasadarao, A. & Hunter, M. pH sensing in the two-pore domain K+ channel, TASK2. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16102–16106 (2005).

20. Ramachandran, R. *et al.* TRPM8 activation attenuates inflammatory responses in mouse models of colitis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7476–81 (2013).

21. Hanna, M. G. Genetic neurological channelopathies. *Nat. Clin. Pract. Neurol.* **2,** 252–263 (2006).

22. Kraev, A. *et al.* Molecular cloning of a third member of the potassium-dependent sodium-calcium exchanger gene family, NCKX3. *J. Biol. Chem.* **276,** 23161–72 (2001).

23. Ismailov, I. I. *et al.* A biologic function for an 'orphan' messenger: D-myo-inositol 3,4,5,6-tetrakisphosphate selectively blocks epithelial calcium-activated chloride channels. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10505–9 (1996).

24. De Bock, M. *et al.* Connexin channels provide a target to manipulate brain endothelial calcium dynamics and blood-brain barrier permeability. *J. Cereb. Blood Flow Metab.* **31,** 1942–1957 (2011).

25. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41,** 334–341 (2009).

26. Debette, S. *et al.* Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat. Genet.* **47,** 78–83 (2015).

27. Law, C. *et al.* Clinical features in a family with an R460H mutation in transforming growth factor beta receptor 2 gene. *J Med Genet* **43,** 908–916 (2006).

28. Bown, M. J. *et al.* Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. *Am. J. Hum. Genet.* **89,** 619–627 (2011).

29. Arndt, A. K. *et al.* Fine mapping of the 1p36 deletion syndrome identifies mutation of PRDM16 as a cause of cardiomyopathy. *Am. J. Hum. Genet.* **93,** 67–77 (2013).

30. Fujimura, M. *et al.* Genetics and Biomarkers of Moyamoya Disease: Significance of RNF213 as a Susceptibility Gene. *J. stroke* **16,** 65–72 (2014).

31. McElhinney, D. B. *et al.* Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome. *Circulation* **106,** 2567–2574 (2002).

32. Bezzina, C. R. *et al.* Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat. Genet.* **45,** 1044–9 (2013).

33. Sinner, M. F. *et al.* Integrating genetic, transcriptional, and functional analyses to identify five novel genes for atrial fibrillation. *Circulation* (2014). doi:10.1161/CIRCULATIONAHA.114.009892

34. Neale, B. M. *et al.* Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc. Natl. Acad. Sci. U. S. A.* **107,** 7395–7400 (2010).

35. Desch, M. *et al.* IRAG determines nitric oxide- and atrial natriuretic peptide-mediated smooth muscle relaxation. *Cardiovasc. Res.* **86,** 496–505 (2010).

36. Lang, N. N., Luksha, L., Newby, D. E. & Kublickiene, K. Connexin 43 mediates endothelium-derived hyperpolarizing factor-induced vasodilatation in subcutaneous resistance arteries from healthy pregnant women. *Am. J. Physiol. Heart Circ. Physiol.* **292,** H1026–H1032 (2007).

37. Dong, H., Jiang, Y., Triggle, C. R., Li, X. & Lytton, J. Novel role for K+-dependent Na+/Ca2+ exchangers in regulation of cytoplasmic free Ca2+ and contractility in arterial smooth muscle. *Am. J. Physiol. Heart Circ. Physiol.* **291,** H1226–H1235 (2006).

38. Yamaji, M., Mahmoud, M., Evans, I. M. & Zachary, I. C. Neuropilin 1 is essential for gastrointestinal smooth muscle contractility and motility in aged mice. *PLoS One* **10,** e0115563 (2015).

39. Lu, X. *et al.* Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nature Genetics* **44,** 890–894 (2012).

40. Hager, J. *et al.* Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis. *PLoS One* **7,** (2012).

41. Coronary, T., Disease, A. & Consortium, G. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43,** 339–44 (2011).

42. Odonnell, C. J. *et al.* Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation* **124,** 2855–2864 (2011).

43. Porcu, E. *et al.* A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet.* **9,** e1003266 (2013).

44. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer disease. *Nature* **Epub ahead,** 448–54 (2014).

45. Kar, R., Riquelme, M. A., Werner, S. & Jiang, J. X. Connexin 43 channels protect osteocytes against oxidative stress-induced cell death. *J. Bone Miner. Res.* **28,** 1611–1621 (2013).

46. Dixit, D., Ghildiyal, R., Anto, N. P. & Sen, E. Chaetocin-induced ROS-mediated apoptosis involves ATM-YAP1 axis and JNK-dependent inhibition of glucose metabolism. *Cell Death Dis.* **5,** e1212 (2014).

47. Chuikov, S., Levi, B. P., Smith, M. L. & Morrison, S. J. Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. *Nat. Cell Biol.* **12,** 999–1006 (2010).

48. Castellano, J. *et al.* Hypoxia stimulates low-density lipoprotein receptor-related protein-1 expression through hypoxia-inducible factor-1α in human vascular smooth muscle cells. *Arterioscler. Thromb. Vasc. Biol.* **31,** 1411–1420 (2011).

49. Schlossmann, J. *et al.* Regulation of intracellular calcium by a signalling complex of IRAG, IP3 receptor and cGMP kinase Ibeta. *Nature* **404,** 197–201 (2000).

50. Nalls, M. a *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **056,** 1–7 (2014).

51. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–8 (2013).

52. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

53. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–86 (2014).

54. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

55. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

56. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2011).

57. Magi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34,** 846–853 (2010).

58. Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44,** 1294–301 (2012).

59. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet.* **6,** (2010).

60. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337,** 1190–1195 (2012).

61. Consortium, T. G. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–5 (2013).

62. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).

63. Chi, J. T. *et al.* Gene expression programs of human smooth muscle cells: Tissue-specific differentiation and prognostic significance in breast cancers. *PLoS Genet.* **3,** 1770–1784 (2007).

64. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28,** 1045–1048 (2010).

65. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

66. Winsvold, B. S. *et al.* Genetic analysis for a shared biological basis between migraine and coronary artery disease. *Neurol. Genet.* **1,** e10–e10 (2015).

67. Malik, R. *et al.* Shared genetic basis for migraine and ischemic stroke: A genome-wide analysis of common variants. *Neurology* **84,** 2132–45 (2015).

68. Ferrari, M. D., Klever, R. R., Terwindt, G. M., Ayata, C. & van den Maagdenberg, A. M. J. M. Migraine pathophysiology: lessons from mouse models and human genetics. *Lancet. Neurol.* **14,** 65–80 (2015).

69. Olesen, J., Burstein, R., Ashina, M. & Tfelt-Hansen, P. Origin of pain in migraine: evidence for peripheral sensitisation. *Lancet Neurol.* **8,** 679–690 (2009).

70. Hadjikhani, N. *et al.* Mechanisms of migraine aura revealed by functional MRI in human visual cortex. *Proc. Natl. Acad. Sci.* **98,** 4687 –4692 (2001).

71. Lauritzen, M. Pathophysiology of the migraine aura. The spreading depression theory. *Brain* **117 (Pt 1,** 199–210 (1994).

72. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia* **33,** 629–808 (2013).

73. Olesen, J. The role of nitric oxide (NO) in migraine, tension-type headache and cluster headache. *Pharmacol Ther* **120,** 157–171 (2008).

74. Ashina, M., Hansen, J. M. & Olesen, J. Pearls and pitfalls in human pharmacological models of migraine: 30 years' experience. *Cephalalgia* **33,** 540–53 (2013).

75. Read, S. J. & Parsons, A. A. Sumatriptan modifies cortical free radical release during cortical spreading depression: A novel antimigraine action for sumatriptan? *Brain Res.* **870,** 44–53 (2000).

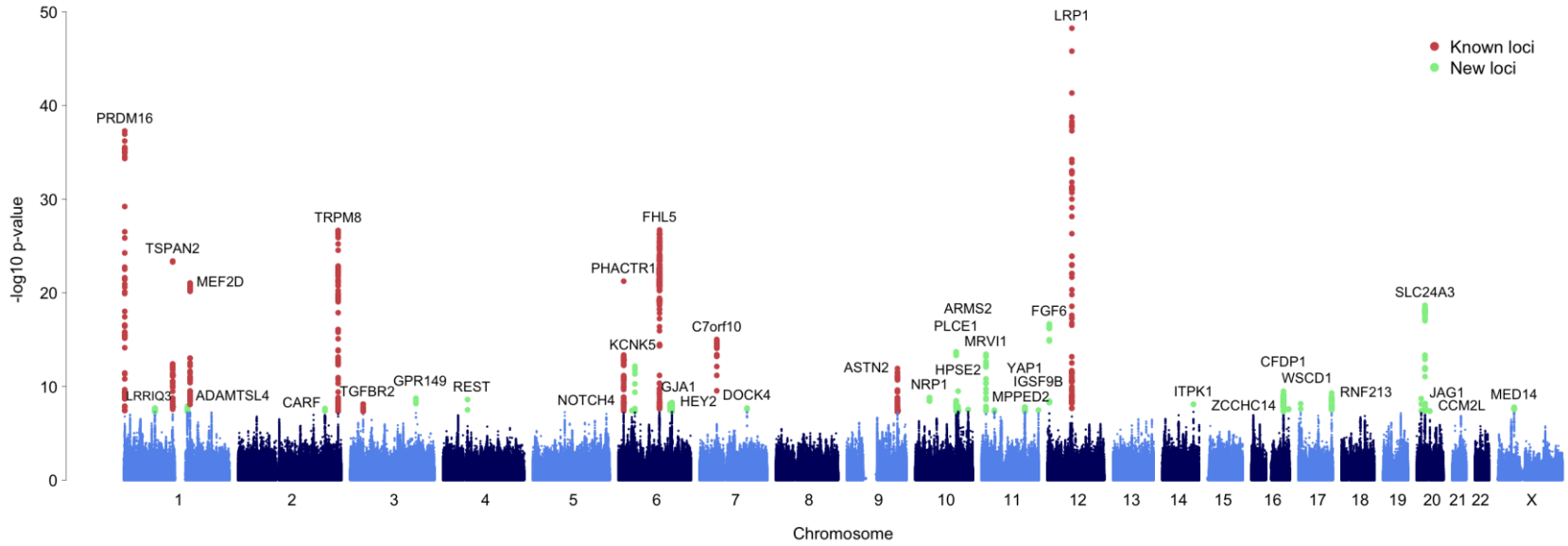


Figure 1. Manhattan plot of the primary meta-analysis of all migraine (59,674 cases *vs.* 316,078 controls). Each marker was tested for association using an additive genetic model by logistic regression adjusted for sex. A fixed-effects meta-analysis was then used to combine the association statistics from all 22 clinic and population-based studies from the IHGC. The horizontal axis shows the chromosomal position and the vertical axis shows the significance of tested markers from logistic regression. Markers with test statistics that reach genome-wide significance ($P < 5 \times 10^{-8}$) at previously known and newly identified loci are highlighted according to the color legend.

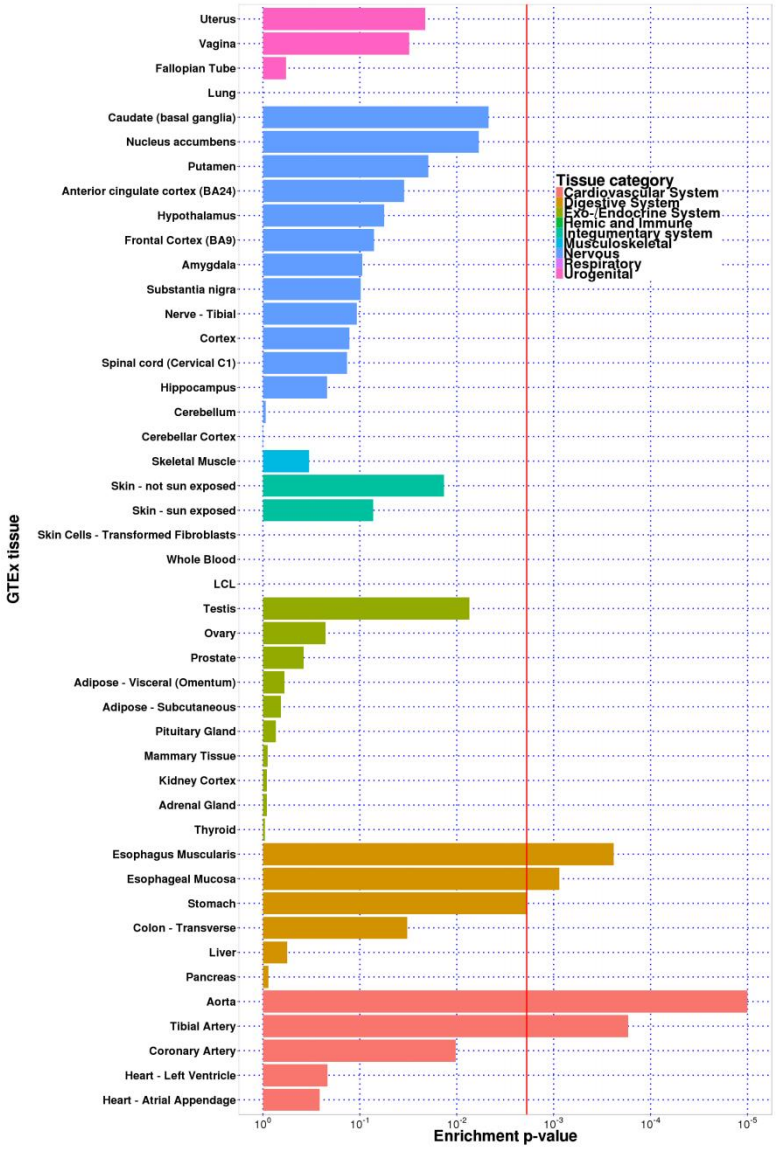


Figure 2. Gene expression enrichment of genes from the 38 migraine loci in GTEx tissues. Expression data from 1,641 samples was obtained using RNAseq for 42 tissues and three cell lines from the GTEx consortium. Enrichment *P*-values were assessed empirically for each tissue using a permutation procedure (100,000 replicates) and the red vertical line shows the significance threshold after adjusting for multiple testing by Bonferroni correction (see **Online Methods**).

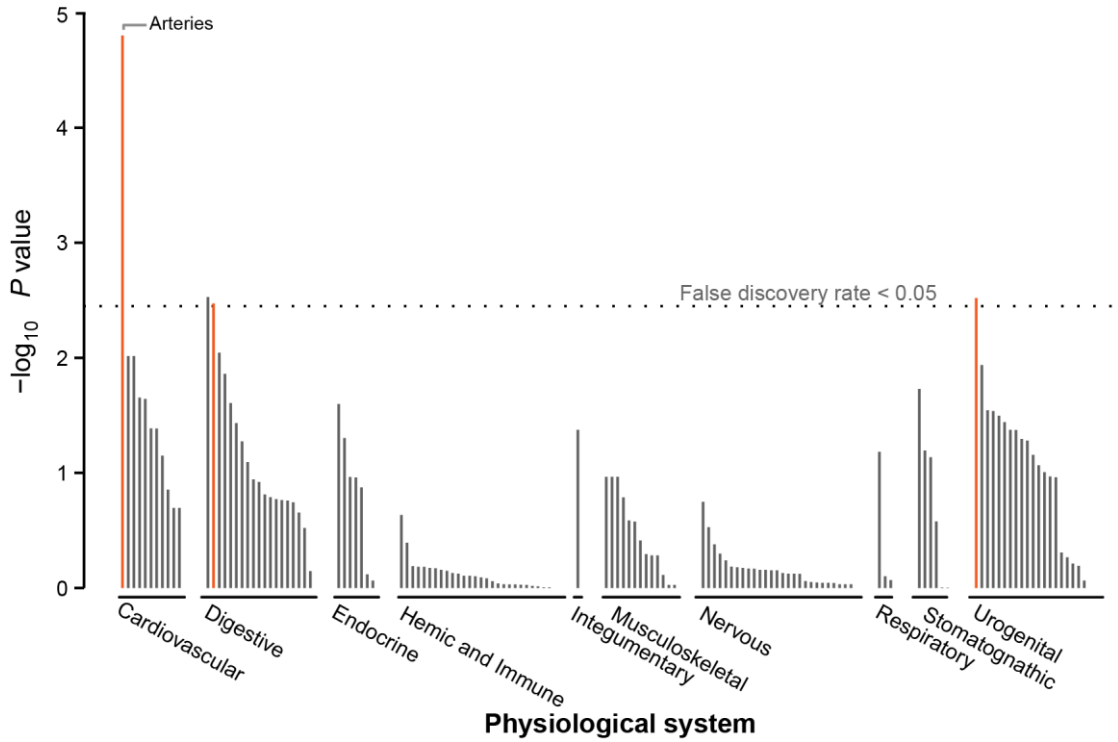
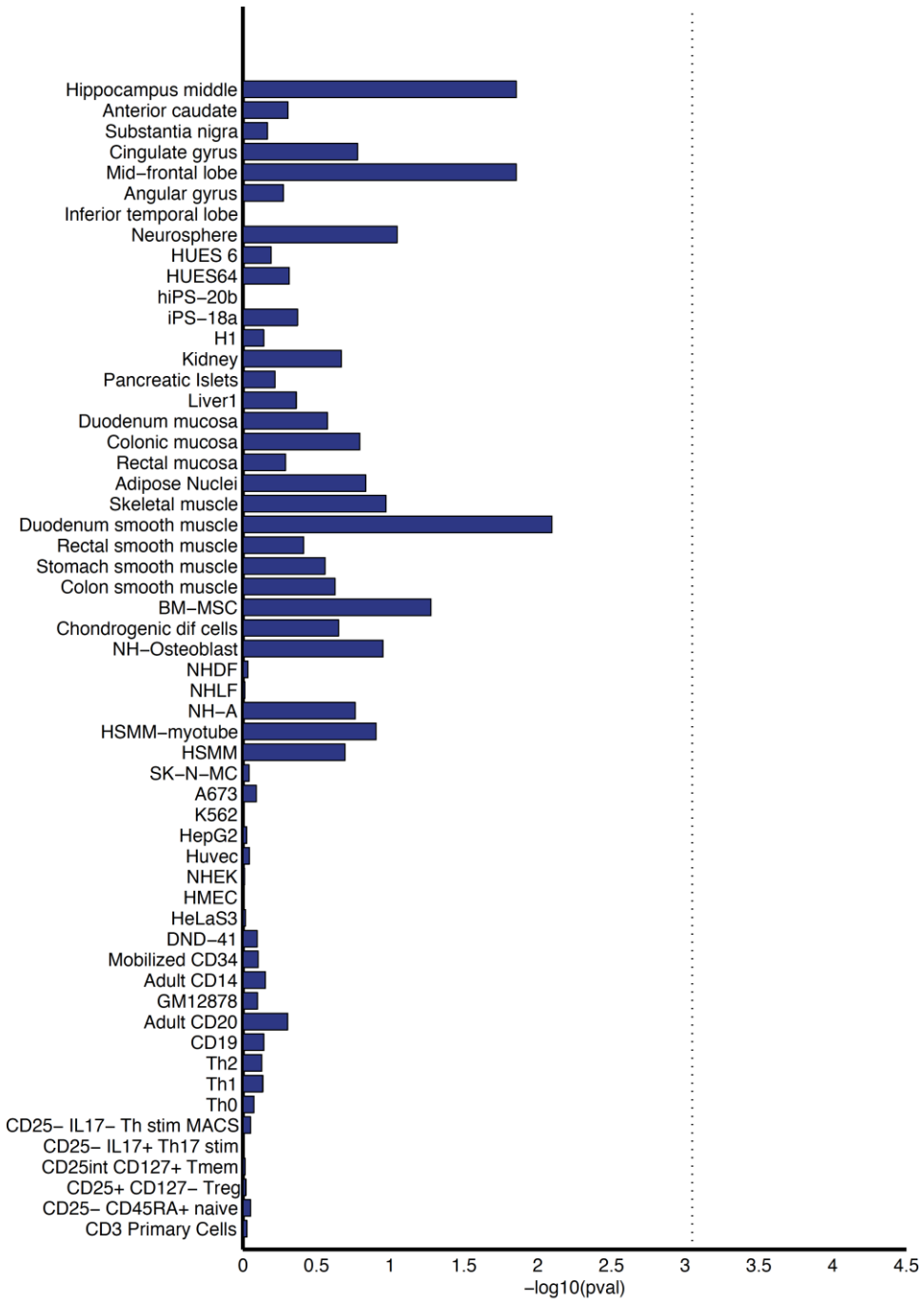
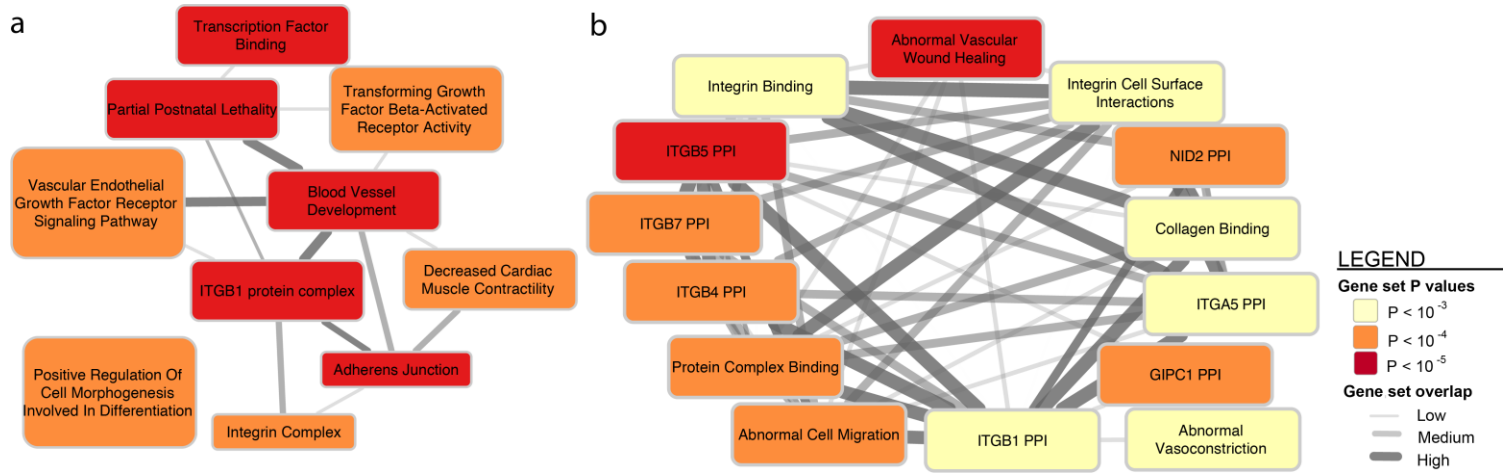


Figure 3. Gene expression enrichment of genes from the 38 migraine loci in 209 tissue/cell type annotations by DEPICT. Expression data was obtained from 37,427 human microarray samples and then genes in the migraine loci were assessed for high expression in each of the annotation categories. Enrichment *P*-values were determined by comparing the expression pattern from the migraine loci to 500 randomly generated loci and the false discovery rate (horizontal dashed line) was estimated to control for multiple testing (see **Online Methods**). A full list of these enrichment results are provided in **Supplementary Table 20**.



715

716 **Figure 4.** Enrichment of the migraine loci in sets of tissue-specific enhancers. We mapped
717 credible sets from the migraine loci to sets of enhancers under active expression in 56 tissues
718 and cell lines (identified by H3K27ac histone marks from the Roadmap Epigenomics[64] and
719 ENCODE[65] projects). Enrichment *P*-values were assessed empirically by randomly generating a
720 background set of matched loci for comparison (10,000 replicates) and the vertical dotted line is
721 the significance threshold after adjusting for 56 separate tests by Bonferroni correction (see
722 **Online Methods**).



723

724 **Figure 5.** DEPICT network of the reconstituted gene sets that were found to be significantly enriched (false discovery rate < 0.05) for
725 genes at the migraine loci (**Online Methods**). Enriched gene sets are represented as nodes with pairwise overlap denoted by the
726 width of the connecting lines and empirical enrichment *P*-value is indicated by color intensity (darker is more significant). The 67
727 significantly enriched gene sets were then clustered by similarity into 10 group nodes as shown in **(a)** where each group node is
728 named after the most representative gene set in the group. **(b)** Shows one example of the enriched reconstituted gene sets that were
729 clustered within the now expanded *ITGB1 PPI* group. A full list of the 67 significantly enriched reconstituted gene sets can be found
730 in **Supplementary Table 24**.

Table 1. Individual IHGC GWA studies listed with cases and control numbers used in the primary analysis (all migraine) and in the subtype analyses (migraine with aura and migraine without aura). Note that chromosome X genotype data was unavailable from three of the individual GWA studies (EGCUT, Rotterdam III, and TwinsUK) and also partially unavailable from some of the control samples (specifically the GSK controls) used for the 'German MO' study, meaning that the number of samples analyzed on chromosome X was 57,756 cases and 299,109 controls. Complete data was available on the autosomes for all samples.

GWA Study ID	Full Name of GWA Study	All migraine		Migraine with aura		Migraine without aura	
		Cases	Controls	Cases	Controls	Cases	Controls
23andMe	23andMe Inc.	30,465	143,147	-	-	-	-
ALSPAC	Avon Longitudinal Study of Parents and Children	3,134	5,103	-	-	-	-
ATM	Australian Twin Migraine	1,683	2,383	-	-	-	-
B58C	1958 British Birth Cohort	1,165	4,141	-	-	-	-
Danish HC	Danish Headache Center	1,771	1,000	775	1,000	996	1,000
DeCODE	deCODE Genetics Inc.	3,135	95,585	366	95,585	608	95,585
Dutch MA	Dutch migraine with aura	734	5,211	734	5,211	-	-
Dutch MO	Dutch migraine without aura	1,115	2,028	-	-	1,115	2,028
EGCUT	Estonian Genome Center, University of Tartu	813	9,850	76	9,850	94	9,850
Finnish MA	Finnish migraine with aura	933	2,715	933	2,715	-	-
German MA	German migraine with aura	1,071	1,010	1,071	1,010	-	-
German MO	German migraine without aura	1,160	1,647	-	-	1,160	1,647

Health 2000	Health 2000	136	1,764	-	-	-	-
HUNT	Nord-Trøndelag Health Study	1,395	1,011	290	1,011	980	1,011
NFBC	Northern Finnish Birth Cohort	756	4,393	-	-	-	-
NTR/NESDA	Netherlands Twin Register and the Netherlands Study of Depression and Anxiety	1,636	3,819	544	3,819	615	3,819
Rotterdam III	Rotterdam Study III	487	2,175	106	2,175	381	2,175
Swedish Twins	Swedish Twin Registry	1,307	4,182	-	-	-	-
Tromsø	The Tromsø Study	660	2,407	-	-	-	-
Twins UK	Twins UK	618	2,334	202	2,334	416	2,334
WGHS	Women's Genome Health Study	5,122	18,108	1,177	18,108	1,826	18,108
Young Finns	Young Finns	378	2,065	58	2,065	157	2,065
	Total:	**59,674**	**316,078**	**6,332**	**144,883**	**8,348**	**139,622**

737

Table 2. Summary of the 38 genomic loci associated with the prevalent types of migraine. Ten loci were previously reported (PubMed IDs listed) and 28 are newly found in this study. For each locus, the nearest coding gene to the index SNP is given. Effect sizes and *P*-values for each SNP were calculated for each study with an additive genetic model using logistic regression adjusted for sex and then combined in a fixed-effects meta-analysis. For loci that contain a secondary LD-independent signal passing genome-wide significance, the secondary index SNP and P-value is given. For the seven loci reaching genome-wide significance in the migraine without aura sub-type analysis, the corresponding index SNP and P-value are also given. Evidence for significant heterogeneity was found at four loci (*TRPM8*, *MRVI1*, *ZCCHC14*, and *CCM2L*) so for those we present the results of a random-effects model.

Locus Rank	Nearest coding gene	Chr	Index SNP	Minor Allele	MAF	All Migraine		Secondary signal		Migraine without aura		Previous Publication PMID
						OR [95% CI]	P	Index SNP	P	Index SNP	P	
1	*LRP1*	12	rs11172113	C	0.42	0.90 [0.89-0.91]	5.6×10^{-49}	rs7961602	2.1×10^{-11}	rs11172113	4.3×10^{-16}	21666692
2	*PRDM16*	1	rs10218452	G	0.22	1.11 [1.10-1.13]	5.3×10^{-38}	rs12135062	3.7×10^{-10}	-	-	21666692
3	*FHL5*	6	rs67338227	T	0.23	1.09 [1.08-1.11]	2.0×10^{-27}	rs2223239	3.2×10^{-10}	rs7775721	1.1×10^{-12}	23793025
4	*TSPAN2*	1	rs2078371	C	0.12	1.11 [1.09-1.13]	4.1×10^{-24}	rs7544256	8.7×10^{-09}	rs2078371	7.4×10^{-09}	23793025
5	*TRPM8*	2	rs10166942	C	0.20	0.94 [0.89-0.99]	1.0×10^{-23}	rs566529	2.5×10^{-09}	rs6724624	1.1×10^{-09}	21666692
6	*PHACTR1*	6	rs9349379	G	0.41	0.93 [0.92-0.95]	5.8×10^{-22}	-	-	rs9349379	2.1×10^{-09}	22683712
7	*MEF2D*	1	rs1925950	G	0.35	1.07 [1.06-1.09]	9.1×10^{-22}	-	-	-	-	22683712
8	*SLC24A3*	20	rs4814864	C	0.26	1.07 [1.06-1.09]	2.2×10^{-19}	-	-	-	-	-
9	*FGF6*	12	rs1024905	G	0.47	1.06 [1.04-1.08]	2.1×10^{-17}	-	-	rs1024905	2.5×10^{-09}	-
10	*C7orf10*	7	rs186166891	T	0.11	1.09 [1.07-1.12]	9.7×10^{-16}	-	-	-	-	23793025
11	*PLCE1*	10	rs10786156	G	0.45	0.95 [0.94-0.96]	2.0×10^{-14}	rs75473620	5.8×10^{-09}	-	-	-
12	*KCNK5*	6	rs10456100	T	0.28	1.06 [1.04-1.07]	6.9×10^{-13}	-	-	-	-	-
13	*ASTN2*	9	rs6478241	A	0.36	1.05 [1.04-1.07]	1.2×10^{-12}	-	-	rs6478241	1.2×10^{-10}	22683712
14	*MRVI1*	11	rs4910165	C	0.33	0.94 [0.91-0.98]	2.9×10^{-11}	-	-	-	-	-
15	*HPSE2*	10	rs12260159	A	0.07	0.92 [0.89-0.94]	3.2×10^{-10}	-	-	-	-	-
16	*CFDP1*	16	rs77505915	T	0.45	1.05 [1.03-1.06]	3.3×10^{-10}	-	-	-	-	-
17	*RNF213*	17	rs17857135	C	0.17	1.06 [1.04-1.08]	5.2×10^{-10}	-	-	-	-	-
18	*NRP1*	10	rs2506142	G	0.17	1.06 [1.04-1.07]	1.5×10^{-09}	-	-	-	-	-

19	*GPR149*	3	rs13078967	C	0.03	0.87 [0.83-0.91]	1.8×10^{-09}	-	-	-	-	-
20	*JAG1*	20	rs111404218	G	0.34	1.05 [1.03-1.07]	2.0×10^{-09}	-	-	-	-	-
21	*REST*	4	rs7684253	C	0.45	0.96 [0.94-0.97]	2.5×10^{-09}	-	-	-	-	-
22	*ZCCHC14*	16	rs4081947	G	0.34	1.03 [1.00-1.06]	2.5×10^{-09}	-	-	-	-	-
23	*HEY2*	6	rs1268083	C	0.48	0.96 [0.95-0.97]	5.3×10^{-09}	-	-	-	-	-
24	*WSCD1*	17	rs75213074	T	0.03	0.89 [0.86-0.93]	7.1×10^{-09}	-	-	-	-	-
25	*GJA1*	6	rs28455731	T	0.16	1.06 [1.04-1.08]	7.3×10^{-09}	-	-	-	-	-
26	*TGFBR2*	3	rs6791480	T	0.31	1.04 [1.03-1.06]	7.8×10^{-09}	-	-	-	-	22683712
27	*ITPK1*	14	rs11624776	C	0.31	0.96 [0.94-0.97]	7.9×10^{-09}	-	-	-	-	-
28	*ADAMTSL4*	1	rs6693567	C	0.27	1.05 [1.03-1.06]	1.2×10^{-08}	-	-	-	-	-
29	*CCM2L*	20	rs144017103	T	0.02	0.85 [0.76-0.96]	1.2×10^{-08}	-	-	-	-	-
30	*YAP1*	11	rs10895275	A	0.33	1.04 [1.03-1.06]	1.6×10^{-08}	-	-	-	-	-
31	*MED14*	X	rs12845494	G	0.27	0.96 [0.95-0.97]	1.7×10^{-08}	-	-	-	-	-
32	*DOCK4*	7	rs10155855	T	0.05	1.08 [1.05-1.12]	2.1×10^{-08}	-	-	-	-	-
33	*LRRIQ3*	1	rs1572668	G	0.48	1.04 [1.02-1.05]	2.1×10^{-08}	-	-	-	-	-
34	*CARF*	2	rs138556413	T	0.03	0.88 [0.84-0.92]	2.3×10^{-08}	-	-	-	-	-
35	*ARMS2*	10	rs2223089	C	0.08	0.93 [0.91-0.95]	3.0×10^{-08}	-	-	-	-	-
36	*IGSF9B*	11	rs561561	T	0.12	0.94 [0.92-0.96]	3.4×10^{-08}	-	-	-	-	-
37	*MPPED2*	11	rs11031122	C	0.24	1.04 [1.03-1.06]	3.5×10^{-08}	-	-	-	-	-
38	*NOTCH4*	6	rs140002913	A	0.06	0.91 [0.88-0.94]	3.8×10^{-08}	-	-	-	-	-

Online Methods

Quality Control. The 22 individual GWA studies were subjected to pre-established quality control (QC) protocols as recommended elsewhere[76,77]. Differences in genotyping chips, DNA quality and genotype calling pipelines necessitated that QC parameters were tuned separately to be appropriate for each individual study. At a minimum, we excluded markers with excessively high missingness rates (> 5%), low minor allele frequency (< 1%), and failing a test of Hardy-Weinberg equilibrium. We also excluded individuals with a high proportion of missing genotypes (> 5%) and used identity-by-descent (IBD) estimates to remove individuals that were highly related (IBD>0.185) to others in the sample. A summary of the genotyping platforms, quality control, imputation protocols and association analysis methods used in each individual GWA study is provided in **Supplementary Table 3**. All case/control sets that were genotyped separately were first quality controlled independently and then again after merging the data.

To control for population stratification within each individual GWA study, we merged the genotypes passing quality control filters with HapMap III genotype data from three populations; European (CEU), Asian (CHB + JPT) and African (YRI). We then performed a principal components analysis on the merged dataset and excluded any (non-European) population outliers from our studies. To control for any further (sub-European) population structure, we performed a second principal components analysis on the genotype data from each GWA study separately to ensure that cases and controls were clustering together. We then tested whether any principal components were significantly associated with the phenotype using logistic regression. Any principal components that were significantly associated were then included as covariates in the model when generating the final association test statistics for the migraine meta-analysis. The specific principal components adjusted for in each individual GWA study are listed in **Supplementary Table 4**.

Imputation. Following GWA study-level QC, the data underwent a phasing step whereby haplotypes for each individual were statistically estimated using (in most instances) the program SHAPEIT[78]. Missing genotypes were then imputed into these haplotypes using the program IMPUTE2[79] and a mixed-population reference panel provided by the 1000 Genomes Project[16]. All study samples were imputed using the March 2012 (phase I, v3 release or later) 1000 Genomes reference panel. A minority of contributing GWA studies used alternative programs for phasing and imputation such as BEAGLE[80], MACH[81], and MINIMAC[82] or some in-house custom software. A full list of software and procedures used are provided in **Supplementary Table 3**.

Statistical Analysis. Individual study association analyses were implemented using logistic regression with an additive model on the imputed dosage of the effect allele. All models were adjusted for sex and other relevant covariates. Age information was not available for individuals from all studies therefore we were not able to adjust for it in our models. However, we note that all of the GWA studies were comprised of adults past the typical age of onset, hence age is at most a non-confounding factor and false positive rates would not be affected by its inclusion/exclusion. Furthermore, including such covariates can be sub-optimal, reducing power to detect genetic associations. To control for sub-European population structure, we also included in the model any principal components that were significantly associated with the phenotype (**Supplementary Table 4**). The programs used for performing the association analyses were either SNPTEST, PLINK or R (see URLs). To combine association summary statistics from all individual studies we used the program GWAMA (URLs) to perform a fixed-effects meta-analysis weighted by the inverse variance to obtain a combined effect size, standard error and p-value at each marker. We excluded markers in any individual study that had low imputation quality scores (IMPUTE2 *INFO* < 0.6 or MACH r^2 < 0.6) or low minor allele frequency (MAF < 0.01). Additionally, we filtered out any marker that was missing from more than half the individual studies (missing from 12 or more out of 22 studies) and also markers exhibiting high levels of heterogeneity as identified by a high heterogeneity index (i^2 > 0.75). After applying all filters, this left 8,045,569 total markers tested in the meta-analysis.

Chromosome X meta-analysis. Due to the different ploidy of males and females, the X chromosome required a different statistical model; we applied a model of X-chromosome inactivation (XCI) that assumes an equal effect of alleles in both males and females. This XCI model was achieved by scaling male dosages to the range 0-2 to match that of females. In total, 57,756 cases and 299,109 controls were available for the X-chromosome analysis (**Supplementary Table 1**). The reduced sample size compared to the autosomal data occurred because some of the individual GWA studies (EGCUT, Rotterdam III, Twins UK, and 846 controls from GSK for the 'German MO' study) did not contribute chromosome X data.

Defining Credible Sets. Within each migraine-associated locus, we defined a credible set of variants that could be considered 99% likely to contain a causal variant. The method has been introduced in detail elsewhere[52,58] and a full derivation is outlined again briefly in the Supplementary Note. This method assumes that there is one and only one causal variant in the

locus. For loci that contain a secondary independent signal, we conservatively mapped only the primary signal.

LD score regression analysis. We conducted a univariate heritability analysis based on summary statistics using LD score regression (LDSC) v1.0.0[55]. For this analysis, high-quality common SNPs were extracted from the summary statistics by filtering the data using the following criteria: presence among the HapMap Project Phase 3 SNPs[83], allele matching to 1000 Genomes data, no strand ambiguity, INFO score > 0.9, MAF >= 1%, and missingness less than two thirds of the 90th percentile of total sample size. The HLA region (chromosome 6, from 25 - 35 Mb) was excluded from the analysis. From this data, we used LDSC to quantify the proportion of the total inflation in chi-square statistics that can be ascribed to polygenic heritability by calculating the ratio of the LDSC intercept estimate and the chi-square mean using the formula described in the original publication[55].

eQTL Credible Set Overlap Analysis. To test whether the association statistics across each of the 38 migraine loci could be explained by credible overlapping eQTL signals, we used two previously published eQTL microarray datasets. The first dataset consisted of 3,754 samples from peripheral venous blood[84] and the second was from a meta-analysis of human brain cortex studies of 550 samples[85]. From both studies we obtained summary statistics from an association test of putative cis-eQTLs between all SNP-transcript pairs within a 1-Mb window of each other. To test for overlapping eQTLs, we used credible sets of markers (see **Defining Credible Sets**) at each of the 38 distinct migraine loci. Then for the most significant eQTLs ($P < 1x10^{-4}$) found to genes within a 1Mb window of each migraine credible set, we created an additional credible set of markers for each eQTL. We then tested (using Spearman's rank correlation) whether there was a significant correlation between the association test-statistics in each migraine credible set compared to the expression test-statistics in each overlapping eQTL credible set. Significant correlation between a migraine credible set and an eQTL credible set was taken as evidence of the migraine locus tagging a real eQTL. We determined the significance threshold to account for multiple testing by Bonferroni correction.

Enhancer Enrichment Analysis. Markers of gene regulation were defined using ChIP-seq datasets produced at the Broad Institute and the University of California at San Diego as part of ENCODE[65] and the NIH Roadmap Epigenome[64] projects. Based on the histone H3K27ac signal, which identifies active enhancers, we processed data from 56 cell lines and tissue

samples to identify celltype- or tissue-specific enhancers, which we define as the 10% of enhancers with the highest ratio of reads in that cell or tissue type divided by the total reads[86]. A description of all 56 tissues/cell types is provided in **Supplementary Table 21** and the raw enhancer data can be downloaded at http://www.epigenomeatlas.org.

We mapped the candidate causal variants at each migraine associated locus to these enhancer sequences, and compared the overlap observed with tissue specific enhancers relative to all enhancers using a background of 10,000 randomly selected sets of SNPs of equal size as the original locus. We restricted the background selection to common variants (MAF>1%) from 1000 Genomes that also passed our quality control filters in the meta-analysis (in other words, to only allow the selection of SNPs that had an *a priori* chance of being associated). The selection procedure then involved randomly selecting regions of the genome that were of equivalent length and containing an equivalent density of enhancers as found in the original locus. Once an appropriate region was found, a set of SNPs was randomly selected to match the number of SNPs in the credible set for that locus. If the selected SNPs happened to fall in an equal number of enhancer sites (of any tissue type) as that of the credible set of SNPs from the original locus, then the selected set of SNPs was accepted and added to the background set of SNPs for comparison. If the number of enhancers overlapping with the selected SNPs didn't match, the randomized selection procedure was repeated until an appropriate comparison set of SNPs was selected. This selection procedure was repeated for each locus 10,000 times to obtain an empirical null distribution. The significance of the observed enrichment was then estimated from the empirical distribution by calculating the proportion of replicates that were greater than the observed value (i.e. *Empirical P-value = [R + 1]/[N + 1]* where *R* is the number out of all *N* replicates that were higher than the observed enrichment). Finally, we used Bonferroni correction to adjust for multiple testing of 56 tissue/cell types ($P < 8.9 \times 10^{-4}$).

DEPICT reconstituted gene-set enrichment analysis. DEPICT[62] (Data-driven Expression Prioritized Integration for Complex Traits) is a computational tool, which, given a set of GWAS summary statistics, allows prioritization of genes in associated loci, enrichment analysis of reconstituted gene sets, and tissue enrichment analysis. DEPICT was run using as input 124 independent genome-wide significant SNPs (PLINK clumping parameters: --clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.6 --clump-kb 250. Note - rs12845494 and rs140002913 could not be mapped). LD distance ($r^2 > 0.5$) was used to define locus boundaries (note that this locus definition is different than used elsewhere in the text) yielding 37 autosomal loci comprising 78

genes. DEPICT was run using default settings, that is, 500 permutations for bias adjustment, 20 replications for false discovery rate estimation, normalized expression data from 77,840 Affymetrix microarrays for gene set reconstitution (see reference[87]), 14,461 reconstituted gene sets for gene set enrichment analysis, and testing 209 tissue/cell types assembled from 37,427 Affymetrix U133 Plus 2.0 Array samples for enrichment in tissue/cell type expression.

DEPICT identified 76 reconstituted gene sets that are significantly enriched (FDR < 5%) for genes found among the 38 migraine associated loci. Post-analysis, we omitted reconstituted gene sets in which genes in the original gene set were not nominally enriched (Wilcoxon rank-sum test) because, by design, genes in the original gene set are expected to be enriched in the reconstituted gene set; lack of enrichment therefore complicates interpretation of the reconstituted gene set because the label of the reconstituted gene set will be inaccurate. Therefore the following reconstituted gene sets were removed from the results (Wilcoxon rank-sum P-values in parentheses): MP:0002089 (0.01), MP:0002190 (0.16), ENSG00000151577 (0.21), ENSG00000168615 (0.94), ENSG00000143322 0.70), ENSG00000112531 (0.04), ENSG00000161021 (0.10), ENSG00000100320 (0.43). We also removed an association identified to another gene set (ENSG00000056345 PPI, $P = 1.7\times10^{-4}$, *FDR* = 0.04) because it is no longer part of the Ensembl database. These post-analysis filtering steps left us with 67 significantly enriched reconstituted gene sets. The Affinity Propagation tool[88] was finally used to cluster related reconstituted gene sets into 10 groups (script to produce the network diagram can be downloaded from https://github.com/perslab/DEPICT).

DEPICT tissue enrichment analysis. For the tissue enrichment analysis, DEPICT incorporates data from 37,427 human microarray samples captured on the Affymetrix HGU133a2.0 platform. These are used to test whether genes in the 38 migraine loci are highly expressed in 209 tissues/cell types with Medical Subject Heading (MeSH) annotations. The annotation procedure and method for normalizing expression profiles across annotations is outlined in the original publication[62]. The tissue/cell type enrichment analysis algorithm is conceptually identical to the gene set enrichment analysis algorithm whereby enrichment *P*-values are calculated empirically using 500 permutations for bias adjustment and 20 replications for false discovery rate estimation.

GTEx tissue enrichment analysis. Credible sets were generated for all 38 migraine loci (with $P < 5 \times 10^{-8}$) and corresponding gene sets for each locus were then generated by taking all

genes within 50 kilobases of a credible set SNP. Genes identified in this way were then analyzed for tissue enrichment using publicly available expression data from pilot phase of the Genotype-Tissue Expression project (GTEx)[61], version 3. In the pilot phase dataset, *postmortem* samples from 42 human tissues and three cell lines across 1,641 samples (**Supplementary Table 16**) have been used for bulk RNA sequencing according to a unified protocol. All samples were sequenced using Illumina 76 base-pair paired-end reads.

Collapsed reads per kilobase per million mapped reads (RPKM) values for each of the 52,577 included transcripts, filtered for unique HGNC IDs (n = 20,932), were organized by tissue and individual ($n_{tissues}$ = 45, $n_{samples}$ = 1,641). By this process we also excluded transcripts from any non-coding RNAs. All transcripts were ranked by mean RPKM across all samples, and 100,000 permutations of each credible set gene list were generated by selecting a random transcript for each entry in the credible set within +/-100 ranks of the transcript for that gene. For each sample, the RPKM values were converted into ranks for that transcript, and sums of ranks within each tissue were computed for each gene. *P*-values for each tissue were calculated by taking the total number of cases where the gene list of interest had a lower sum of ranks than the permuted sum of ranks, and dividing by the total number of permutations. To assess the significance of the enrichment after testing multiple tissues, we used a Bonferroni correction adjusted for the number of independent tissues, estimated via the matSpD tool[89] to arrive at 27 independent tests and a significance threshold of $P < 1.90 \times 10^{-3}$.

Specificity of individual gene expression in GTEx tissues. For the individual-gene expression analysis, we selected the closest gene to the index SNP at each migraine locus and then investigated expression activity of each of these genes in the collection of available tissues. As the number of samples for some tissues was small, we grouped individual tissues into four categories; brain, vascular, gastrointestinal, and other tissues (**Supplementary Table 16**). Then for each selected gene, we tested whether the average expression (mean RPKM) was significantly higher in a particular tissue group compared to the “other tissues” category. We assessed significance using a one-tailed t-test and used Bonferroni correction to control for multiple testing for all 114 tests (38 genes × 3 tissue groups). While some genes were observed to be significantly expressed in multiple tissue groups, we determined that a gene was tissue-specific if it was only expressed highly in one tissue group (i.e. brain, vascular, or gastrointestinal, **Supplementary Table 25)**.

eQTL credible set analysis in GTEx tissues. For all tissues and transcripts, we identified genome-wide significant ($P < 2 \times 10^{-13}$) *cis*-eQTLs within a 1Mb window of each transcript and created credible sets (see **Defining Credible Sets**) for each eQTL locus identified in each tissue. Then, for each eQTL credible set that contained markers that overlapped with a migraine credible set, we tested using Spearman's rank correlation if the test statistics between the two overlapping credible sets were significantly correlated. Significant correlation between a migraine credible set and an eQTL credible set was taken as evidence of the migraine locus tagging a real eQTL. Multiple testing was controlled for using Bonferroni correction.

Across the GTEx collection of tissues we found 35 significant *cis*-eQTLs within a 1Mb window of the 38 migraine loci, however, upon creating credible sets, seven of these still contained SNPs that overlapped with any of the migraine credible sets. Testing these seven eQTL credible sets as described above found that the correlation was significant ($P < 7.1 \times 10^{-3}$) for eQTLs to four tissues (Lung, Thyroid, Tibial Artery, and Aorta) at two migraine loci (*HPSE2* and *HEY2*) **Supplementary Table 19** and **Supplementary Figure 15**.

Heterogeneity analysis of migraine subtypes. To discover if heterogeneity between the migraine subtypes might have affected our ability to identify new loci, we performed an additional meta-analysis using a subtype-differentiated approach that allows for different allelic effects between the two groups[57]. Since a large proportion of the controls were shared in the original migraine with aura and migraine without aura samples (see **Table 1**), for this analysis we created two additional subsets of the migraine subtype data that contained no overlapping controls between the two new subsets (**Supplementary Table 12**). The new migraine with aura subset consisted of 4,837 cases and 49,174 controls and the new migraine without aura subset consisted if 4,833 cases and 106,834 controls. Then using the association test statistics from each of the individual GWA studies listed, we performed the subtype-differentiated meta-analysis as implemented in GWAMA (see URLs).

To assess the amount of heterogeneity observed, we chose the 44 LD independent SNPs that were associated with migraine and examined the results of the subtype-differentiated meta-analysis. We observed that only seven out of the 44 SNPs showed evidence for heterogeneity in the subtype-differentiated test (Heterogeneity *P*-value < 0.05, **Supplementary Table 13**). This suggests that most of the identified loci are truly affecting risk for both migraine with aura

and migraine without aura even though we may not yet have power to detect their association in the subset meta-analyses.

Methods references

76. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5,** 1564–1573 (2010).

77. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–1212 (2014).

78. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature Methods* **9,** 179–181 (2011).

79. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* **44,** 955–959 (2012).

80. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).

81. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).

82. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31,** 782–784 (2015).

83. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).

84. Wright, F. a *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46,** 430–7 (2014).

85. Richards, A. L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol. Psychiatry* **17,** 193–201 (2012).

86. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518,** 337–343 (2014).

87. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47,** 115–125 (2015).

88. Frey, B. J. & Dueck, D. Clustering by Passing Messages Between Data Points. *Science (80-.).* **315,** 972–976 (2007).

89. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74,** 765–769 (2004).