nature genetics

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

Padhraig Gormley[1–4,81], Verneri Anttila[2,3,5,81], Bendik S Winsvold[6–8], Priit Palta[9], Tonu Esko[2,10,11], Tune H Pers[2,11–13], Kai-How Farh[2,5,14], Ester Cuenca-Leon[1–3,15], Mikko Muona[9,16–18], 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 H H 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 W J H Penninx[41], Markus Färkkilä[26], Ville Artto[26], Mari Kaunisto[9], Salli Vepsäläinen[26], Rainer Malik[28], Andrew C Heath[42], Pamela A F Madden[42], Nicholas G Martin[30], Grant W Montgomery[30], Mitja I Kurki[1–3,9,43], 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[44], Jie Huang[4], Evie Stergiakouli[23], Phil H Lee[1–3], Cynthia Sandor[45], Caleb Webber[45], Zameel Cader[46,47], Bertram Muller-Myhsok[48,76,93], Stefan Schreiber[49], Thomas Meitinger[50,51], Johan G Eriksson[52,53], Veikko Salomaa[53], Kauko Heikkilä[54], Elizabeth Loehrer[34,55], Andre G Uitterlinden[56], Albert Hofman[34], Cornelia M van Duijn[34], Lynn Cherkas[33], Linda M Pedersen[6], Audun Stubhaug[57,58], Christopher S Nielsen[57,59], Minna Männikkö[32], Evelin Mihailov[10], Lili Milani[10], Hartmut Göbel[60], Ann-Louise Esserlind[61], Anne Francke Christensen[61], Thomas Folkmann Hansen[62], Thomas Werge[63–65], International Headache Genetics Consortium[66], Jaakko Kaprio[9,37,67], Arpo J Aromaa[53], Olli Raitakari[68,69], M Arfan Ikram[34,35,70], Tim Spector[33], Marjo-Riitta Järvelin[32,71–73], Andres Metspalu[10], Christian Kubisch[74], David P Strachan[75], Michel D Ferrari[25], Andrea C Belin[29], Martin Dichgans[28,76], Maija Wessman[9,16], Arn M J M van den Maagdenberg[25,77], John-Anker Zwart[6–8], Dorret I Boomsma[24], George Davey Smith[23], Kari Stefansson[22,78], Nicholas Eriksson[19], Mark J Daly[2,3,5], Benjamin M Neale[2,3,5,82], Jes Olesen[61,82], Daniel I Chasman[21,38,82], Dale R Nyholt[79,82] & Aarno Palotie[1–5,9,80,82]

Migraine is a debilitating neurological disorder affecting around one in seven people worldwide, but its molecular mechanisms remain poorly understood. There is some debate about 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 carried out a genetic study of migraine on 59,674 affected subjects 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 mapped to 38 distinct genomic loci, including 28 loci not previously reported and a locus that to our knowledge is the first to be 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 the third most common disease worldwide, with a lifetime prevalence of 15–20%, affecting up to 1 billion people across the globe[1,2]. It ranks as the seventh most disabling disease worldwide (and the most disabling neurological disease) in terms of years of life lost to disability[1], and it is the third most costly neurological disorder, after dementia and stroke[3]. There is debate about whether migraine is a disease of vascular dysfunction or of neuronal dysfunction with vascular changes representing downstream effects that are not themselves causative of migraine[4,5]. However, genetic evidence favoring one theory over the other is lacking. At the phenotype level, migraine is defined by diagnostic criteria from the International Headache Society[6]. There are two prevalent subforms: migraine without aura, which is characterized by recurrent attacks of moderate or severe headache associated with nausea or hypersensitivity to light and

A full list of authors and affiliations appears at the end of the paper.

Received 27 October 2015; accepted 26 May 2016; published online 20 June 2016; corrected online 18 July 2016 (details online); doi:10.1038/ng.3598

sound, and migraine with aura, which is characterized by transient visual, sensory, 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 = 36–47%) for migraine[7], pointing to a genetic component of the disease. Despite this, genetic association studies have uncovered relatively little about the molecular mechanisms that contribute to migraine's pathophysiology. Understanding has been limited partly because so far only 13 genome-wide significant risk loci have been identified for the prevalent forms of migraine[8–11]. For 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 the more prevalent forms of migraine[15].

We conducted a meta-analysis of 22 GWA studies, including data for a total of 59,674 affected subjects and 316,078 controls collected from six tertiary headache clinics and 27 population-based cohorts through our worldwide collaboration with the International Headache Genetics Consortium (IHGC). This combined data set contained more than 35,000 new migraine cases not included in previously published GWA studies. Here we present the findings of this 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 carried out on all samples from subjects with migraine available through the IHGC, regardless of ascertainment. These case samples came from both individuals diagnosed by a doctor and individuals with self-reported migraine as stated on questionnaires. Study design and sample ascertainment for each individual study are outlined in the **Supplementary Note** (and summarized in **Supplementary Table 1**). The final combined sample consisted of 59,674 case samples and 316,078 controls in 22 non-overlapping case–control data sets (**Table 1**). All subjects were of European ancestry (EUR). Before including the largest study from 23andMe, we confirmed that it did not contribute any additional heterogeneity compared with the other population and clinic-based studies (**Supplementary Table 2**).

The 22 individual GWA studies included 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 carried out within each study using logistic regression on the imputed marker dosages, with adjustments made for sex and other covariates where necessary (Online Methods and **Supplementary Table 4**). The association results were combined in 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 the variants in the primary analysis, we identified 44 genome-wide significant SNP associations ($P < 5 \times 10^{-8}$; **Supplementary Fig. 1**) that were independent ($r^2 < 0.1$) with regard to linkage disequilibrium (LD). We validated these 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, we defined an associated locus as the genomic region bounded by all markers in LD ($r^2 > 0.6$ in 1000 Genomes Project, phase 1, EUR individuals) with each of the 44 index SNPs, and in addition, all such regions in close proximity (<250 kb) were merged. On the basis of these defined regions, we implicated 38 genomic loci for the prevalent forms of migraine, 28 of which had not been reported previously (**Fig. 1**).

These 38 loci replicated 10 of the 13 previously reported genome-wide associations with migraine, and 6 loci contained a secondary genome-wide significant SNP 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–STAT6–SDR9C7*, *PRDM16*, *FHL5–UFL1*, *TRPM8–HJURP*, and near *TSPAN2–NGF*), whereas the sixth was found within one of the 28 new loci (*PLCE1*). Therefore, out of the 44 independent SNPs reported here, 34 represent new associations with migraine. Three previously reported loci that were associated with subtypes of migraine (rs1835740 near *MTDH* for migraine with aura, rs10915437 near *AJAP1* for migraine clinical samples, and rs10504861 near *MMP16* for migraine without aura)[8,11] showed 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 with specific phenotypic features of migraine[17], and therefore a more phenotypically homogeneous sample may be required for an accurate assessment of association. Four out of 44 SNPs (*MRVI1*, at *TRPM8–HJURP*, near *ZCCHC14*, and near *CCM2L–HCK*) showed moderate heterogeneity across the individual GWA studies (Cochran's Q test, $P < 0.05$; **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 overlapped with transcripts from protein-coding genes, and 17 (45%) of these regions contained just a single gene (see **Supplementary Fig. 2** for regional association plots and **Supplementary Table 8** for additional locus information). Among the 38 loci, only two contained ion channel genes (*KCNK5* and *TRPM8*)[19,20]. Thus, despite previous hypotheses that migraine is a potential channelopathy[5,21], the loci identified to date do not support the idea of common variants in ion channel genes being strong susceptibility components in prevalent forms of migraine. However, three other loci do contain genes involved more generally in ion homeostasis[22–24] (*SLC24A3*, *ITPK1*, and *GJA1*; **Supplementary Table 9**).

Several of the identified genes have previously been associated with vascular disease (*PHACTR1*, *TGFBR2*, *LRP1*, *PRDM16*, *RNF213*, *JAG1*, *HEY2*, *GJA1*, and *ARMS2*)[25–34] or are involved in smooth muscle contractility and regulation of vascular tone (*MRVI1*, *GJA1*, *SLC24A3*, and *NRP1*)[35–38]. Three of the 44 migraine index SNPs had previously reported associations in the National Human Genome Research Institute GWA study catalog at exactly the same SNP (rs9349379 at *PHACTR1* with coronary heart disease[39–41], coronary artery calcification[42], and cervical artery dissection[26]; rs11624776 near *ITPK1* with thyroid hormone levels[43]; and rs11172113 at *LRP1–STAT6–SDR9C7* with pulmonary function[44]; **Supplementary Table 10**). Six of the loci harbor genes that are involved in nitric oxide (NO) signaling and oxidative stress (*REST*, *GJA1*, *YAP1*, *PRDM16*, *LRP1*, and *MRVI1*)[45–50].

For each locus, we chose the gene nearest to the index SNP to assess gene expression activity in tissues from the Genotype-Tissue Expression (GTEx) Consortium project (**Supplementary Fig. 3**). Although we found that most of the putative migraine loci genes were expressed in many different tissue types, we were able to detect tissue specificity in certain instances in which some genes showed

Table 1 Individual IHGC GWA studies with numbers of case and control samples used in the primary analysis (all migraine) and in the subtype analyses (migraine with aura and migraine without aura)

		All migraine		Migraine with aura		Migraine without aura	
GWA study ID	Full name of GWA study	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**

Note that chromosome X genotype data were unavailable from three of the individual GWA studies (EGCUT, Rotterdam III, and Twins UK) and were partially unavailable from some of the control samples (specifically, the GSK controls) used for the German MO study, meaning that the samples analyzed for chromosome X represented 57,756 cases and 299,109 controls. Complete data were available on the autosomes for all samples.

significantly higher expression in a particular tissue group relative to the others. For instance, four genes were expressed more actively in brain (*GPR149*, *CFDP1*, *DOCK4*, and *MPPED2*) than in other tissues, and eight genes were specifically active in vascular tissues (*PRDM16*, *MEF2D*, *FHL5*, *C7orf10*, *YAP1*, *LRP1*, *ZCCHC14*, and *JAG1*). Many 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 harbored true associations with migraine, rather than reflecting systematic differences between case and control samples (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 deviated from the null (genomic inflation factor $\lambda_{GC} = 1.24$; **Supplementary Fig. 4**), which is in line with other large GWA study meta-analyses[51–54]. Because much of the inflation for a polygenic trait arises from LD between the causal SNPs and many other neighboring SNPs in the local region, we LD-pruned the data to create a set of LD-independent markers (in PLINK[55] with a 250-kb sliding window and $r^2 > 0.2$). The resulting genomic inflation was comparatively reduced ($\lambda_{GC} = 1.15$; **Supplementary**

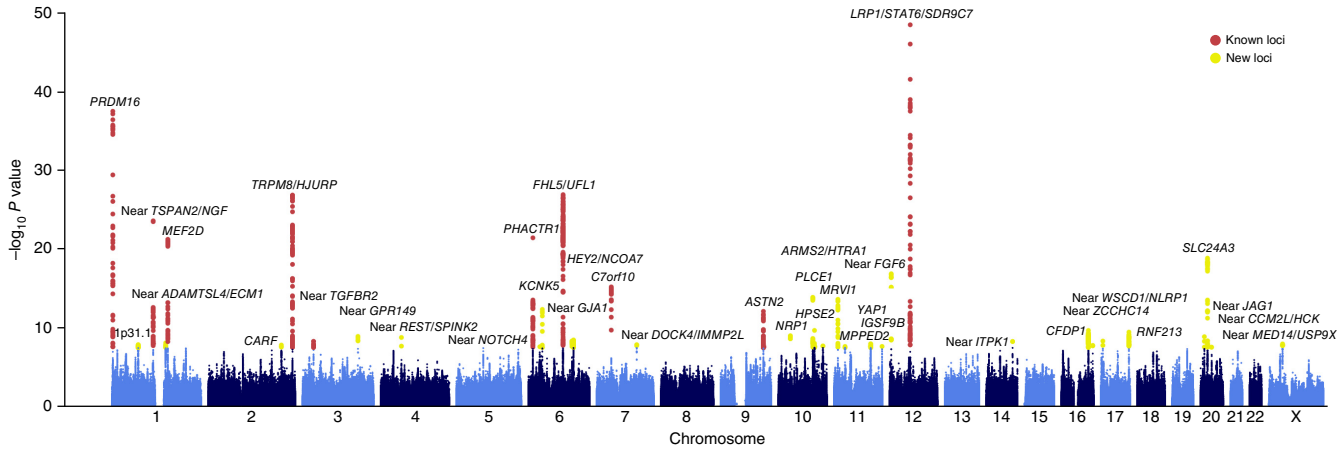

Figure 1 Manhattan plot showing results of the primary meta-analysis of all migraine samples (59,674 case and 316,078 control). The horizontal axis shows the chromosomal position, and the vertical axis shows the significance of tested markers combined in a fixed-effects meta-analysis. Markers that reached genome-wide significance ($P < 5 \times 10^{-8}$, chi-square test, 1 d.f.) at previously known and/or newly identified loci are highlighted.

Fig. 5) and probably reflects the inflation remaining owing to the polygenic signal at many independent loci, including those not yet significantly associated.

To confirm that the observed inflation came primarily from true polygenic signal, we analyzed the data from all imputed markers using LD-score regression[56]. 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 Fig. 6**) and suggest 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 demonstrated previously by

Table 2 Summary of the 38 genomic loci associated with the prevalent types of migraine

						All migraine		Secondary signal		Migraine without aura		Previous publication
Locus rank	Locus	Chr	Index SNP	Minor allele	MAF	OR (95% CI)	*P*	Index SNP	*P*	Index SNP	*P*	PMID
1	*LRP1–STAT6–SDR9C7*	12	rs11172113	C	0.42	0.90 (0.89–0.91)	5.6×10^{-49}	rs11172055	1.3×10^{-9}	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–UFL1*	6	rs67338227	T	0.23	1.09 (1.08–1.11)	2.0×10^{-27}	rs4839827	5.7×10^{-10}	rs7775721	1.1×10^{-12}	23793025
4	Near *TSPAN2–NGF*	1	rs2078371	C	0.12	1.11 (1.09–1.13)	4.1×10^{-24}	rs7544256	8.7×10^{-9}	rs2078371	7.4×10^{-9}	23793025
5	*TRPM8–HJURP*	2	rs10166942	C	0.20	0.94 (0.89–0.99)	1.0×10^{-23}	rs566529	2.5×10^{-9}	rs6724624	1.1×10^{-9}	21666692
6	*PHACTR1*	6	rs9349379	G	0.41	0.93 (0.92–0.95)	5.8×10^{-22}	-	-	rs9349379	2.1×10^{-9}	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	Near *FGF6*	12	rs1024905	G	0.47	1.06 (1.04–1.08)	2.1×10^{-17}	-	-	rs1024905	2.5×10^{-9}	-
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^{-9}	-	-	-
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^{-9}	-	-	-	-	-
19	Near *GPR149*	3	rs13078967	C	0.03	0.87 (0.83–0.91)	1.8×10^{-9}	-	-	-	-	-
20	Near *JAG1*	20	rs111404218	G	0.34	1.05 (1.03–1.07)	2.0×10^{-9}	-	-	-	-	-
21	Near *REST–SPINK2*	4	rs7684253	C	0.45	0.96 (0.94–0.97)	2.5×10^{-9}	-	-	-	-	-
22	Near *ZCCHC14*	16	rs4081947	G	0.34	1.03 (1.00–1.06)	2.5×10^{-9}	-	-	-	-	-
23	*HEY2–NCOA7*	6	rs1268083	C	0.48	0.96 (0.95–0.97)	5.3×10^{-9}	-	-	-	-	-
24	Near *WSCD1–NLRP1*	17	rs75213074	T	0.03	0.89 (0.86–0.93)	7.1×10^{-9}	-	-	-	-	-
25	Near *GJA1*	6	rs28455731	T	0.16	1.06 (1.04–1.08)	7.3×10^{-9}	-	-	-	-	-
26	Near *TGFBR2*	3	rs6791480	T	0.31	1.04 (1.03–1.06)	7.8×10^{-9}	-	-	-	-	22683712
27	Near *ITPK1*	14	rs11624776	C	0.31	0.96 (0.94–0.97)	7.9×10^{-9}	-	-	-	-	-
28	Near *ADAMTSL4–ECM1*	1	rs6693567	C	0.27	1.05 (1.03–1.06)	1.2×10^{-8}	-	-	-	-	-
29	Near *CCM2L–HCK*	20	rs144017103	T	0.02	0.85 (0.76–0.96)	1.2×10^{-8}	-	-	-	-	-
30	*YAP1*	11	rs10895275	A	0.33	1.04 (1.03–1.06)	1.6×10^{-8}	-	-	-	-	-
31	Near *MED14–USP9X*	X	rs12845494	G	0.27	0.96 (0.95–0.97)	1.7×10^{-8}	-	-	-	-	-
32	Near *DOCK4–IMMP2L*	7	rs10155855	T	0.05	1.08 (1.05–1.12)	2.1×10^{-8}	-	-	-	-	-
33	1p31.1[a]	1	rs1572668	G	0.48	1.04 (1.02–1.05)	2.1×10^{-8}	-	-	-	-	-
34	*CARF*	2	rs138556413	T	0.03	0.88 (0.84–0.92)	2.3×10^{-8}	-	-	-	-	-
35	*ARMS2–HTRA1*	10	rs2223089	C	0.08	0.93 (0.91–0.95)	3.0×10^{-8}	-	-	-	-	-
36	*IGSF9B*	11	rs561561	T	0.12	0.94 (0.92–0.96)	3.4×10^{-8}	-	-	-	-	-
37	*MPPED2*	11	rs11031122	C	0.24	1.04 (1.03–1.06)	3.5×10^{-8}	-	-	-	-	-
38	Near *NOTCH4*	6	rs140002913	A	0.06	0.91 (0.88–0.94)	3.8×10^{-8}	-	-	-	-	-

[a]The nearest coding gene (*LRRIQ3*) to this locus is 592 kb away.

Ten loci were previously reported (PubMed IDs (PMID) are listed for these loci), and 28 were newly found in this study. Each locus is labeled with protein-coding genes that overlap the region. Intergenic loci are also labeled as "near" to highlight the additional uncertainty in identifying relevant genes. 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 are given. For the seven loci reaching genome-wide significance in the 'migraine without aura' subtype analysis, the corresponding index SNP and *P* value are also given. Evidence for significant heterogeneity was found at four loci (*TRPM8–HJURP*, *MRVI1*, near *ZCCHC14*, and near *CCM2L–HCK*), so for those we present the results of a random-effects model. Chr, chromosome; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

Figure 2 Expression enrichment of genes from the migraine loci in GTEx tissue samples. Expression data from 1,641 samples were obtained using RNA-seq for 42 tissues and three cell lines. 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 adjustment for multiple testing by Bonferroni correction (Online Methods). LCL, Epstein–Barr transformed lymphoblastoid cell lines.

GWA studies using samples of similar size with both simulated and real data[57].

Migraine subtype analyses

To elucidate pathophysiological mechanisms underpinning the migraine aura, we carried out a secondary analysis of two created subsets that included only samples with the subtypes 'migraine with aura' and 'migraine without aura'. These subsets included samples only from studies in which sufficient information was available to assign a diagnosis of one or the other subtype according to classification criteria standardized by the International Headache Society[6]. For the population-based studies, this involved questionnaires, whereas for the clinic-based studies the diagnosis was assigned on the basis of a structured interview by telephone or in person. A stricter diagnosis was required for the migraine subtypes, as it is often challenging to distinguish migraine aura from other neurological features that can present as symptoms from unrelated conditions. As a result, the migraine subtype analyses involved considerably smaller sample sizes than the main analysis (6,332 case samples and 144,883 controls for migraine with aura; 8,348 case samples and 139,622 controls for migraine without aura; **Table 1**).

As with the primary analysis, the test statistics for migraine with aura and for migraine without aura were consistent with underlying polygenic architecture rather than other potential sources of inflation (**Supplementary Figs. 7** and **8**). In our analysis for migraine without aura, we found seven significantly ($P < 5 \times 10^{-8}$) associated genomic loci (near *TSPAN2*, *TRPM8*, *PHACTR1*, *FHL5*, *ASTN2*, near *FGF6*, and *LRP1*; **Supplementary Table 11** and **Supplementary Fig. 9**). All seven of these loci had already been identified in the primary analysis, possibly reflecting the fact that migraine without aura is the most common form of migraine (around two in three cases) and probably drove these association signals in the primary analysis. Notably, no loci were associated with migraine with aura in the other subset analysis (**Supplementary Fig. 10**).

To investigate whether excess heterogeneity could be contributing to the lack of associations for migraine with aura, we carried out a heterogeneity analysis of the two subgroups (Online Methods and **Supplementary Table 12**). We selected the 44 LD-independent SNPs 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 allowed for heterogeneity between the two migraine groups[58]. We found little heterogeneity, with only 7 of the 44 loci (at *MEF2D*, *PHACTR1*, near *REST–SPINK2*, *ASTN2*, *PLCE1*, *MPPED2*, and near *MED14–USP9X*) exhibiting 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[59]. This method incorporated evidence from association-test statistics and the LD structure between SNPs in a locus (Online Methods). A list of the credible set of SNPs obtained for each locus is provided in **Supplementary Table 14**. We found three loci (in *RNF213*, *PLCE1*, and *MRVI1*) for which the association signal could be credibly attributed to exonic mis-sense 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 disruption of protein structure[60,61].

Overlap with eQTLs in specific tissues

To identify migraine loci that might influence gene expression, we used previously published data sets cataloging expression quantitative

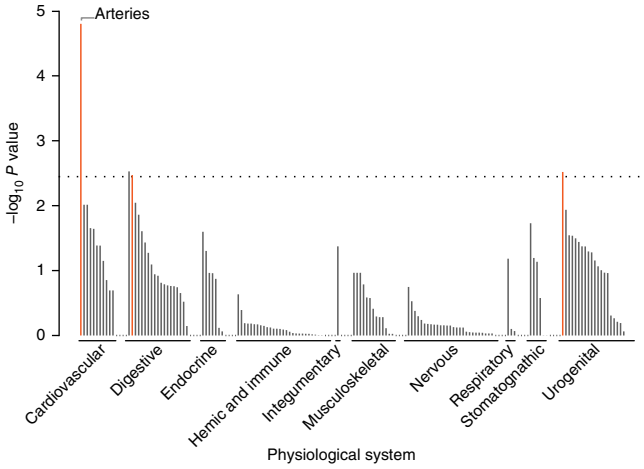

Figure 3 Expression enrichment of genes from the migraine loci in 209 tissue or cell type annotations by DEPICT. Expression data were obtained from 37,427 human microarray samples, and then genes in the migraine loci were assessed for high expression in each of the annotation categories. We determined enrichment *P* values by comparing the expression patterns from the migraine loci to those from 500 randomly generated loci, and we estimated the false discovery rate (horizontal dashed line denotes FDR < 0.05) to control for multiple testing (Online Methods). Red bars indicate categories that were significant after FDR correction. A full list of these enrichment results is provided in **Supplementary Table 20**.

trait loci (eQTLs) in either of two microarray-based studies of samples from peripheral venous blood (n = 3,754) or from human brain cortex (n = 550). Additionally, we used a third study based on RNA-seq data from a collection of 42 tissues and three cell lines (n = 1,641) from the GTEx Consortium project[62]. Although these data offered the advantage of a diverse tissue catalog, the number of samples per tissue was relatively small (**Supplementary Table 16**) compared with the two microarray data sets, which may have resulted in reduced power to detect significant eQTLs in some tissues. Using these data sets, 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 whether 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** and **18** and **Supplementary Fig. 11**). In the GTEx data, 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–NCOA7* locus (**Supplementary Table 19** and **Supplementary Fig. 12**).

In summary, from three data sets we found evidence implicating eQTL signals at only two loci (*HPSE2* and *HEY2*). This low number (2 out of 38) is consistent with previous studies noting that available eQTL catalogs currently lack sufficient tissue specificity and developmental diversity to provide enough power for meaningful biological insight[53]. No plausibly causal eQTLs were observed in expression data from brain tissue samples.

Gene expression enrichment in specific tissues

To understand whether the 38 migraine loci as a group are enriched for expression in certain tissue types, we again used the GTEx pilot data[62] (Online Methods). We found four tissues that were significantly enriched (after Bonferroni correction) for expression of the migraine-associated genes (**Fig. 2**). The two most strongly enriched tissues were part of the cardiovascular system (aorta and tibial artery). The two other significantly enriched tissues were from the digestive system (esophagus muscularis and esophageal mucosa). We replicated these enrichment results using the DEPICT[63] tool and an independent microarray-based gene expression data set (Online Methods). DEPICT highlighted four tissues (**Fig. 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 implicated arterial and gastrointestinal tissues. To discover whether 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[64]. Overall, migraine loci genes were not significantly enriched in a particular class of smooth muscle (**Supplementary Figs. 13–15**). This suggests that the enrichment of migraine risk variants in genes expressed in tissues with a smooth muscle component is not specific to blood vessels, the stomach, or the gastrointestinal tract; rather, it seems to be generalizable across vascular and visceral smooth muscle types.

Combined, these enrichment results suggest that some of the genes affected by migraine-associated variants have high expression in vascular tissues, and their dysfunction could have a role in migraine. Furthermore, the results suggest that other tissue types (for example, smooth muscle) could also have a role, which 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. Using candidate causal variants from the migraine loci, we examined their enrichment in cell-type-specific enhancers from 56 primary human tissues and cell types from the Roadmap Epigenomics[65] and ENCODE projects[66] (Online Methods and **Supplementary Table 21**). These variants showed the greatest enrichment in tissues from the mid-frontal lobe and duodenum smooth muscle, but their enrichment was not significant after adjustment for multiple testing (**Fig. 4**).

Gene set enrichment analyses

To implicate underlying biological pathways involved in migraine, we applied a Gene Ontology over-representation analysis of the 38 migraine loci (Online Methods). We found nine vascular-related biological function categories that were significantly enriched (adjusted $P < 0.05$) after correction for multiple testing (**Supplementary Table 22**). Notably, data for the identified loci provided little statistical support for some molecular processes that have been previously linked to migraine, including ion homeostasis, glutamate signaling, serotonin signaling, NO signaling, and oxidative stress. However, a possible explanation for the lack of enrichment for these functions is that current annotations for many genes and pathways are far from comprehensive, or that larger numbers of migraine loci need to be identified before sensitivity can be sufficient to detect enrichment for these mechanisms.

For a more comprehensive pathway analysis, we used DEPICT, which incorporates gene coexpression information from microarray data to implicate additional, functionally less well-characterized genes in known biological pathways, protein–protein complexes, and mouse phenotypes[63] (by forming so-called reconstituted gene sets). From DEPICT, we identified 67 reconstituted gene sets that were

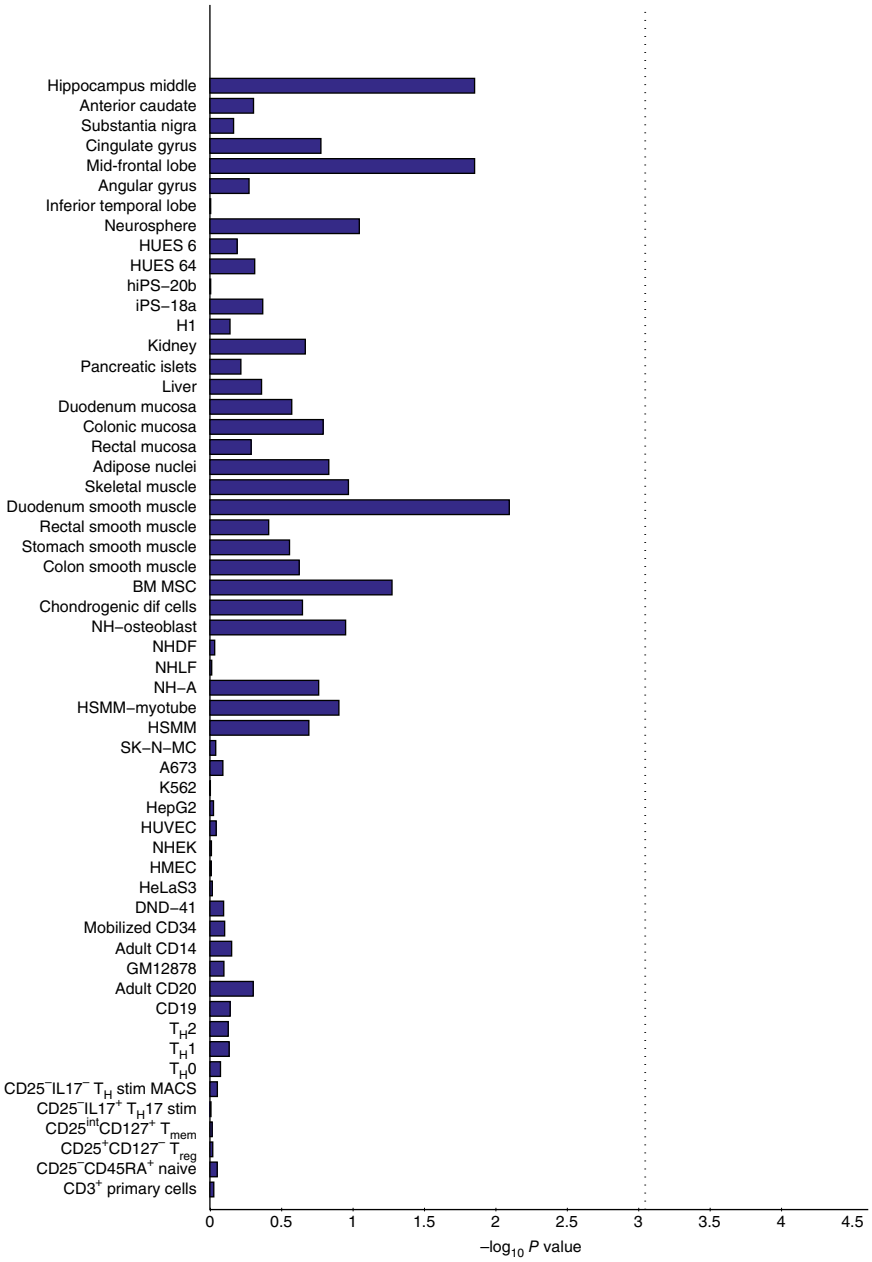

Figure 4 Enrichment of the migraine loci in sets of tissue- and cell-type-specific enhancers. Credible sets from the migraine loci were mapped to sets of enhancers under active expression in 56 tissues and cell lines (identified by H3K27ac histone marks from the Roadmap Epigenomics[65] and ENCODE[66] projects). We assessed enrichment *P* values empirically by randomly generating a background set of matched loci for comparison (10,000 replicates); the vertical dotted line indicates the significance threshold after adjustment for 56 separate tests by Bonferroni correction (Online Methods). BM MSC, bone marrow mesenchymal stem cells; dif, differentiated; NHDF, normal human dermal fibroblasts; NHLF, normal human lung fibroblasts; NH-A, normal human astrocytes; MACS, magnetic-activated cell sorting; HSMM, human skeletal muscle myoblast; HUVEC, human umbilical vein endothelial cells; NHEK, normal human epidermal keratinocyte; HMEC, human mammary epithelial cells; T_H, helper T cell; T_{mem}, memory T cell; T_{reg}, regulatory T cell; stim, stimulated.

significantly enriched (false discovery rate (FDR) < 5%) for genes found among the migraine-associated loci (**Supplementary Table 23**). Because the reconstituted gene sets had genes in common, we clustered them into ten distinct groups (**Fig. 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* protein-protein interaction, adherens junction, and integrin complex). Several of the other gene set clusters were also related to vascular biology (**Fig. 5** and **Supplementary Table 23**). We still did not observe any support for molecular processes with hypothesized links to migraine (**Supplementary Table 24**); however, this might again have been due to the reasons outlined above.

DISCUSSION

In what is to our knowledge the largest genetic study of migraine so far, 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 in both arterial and smooth muscle function. Two separate analyses, the DEPICT and GTEx gene expression enrichment analyses, pointed to the involvement of vascular and smooth muscle tissues in common variant susceptibility to migraine. The vascular finding is consistent with known comorbidities and previously reported shared polygenic risk among migraine, stroke, and cardiovascular diseases[67,68]. Furthermore, a recent GWA study of cervical artery dissection identified a genome-wide significant association at the same index SNP (rs9349379 in the *PHACTR1* locus) as is associated with migraine, suggesting the possibility of partially shared genetic components between migraine and cervical artery dissection[26]. These results suggest that vascular dysfunction and possibly also smooth muscle dysfunction are likely to have 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 data sets. 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[5,69–72]. We found some support for this in our examination of the expression of individual genes at the 38 loci (**Supplementary Fig. 3** and **Supplementary Table 25**), which showed that several genes were specifically active in brain tissues. We did not observe statistically significant enrichment in brain across all loci, but it may be that more associated loci are needed in order for this to be detected. Alternatively, the lack of significant enrichment could be due to difficulties in collecting appropriate brain tissue samples with enough specificity, or other technical

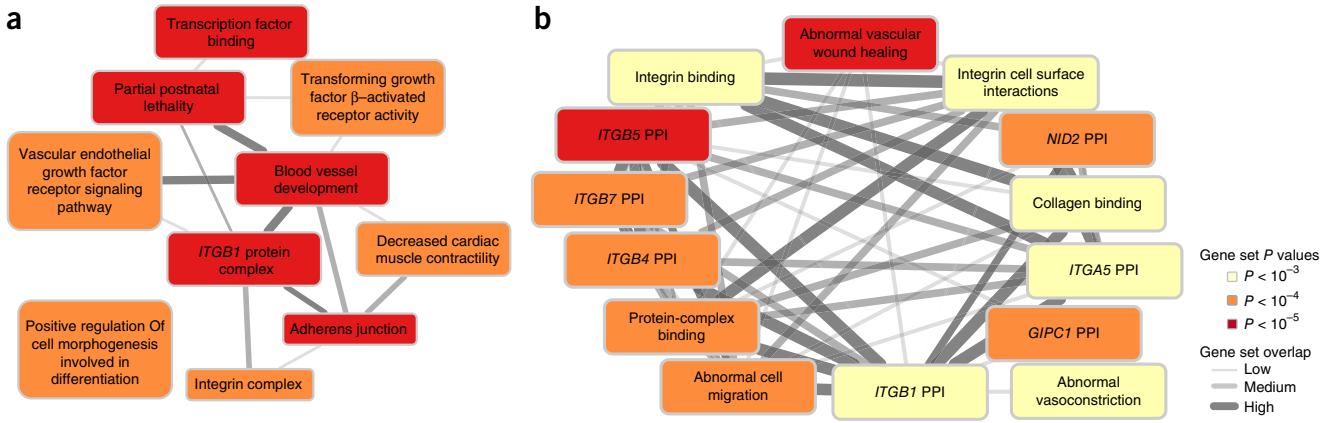

Figure 5 DEPICT network of the reconstituted gene sets that were significantly enriched (FDR < 0.05, determined empirically by permutation) for genes at the migraine loci (Online Methods). Enriched gene sets are represented as nodes, with the extent of pairwise overlap denoted by the width of the connecting lines, and empirical enrichment *P* values are indicated by color according to the key. (**a**) The 67 significantly enriched gene sets clustered by similarity into ten group nodes, with each group node named for the most representative gene set in the group. (**b**) One example of gene sets that were clustered within the now expanded *ITGB1* protein-protein interaction (PPI) group. A full list of the 67 significantly enriched reconstituted gene sets can be found in **Supplementary Table 23**.

challenges. Additionally, there is less clarity regarding the biological mechanisms for a neurological disease like migraine compared with some other common diseases, such as autoimmune or cardiometabolic diseases for which intermediate risk factors and underlying mechanisms are better understood.

Interestingly, some of the analyses highlighted gastrointestinal tissues. Although migraine attacks may include gastrointestinal symptoms (e.g., nausea, vomiting, diarrhea)[6], 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, but specific smooth muscle subtypes were not available to test this hypothesis in our primary enrichment analyses. Instead we examined a range of 60 smooth muscle subtypes and found that the migraine loci were expressed in many types of smooth muscle, including vascular (**Supplementary Figs. 14** and **15**). These results, although not conclusive, suggest that the enrichment of the migraine loci in smooth muscle is not specific to the stomach and gastrointestinal tract.

Our results implicate cellular pathways and provide an opportunity to determine whether the genomic data support previously presented hypotheses of mechanisms linked to migraine. One prevailing hypothesis, stimulated by findings in FHM, has been that migraine is a channelopathy[5,21]. Among the 38 migraine loci in the current study, only two harbor known ion channels (*KCNK5* and *TRPM8*)[19,20], and three others (*SLC24A3*, near *ITPK1*, and near *GJA1*) can be linked to ion homeostasis[22–24]. 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 genetic susceptibility. Moreover, we cannot exclude that ion channels might still be important contributors in migraine with aura, the form most closely resembling FHM, as identifying loci for this subgroup is more challenging. Another hypothesis relates to oxidative stress and NO signaling[73–75]. Six genes with known links to oxidative stress and NO were identified in these 38 loci (*REST*, *GJA1*, *YAP1*, *PRDM16*, *LRP1*, and *MRVI1*)[45–50]. This is in line with previous findings[11]; however, the DEPICT pathway analysis did not show an association between NO-related reconstituted gene sets and migraine (FDR > 0.54; **Supplementary Table 24**).

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** and **13**) demonstrated that most of the identified loci were implicated in both migraine subtypes. This suggests that the absence of significant loci in the analysis for migraine with aura was mainly due to a lack of power owing to the smaller sample size. Additionally, as shown by the LD-score analysis (**Supplementary Figs. 6–8**), the amount of heritability captured by the data set for migraine with aura was considerably lower than that for migraine without aura, such that in order for comparable power to be achieved, a sample size two to three times larger would be required. This might reflect a higher degree of heterogeneity in the clinical capture, more complex underlying biology, or even a greater contribution to risk from low-frequency and rare variation 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, https://github.com/perslab/DEPICT; credible set fine-mapping script, https://github.com/hailianghuang/FM-summary; GTEx, http://www.gtexportal.org; GWAMA, http://www.well.ox.ac.uk/gwama/; IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; IHGC, 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; PANTHER Gene Ontology enrichment, http://geneontology.org/page/go-enrichment-analysis; PLINK, https://www.cog-genomics.org/plink2; ProbABEL, http://www.genabel.org/packages/ProbABEL; R, https://www.r-project.org/; Roadmap Epigenomics data, http://www.epigenomeatlas.org; SHAPEIT, http://www.shapeit.fr; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

P.G., V. Anttila, G.W.M., M.I.K., M. Kals, R. Mägi, 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. Anttila, B.S.W., P.P., T.E., T.H.P., K.-H.F., M. Muona, N.A.F., A.I., G.M., L.L., S.G.G., S. Steinberg, L.Q., H.H.H.A., D.A.H., J.-J.H., R. Malik, 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. Anttila, B.S.W., P.P., T.E., T.H.P., K.-H.F., E.C.-L., N.A.F., A.I., G.M., L.L., M. Kallela, T.M.F., S.G.G., S. Steinberg, M. Koiranen, L.Q., H.H.H.A., T.L., J.W., D.A.H., S.M.R., M.F., V. Artto, M. Kaunisto, S.V., R. Malik, M.I.K., M. Kals, R. Mägi, 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 and/or interpretation. P.G., V. Anttila, B.S.W., T.H.P., K.-H.F., E.C.-L., T.K., G.M.T., M. Kallela, C.R., A.H.S., G.B., M. Koiranen, T.L., M.S., M.G.H., M.F., V. Artto, M. Kaunisto, S.V., R. Malik, 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. Schreiber, 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., A.-P.S., J.E.B., P.M.R., and A.P. contributed materials and/or 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. Männikkö, 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. Anttila, 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.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

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* **386**, 743–800 (2015).
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. *Nat. Rev. Neurosci.* **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. Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia* **33**, 629–808 (2013).
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**, 912–917 (2013).
12. Ophoff, R.A. *et al.* Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca^{2+} channel gene *CACNL1A4*. *Cell* **87**, 543–552 (1996).
13. De Fusco, M. *et al.* Haploinsufficiency of *ATP1A2* encoding the Na^+/K^+ pump $\alpha 2$ 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. 1000 Genomes Project Consortium *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**, e1004366 (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. USA* **102**, 16102–16106 (2005).
20. Ramachandran, R. *et al.* TRPM8 activation attenuates inflammatory responses in mouse models of colitis. *Proc. Natl. Acad. Sci. USA* **110**, 7476–7481 (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**, 23161–23172 (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. USA* **93**, 10505–10509 (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 β 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**, 1044–1049 (2013).
33. Sinner, M.F. *et al.* Integrating genetic, transcriptional, and functional analyses to identify five novel genes for atrial fibrillation. *Circulation* **130**, 1225–1235 (2014).
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. USA* **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^+/Ca^{2+} exchangers in regulation of cytoplasmic free Ca^{2+} 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. *Nat. Genet.* **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**, e38663 (2012).
41. The Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011).
42. O'Donnell, 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. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* **43**, 1082–1090 (2011).
45. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer disease. *Nature* **507**, 448–454 (2014).
46. 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).
47. 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).
48. 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).
49. 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).
50. Schlossmann, J. *et al.* Regulation of intracellular calcium by a signalling complex of IRAG, IP3 receptor and cGMP kinase Iβ. *Nature* **404**, 197–201 (2000).
51. 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.* **46**, 989–993 (2014).
52. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
53. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
54. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
55. 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).
56. Bulik-Sullivan, B.K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
57. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
58. Magi, R., Lindgren, C.M. & Morris, A.P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34**, 846–853 (2010).
59. Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
60. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
61. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
62. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
63. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
64. 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).
65. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
66. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
67. Winsvold, B.S. *et al.* Genetic analysis for a shared biological basis between migraine and coronary artery disease. *Neurol. Genet.* **1**, e10 (2015).
68. Malik, R. *et al.* Shared genetic basis for migraine and ischemic stroke: a genome-wide analysis of common variants. *Neurology* **84**, 2132–2145 (2015).
69. 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).
70. Olesen, J., Burstein, R., Ashina, M. & Tfelt-Hansen, P. Origin of pain in migraine: evidence for peripheral sensitisation. *Lancet Neurol.* **8**, 679–690 (2009).
71. Hadjikhani, N. *et al.* Mechanisms of migraine aura revealed by functional MRI in human visual cortex. *Proc. Natl. Acad. Sci. USA* **98**, 4687–4692 (2001).
72. Lauritzen, M. Pathophysiology of the migraine aura. The spreading depression theory. *Brain* **117**, 199–210 (1994).
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**, 540–553 (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).

[1]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [2]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 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, Massachusetts, USA. [6]FORMI, Oslo University Hospital, Oslo, Norway. [7]Department of Neurology, Oslo University Hospital, Oslo, Norway. [8]Institute of Clinical Medicine, University of Oslo, 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, Massachusetts, USA. [12]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [13]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. [14]Illumina, San Diego, California, USA. [15]Pediatric Neurology, Vall d'Hebron Research Institute, Barcelona, Spain. [16]Folkhälsan Institute of Genetics, Helsinki, Finland. [17]Neuroscience Center, University of Helsinki, Helsinki, Finland. [18]Molecular Neurology Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland. [19]23andMe, Inc., Mountain View, California, USA. [20]Institute of Public Health, Charité–Universitätsmedizin Berlin, Berlin, Germany. [21]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [22]deCODE Genetics, Reykjavik, Iceland. [23]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [24]Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands. [25]Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands. [26]Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland. [27]Department of Neurology and Epileptology, Hertie-Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany. [28]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität München, Munich, Germany. [29]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [30]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [31]Institute of Human Genetics, Ulm University, Ulm, Germany. [32]Center for Life Course Epidemiology and Systems Medicine, University of Oulu, Oulu, Finland. [33]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [34]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [35]Department of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [36]Department of Clinical Chemistry, Fimlab Laboratories, School of Medicine, University of Tampere, Tampere, Finland. [37]Department of Public Health, University of Helsinki, Helsinki, Finland. [38]Harvard Medical School, Boston, Massachusetts, USA. [39]Department of Neurology, University Duisburg–Essen, Essen, Germany. [40]Landspitali University Hospital, Reykjavik, Iceland. [41]Department of Psychiatry, VU University Medical Centre, Amsterdam, the Netherlands. [42]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [43]Department of Neurosurgery, NeuroCenter, Kuopio University Hospital, Kuopio, Finland. [44]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [45]MRC Functional Genomics Unit, Department of Physiology, Anatomy & Genetics, Oxford University, Oxford, UK. [46]Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK. [47]Oxford Headache Centre, John Radcliffe Hospital, Oxford, UK. [48]Max Planck Institute of Psychiatry, Munich, Germany. [49]Institute of Clinical Molecular Biology, Christian Albrechts University, Kiel, Germany. [50]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. [51]Institute of Human Genetics, Technische Universität München, Munich, Germany. [52]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [53]National Institute for Health and Welfare, Helsinki, Finland. [54]Institute of Clinical Medicine, University of Helsinki, Helsinki, Finland. [55]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [56]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [57]Department of Pain Management and Research, Oslo University Hospital, Oslo, Norway. [58]Medical Faculty, University of Oslo, Oslo, Norway. [59]Department of Ageing and Health, Norwegian Institute of Public Health, Oslo, Norway. [60]Kiel Pain and Headache Center, Kiel, Germany. [61]Danish Headache Center, Department of Neurology, Rigshospitalet, Glostrup Hospital, University of Copenhagen, Copenhagen, Denmark. [62]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, University of Copenhagen, Roskilde, Denmark. [63]Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Copenhagen, Denmark. [64]Institute of Clinical Sciences, Faculty of Medicine and Health Sciences, University of Copenhagen, Copenhagen, Denmark. [65]iPSYCH—The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Copenhagen, Denmark. [66]A full list of members and affiliations appears at the end of the paper and in the **Supplementary Note**. [67]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [68]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,

Finland. [69]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [70]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. [71]Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPE) Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [72]Biocenter Oulu, University of Oulu, Oulu, Finland. [73]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [74]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [75]Population Health Research Institute, St George's, University of London, London, UK. [76]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [77]Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands. [78]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [79]Statistical and Genomic Epidemiology Laboratory, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland, Australia. [80]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA. [81]These authors contributed equally to this work. [82]These authors jointly supervised this work. Correspondence should be addressed to A.P. (aarno.palotie@helsinki.fi).

The International Headache Genetics Consortium

Verneri Anttila[2,3], Ville Artto[26], Andrea Carmine Belin[29], Dorret I Boomsma[24], Sigrid Børte[83], Daniel I Chasman[38], Lynn Cherkas[33], Anne Francke Christensen[61], Bru Cormand[84], Ester Cuenca-Leon[2,3], George Davey Smith[23], Martin Dichgans[28], Cornelia van Duijn[34], Else Eising[77], Tonu Esko[10], Ann-Louise Esserlind[61], Michel Ferrari[25], Rune R Frants[77], Tobias M Freilinger[27], Nicholas A Furlotte[19], Padhraig Gormley[2,3], Lyn Griffiths[85], Eija Hamalainen[9], Thomas Folkmann Hansen[61], Marjo Hiekkala[16], M Arfan Ikram[34,35,70], Andres Ingason[22], Marjo-Riitta Järvelin[32,72,73], Risto Kajanne[9], Mikko Kallela[26], Jaakko Kaprio[9], Mari Kaunisto[16], Christian Kubisch[74], Mitja Kurki[2,3], Tobias Kurth[38], Lenore Launer[86], Terho Lehtimaki[36], Davor Lessel[74], Lannie Ligthart[24], Nadia Litterman[19], Arn M J M van den Maagdenberg[25,77], Alfons Macaya[15], Rainer Malik[28], Massimo Mangino[33], George McMahon[23], Bertram Muller-Myhsok[48,76,93], Benjamin M Neale[2,3], Carrie Northover[19], Dale R Nyholt[85], Jes Olesen[61], Aarno Palotie[2,3], Priit Palta[9], Linda M Pedersen[6], Nancy Pedersen[87], Danielle Posthuma[88], Patricia Pozo-Rosich[89,94], Alice Pressman[90], Lydia Quaye[33], Olli Raitakari[91], Markus Schürks[38], Celia Sintas[84], Kari Stefansson[22], Hreinn Stefansson[22], Stacy Steinberg[22], David Strachan[75], Gisela M Terwindt[25], Marta Vila-Pueyo[15], Maija Wessman[16], Bendik S Winsvold[6–8], William Wrenthal[92], Huiying Zhao[85] & John-Anker Zwart[6–8]

[83]Department of Neurology, Oslo University Hospital and University of Oslo, Oslo, Norway. [84]Department of Genetics, University of Barcelona, Barcelona, Spain. [85]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [86]National Institute on Aging, Bethesda, Maryland, USA. [87]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [88]Department of Clinical Genetics, Vrije Universiteit, Amsterdam, the Netherlands. [89]Headache Research Group, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain. [90]Sutter Health, Sacramento, California, USA. [91]Department of Medicine, University of Turku, Turku, Finland. [92]Broad Institute of MIT and Harvard, Cambridge, USA. [93]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [94]Headache Unit, Neurology Department, Vall d'Hebron University Hospital, Barcelona, Spain.

ONLINE METHODS

Study design and phenotyping. A description of the study design, ascertainment, and phenotyping for each GWA study is provided in the **Supplementary Note**.

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 calling pipelines necessitated that the QC parameters be tuned separately for each study. At a minimum, we excluded markers that had high 'missingness' rates (>5%), that had low minor allele frequency (<1%), and that failed a test of Hardy–Weinberg equilibrium. We also excluded individuals with a high proportion of missing genotypes (>5%) and used identity-by-descent estimates to remove related individuals (identity by descent > 0.185). A summary of the genotyping platforms, QC, and software used in each study is provided in **Supplementary Table 3**. To control for population stratification within each study, we merged the genotypes passing QC filters with HapMap III data from three populations: European (CEU), Asian (CHB+JPT), and African (YRI). We then performed a principal-components analysis on the merged data set and excluded any (non-European) population outliers. To control for any sub-European population structure, we performed a second principal-components analysis within each study to ensure that cases and controls were clustering together. Any principal components that were significantly associated with the phenotype were included as covariates in the model when we calculated test statistics for the meta-analysis (**Supplementary Table 4**).

Imputation. After study-level QC, estimated haplotypes were phased for each individual using (in most instances) the program SHAPEIT[78]. Missing genotypes were then imputed into these haplotypes using the program IMPUTE2 (ref. 79) and a mixed-population 1000 Genomes Project[16] reference panel (March 2012, phase 1, v3 release or later). A minority of contributing studies used alternative programs for phasing and imputation (BEAGLE[80], MACH[81], MINIMAC[82], or in-house custom software). A summary of software and procedures used is 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 when appropriate (**Supplementary Table 4**). Because age information was not available for all individuals from all studies, we were not able to adjust for it in our models. However, we note that all of the GWA studies included adults past the typical age of onset; thus age was at most a non-confounding factor, and false positive rates would therefore not be affected by its inclusion or exclusion. For the within-study association analyses, we used SNPTEST, PLINK, or R. The program GWAMA was then used to perform a fixed-effects meta-analysis weighted by the inverse variances to obtain a combined effect size, standard error, and *P* value at each marker. We excluded markers in any 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 markers that were missing from more than half of all studies (12 or more) or that exhibited high heterogeneity (heterogeneity index I^2 > 0.75). After filtering, 8,045,569 total markers were tested in the meta-analysis.

Chromosome X meta-analysis. Because of the different ploidy of males and females on chromosome X, we implemented a model of X-chromosome inactivation that assumes an equal effect of alleles in both males and females. We achieved this by scaling male dosages to the range of 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 sample size was smaller than that for the autosomal data 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.

LD-score regression analysis. We conducted a univariate heritability analysis based on summary statistics using LD-score regression (LDSC)[56] v1.0.0. For this analysis, we extracted high-quality common SNPs from the summary statistics by filtering the data according to the following criteria: presence among the HapMap Project Phase 3 SNPs[83], allele matching to 1000 Genomes Project data, no strand ambiguity, INFO score > 0.9, MAF ≥ 1%, and missingness less than two-thirds of the 90th percentile of the total sample size. The HLA region (chromosome 6, 25–35 Mb) was excluded from the analysis. With these data, we used LDSC to quantify the proportion of the total inflation in chi-square statistics that could 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[56].

Heterogeneity analysis of migraine subtypes. To determine whether 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 two groups[58]. Because a large proportion of the controls were shared between the original data sets for migraine with aura and migraine without aura (**Table 1**), for this analysis we created two additional subsets of the migraine subtype data that contained no overlapping controls (**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 of 4,833 cases and 106,834 controls. To assess the heterogeneity observed, we chose the 44 index SNPs from the primary meta-analysis and applied the subtype-differentiated meta-analysis method to them. We observed that only 7 out of the 44 SNPs exhibited heterogeneity in the subtype-differentiated test (heterogeneity *P* value < 0.05; **Supplementary Table 13**), suggesting that most loci probably affect risk for both subtypes.

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 described in detail elsewhere[53,59] and is outlined briefly here. Assume D represents the data including the genotype matrix X for all of the P variants (the genotype for variant j is denoted as $\boldsymbol{x}_j$) and disease status Y (for N individuals), and β represents the model parameters. We define the model, denoted by A, as the causal status for all of the P variants in the locus: $A \equiv \{a_j\}$, in which a_j is the causal status for variant j. $a_j = 1$ if the variant j is causal, and $a_j = 0$ if it is not. We assume that there is one and only one genuine signal for each locus; therefore, one and only one of the P variants is causal: $\Sigma_j a_j = 1$. For convenience, we define A_j as the model in which only variant j is causal, and A_0 as the model in which no variant is causal (null model). The probability of model A_j (where variant j is the only causal variant in the locus) given the data can be calculated using Bayes's rule:

$$\Pr(A_j \mid D) = \int_{\beta} \Pr(D, \beta \mid A_j) \cdot \frac{\Pr(A_j)}{\Pr(D)} \cdot d\beta \qquad (1)$$

We estimate equation (1) using the steepest-descent approach[84]. Making the assumption of a flat prior on the model parameters, we approximate the integral over the model parameters using their maximum likelihood estimator ($\hat{\beta}_j$):

$$\hat{\beta}\ \Pr(A_j \mid D) \approx \Pr(D \mid A_j, \hat{\beta}_j) \cdot N^{-|\beta_j|/2} \cdot \frac{\Pr(A_j)}{\Pr(D)} \qquad (2)$$

where the sample size is denoted by N and the number of fitted parameters for model A_j is denoted by $|\beta_j|$. $|\beta_j|$ is a constant because model A_j has the same number of parameters across all variants. In the framework of a generalized linear model, the deviance for two nested models follows an approximate chi-square distribution. We therefore define χ_j^2 as the deviance comparing the null model and the model in which variant j is causal:

$$\chi_j^2 \equiv -2 \log \frac{\Pr(D \mid A_0, \hat{\beta}_0)}{\Pr(D \mid A_j, \hat{\beta}_j)} \qquad (3)$$

We further show that χ_j^2 can be calculated as the chi-square statistic of fitting a binomial model with the disease status (Y) as the dependent variable and the genotype of variant j as the explanatory variable:

$$\begin{aligned}\hat{\beta}\ \chi_j^2 &= -2\log\frac{\Pr(\{x_i\},Y\mid\{a_i=0\},\{\hat{\beta}_{i,0}\})}{\Pr(\{x_i\},Y\mid\{a_j=1,a_{-j}=0\},\{\hat{\beta}_j,\hat{\beta}_{-j,0}\})}\\ &= -2\log\frac{\prod_i\Pr(x_i,Y\mid a_i=0,\hat{\beta}_{i,0})}{\Pr(x_j,Y\mid a_j=1,\hat{\beta}_j)\prod_{i\neq j}\Pr(x_i,Y\mid a_i=0,\hat{\beta}_{i,0})}\\ &= -2\log\frac{\Pr(x_j,Y\mid a_j=0,\hat{\beta}_{j,0})}{\Pr(x_j,Y\mid a_j=1,\hat{\beta}_j)}\end{aligned} \quad (4)$$

$\Pr(A_j|D)$ in equation (2) is then a function of χ_j^2:

$$\Pr(A_j\mid D)\approx\exp\left(\frac{\chi_j^2}{2}\right)\cdot l_0\cdot N^{-|\beta_j|/2}\cdot\frac{\Pr(A_j)}{\Pr(D)} \quad (5)$$

where $l_0=\Pr(D\mid A_0,\hat{\beta}_0)$. We make the assumption that the prior causal probability for all variants is equal, that is, $\Pr(A_j)$ is the same across all variants j. Equation (5) can then be simplified with a constant for the term $l_0\cdot N^{-|\beta_j|/2}\cdot(\Pr(A_j)/\Pr(D))$, and the probability that variant j is causal can be calculated using

$$\Pr(A_j\mid D)\propto\exp\left(\frac{\chi_j^2}{2}\right) \quad (6)$$

which can be normalized across all variants as

$$\mathrm{P}(A_j)\equiv\Pr(A_j\mid D)/\Sigma_k\Pr(A_k\mid D) \quad (7)$$

Finally, the 99% credible set of variants is defined as the smallest set of models, with each model designating one causal variant, $\mathrm{S}=\{A_j\}$, such that

$$\sum_{A_j\in\mathrm{S}}\mathrm{P}(A_j)\geq 99\% \quad (8)$$

This credible set of variants has 99% probability of containing the causal variant, given the assumption that there is a true association and that all possible causal variants have been genotyped (both assumptions are likely to be valid in genome-wide significant regions of data that have been imputed to the 1000 Genomes Project). We have made the R script for implementing the method freely available online (see "URLs").

eQTL credible set overlap analysis. To assess whether the association statistics in the 38 migraine loci could be explained by credible overlapping eQTL signals, we used two eQTL microarray data sets. The first consisted of 3,754 samples from peripheral venous blood[85], and the second was from a meta-analysis of human brain cortex studies of a total of 550 samples[86]. 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. Then, for the most significant eQTLs ($P<1\times10^{-4}$) found for genes within a 1-Mb window of migraine credible set variants (see "Defining credible sets"), 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 that the migraine locus tagged a real eQTL. An appropriate significance threshold for multiple testing was determined by Bonferroni correction.

GTEx tissue enrichment analysis. We obtained gene sets for each locus by taking all genes within 50 kb of credible-set SNPs. We then analyzed identified genes for tissue enrichment using publicly available expression data from the pilot phase of the GTEx project[62], version 3. In this data set, postmortem samples from 42 human tissues and three cell lines across 1,641 samples (**Supplementary Table 16**) were 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) for 52,577 transcripts were filtered for those with unique HGNC IDs (n = 20,932). We also excluded transcripts from any noncoding RNAs. We ranked all transcripts by mean RPKM across all samples, and we generated 100,000 permutations of each credible set gene list 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. We calculated enrichment P values for each tissue by taking the total number of instances when the gene list of interest had a lower sum of ranks than the permuted sum of ranks (divided by the total number of permutations). We estimated the number of independent tissues using the matSpD tool[87] and then used Bonferroni correction to adjust for 27 independent tests ($P<1.90\times10^{-3}$).

Specificity of individual genes in GTEx tissues. We selected the nearest gene to the index SNP at each migraine locus and then investigated the individual expression activity of each of the selected genes. Because 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**). For each selected gene, we then tested whether the average expression (mean RPKM) was significantly higher in a particular tissue group compared with the 'other tissues' category. We assessed significance using a one-tailed t-test and used Bonferroni correction to adjust for 114 tests (38 genes × 3 tissue groups). Although some genes were observed to be significantly expressed in multiple tissue groups, we determined that a gene was tissue specific if it had high expression in only one tissue group (i.e., brain, vascular, or gastrointestinal; **Supplementary Table 25**).

eQTL credible set analysis in GTEx tissues. For all tissues and transcripts (filtered as described above), we identified genome-wide significant ($P<2\times10^{-13}$) *cis*-eQTLs within a 1-Mb window of each transcript and created credible sets (see "Defining credible sets") for each eQTL identified in each tissue. We found a total of 35 of these significant eQTL credible sets within a 1-Mb window of the migraine loci; however, only 7 out of the 35 contained variants that overlapped with a migraine credible set. For these seven eQTL credible sets, we then tested (using Spearman's rank correlation) whether 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 that the migraine locus tagged a real eQTL. Multiple testing was controlled for the use of Bonferroni correction (i.e., for seven tests at $P<7.1\times10^{-3}$).

Enhancer enrichment analysis. Markers of gene regulation were defined using ChIP-seq data sets from ENCODE[66] and the NIH Roadmap Epigenome[65] projects. On the basis of the histone H3K27ac signal, which identifies active enhancers, we processed data from 56 cell lines and tissue samples to identify cell-type-specific and 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[88]. The raw data are publicly available (see "URLs"), and a description of the 56 tissues and cell types is provided in **Supplementary Table 21**. We mapped the credible set variants at each migraine locus to these enhancer sites and compared the overlap observed with tissue-specific enhancers relative to a background of 10,000 randomly selected sets of SNPs of equal size. We restricted the background selection to 1000 Genomes Project variants (MAF > 1%) that also passed QC filters in the meta-analysis (to allow the selection of only SNPs that had an *a priori* chance of being associated). The selection procedure then involved randomly selecting genomic regions with length and density of enhancers equivalent to those 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 mapped to an equal number of enhancer sites (of any tissue type) as credible SNPs from the original locus, then they were added to the background set of SNPs for comparison. If the selected SNPs did not map to

the correct number of enhancers, the selection procedure was repeated until an appropriate set was found. This procedure was repeated 10,000 times for each locus to obtain an empirical null distribution. We then estimated the enrichment significance empirically by calculating the proportion of replicates that were greater than the observed value. Finally, we used Bonferroni correction to adjust for multiple testing of 56 tissue and cell types ($P < 8.9 \times 10^{-4}$).

GO enrichment analysis. The set of 38 genes nearest to the index SNP in each migraine locus was chosen and tested for over-representation in Gene Ontology (GO) annotations. We used the PANTHER[89] tool (see "URLs") to perform the analysis, implementing a binomial test to determine whether the number of genes from the migraine test set found in each GO pathway was likely to have occurred by chance alone. The association *P* values were adjusted for the number of pathways tested by Bonferroni correction.

DEPICT reconstituted gene set enrichment analysis. DEPICT[63] (Data-driven Expression Prioritized Integration for Complex Traits) is a computational tool that, given a set of GWA study summary statistics, allows prioritization of genes in associated loci, enrichment analysis of reconstituted gene sets, and tissue enrichment analysis. DEPICT was run using 124 independent genome-wide significant SNPs as input (PLINK clumping parameters: --clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.6 --clump-kb 250; note that 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 that 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 FDR estimation, normalized expression data from 77,840 Affymetrix microarrays for gene set reconstitution (see ref. 90), 14,461 reconstituted gene sets for gene set enrichment analysis, and testing of 209 tissue/cell types assembled from 37,427 Affymetrix U133 Plus 2.0 array samples for enrichment in tissue/cell-type expression.

After the 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 were expected to be enriched in the reconstituted gene set. Therefore, a lack of enrichment complicated interpretation because the label of the reconstituted gene set may have been inaccurate. Hence eight reconstituted gene sets were removed from the results: MP:0002089, MP:0002190, ENSG00000151577, ENSG00000168615, ENSG00000143322, ENSG00000112531, ENSG00000161021, and ENSG00000100320. We also removed an association identified for another gene set (ENSG00000056345 PPI, $P = 1.7 \times 10^{-4}$, FDR = 0.04) because it was no longer part of the Ensembl database. Finally, we used the Affinity Propagation tool[91] to cluster related reconstituted gene sets into ten groups (see "URLs").

DEPICT tissue enrichment analysis. DEPICT used data from 37,427 human microarray samples captured on the Affymetrix HGU133a2.0 platform to test whether genes in the 38 migraine loci were highly expressed in 209 tissues and 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[63]. The tissue/cell-type enrichment analysis algorithm was conceptually identical to the gene set enrichment analysis whereby enrichment *P* values were calculated empirically using 500 permutations for bias adjustment and 20 replications for FDR estimation.

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/content/datasets-and-cohorts). For access to deeper-level data, please contact the data access committee (fimm-dac@helsinki.fi).

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. *Nat. 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. *Nat. Genet.* **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. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
85. Wright, F.A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437 (2014).
86. 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).
87. 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).
88. Farh, K.K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
89. Mi, H., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
90. Fehrmann, R.S.N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
91. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).