Migraine is a highly prevalent brain disorder characterized by disabling attacks of moderate-to-severe pulsating and usually one-sided headache that may be aggravated by physical activity, and can be associated with symptoms such as a hypersensitivity to light and sound, nausea and vomiting. Migraine has a lifetime prevalence of 15–20% and is ranked as the second most disabling condition in terms of years lived with disability.

Migraine affects over a billion individuals worldwide but its genetic underpinning remains largely unknown. Here, we performed a genome-wide association study of 102,084 migraine cases and 771,257 controls and identified 123 loci, of which 86 are previously unknown. These loci provide an opportunity to evaluate shared and distinct genetic components in the two main migraine subtypes: migraine with aura and migraine without aura. Stratification of the risk loci using 29,679 cases with subtype information indicated three risk variants that seem specific for migraine with aura (in \textit{HMOX2}, \textit{CACNA1A} and \textit{MPPED2}), two that seem specific for migraine without aura (near \textit{SPINK2} and near \textit{FECH}) and nine that increase susceptibility for migraine regardless of subtype. The new risk loci include genes encoding recent migraine-specific drug targets, namely calcitonin gene-related peptide (\textit{CALCA}/\textit{CALCB}) and serotonin 1F receptor (\textit{HTR1F}). Overall, genomic annotations among migraine-associated variants were enriched in both vascular and central nervous system tissue/cell types, supporting unequivocally that neurovascular mechanisms underlie migraine pathophysiology.
of migraine signals near activating histone marks specific to car-
vascular and central nervous system (CNS) tissues, as well as by loci that are shared by both subtypes. Our findings
confirmed both by risk loci that seem specific for only one subtype in MA. Our subtype data show compellingly that migraine risk is
underlying causes are. Prevailing theories about migraine patho-
physiology emphasize neuronal and/or vascular dysfunction. Current knowledge of disease mechanisms comes largely from studies of a rare monogenic subform of MA—familial hemiplegic migraine—for which three ion transporter genes (CACNA1A, ATPIA2 and SCN1A) have been identified. The common forms of migraine (MA and MO) instead have a complex polygenic architecture with an increased familial relative risk, increased concordance in monzygotic twins and a heritability of 40–60%. The largest genome-wide association study (GWAS) thus far, with 59,674 cases and 316,078 controls, reported 38 genomic loci that confer migraine risk. Subsequent analyses of these GWAS data showed enrichment of migraine signals near activating histone marks specific to cardiovascular and central nervous system (CNS) tissues, as well as for genes expressed in vascular and smooth muscle tissues. Other smaller GWAS have suggested ten additional loci. Of note, the previous datasets were too small to perform a meaningful comparison of the genetic background between migraine subtypes.

As migraine is globally the second largest contributor to years lived with disability, there is clearly a large need for new treatments. Triptans, that is, serotonin 5-HT1B/1D receptor agonists, are migraine-specific acute treatments for the headache phase but are not effective in every patient, whereas preventive medication is far from satisfactory. Recent promising alternatives for acute treatment are serotonin 5-HT1B receptor agonists ('dihydroergotamine' triptans) and small-molecule calcitonin gene-related peptide (CGRP) receptor antagonists ('gepants'). For preventive treatment, monoclonal antibodies (mAbs) targeting CGRP or its receptor have recently proven effective, and new gepants for migraine prevention are under development. Still, there remains an urgent need for treatment options for patients who do not respond to existing treatments. Genetics has proven promising in developing new therapeutic hypotheses in other prevalent complex diseases, such as cardiovascular disease and type 2 diabetes, and we anticipate that large genetic studies of migraine could also yield similar insights.

We conducted a GWAS meta-analysis of migraine by adding to the previous meta-analysis 42,410 new migraine cases from four study collections (Table 1). This increased the number of migraine cases by 71% for a total sample of 102,084 cases and 771,257 controls. Furthermore, we assessed the subtype specificity of the risk loci in 8,292 new MA and 6,707 new MO cases in addition to the 6,332 MA and 8,348 MO cases used previously (Table 2). Here, we report 123 genomic loci, of which 86 are previously unknown, and include the first four loci that reach genome-wide significance ($P < 5 \times 10^{-8}$) in MA. Our subtype data show compellingly that migraine risk is conferred both by risk loci that seem specific for only one subtype as well as by loci that are shared by both subtypes. Our findings also include new risk loci containing target genes of recent migraine drugs acting on the CGRP pathway and the serotonin 5-HT1B receptor. Finally, our data support the concept that migraine is brought about by both neuronal and vascular genetic factors, strengthening the view that migraine is truly a neurovascular disorder.

**Results**

**Genome-wide meta-analysis.** We combined data on 873,341 individuals of European ancestry (102,084 cases and 771,257 controls) from five study collections (Table 1 and Supplementary Table 1) and analyzed 10,843,197 common variants (Methods). Despite different approaches to the ascertainment of migraine cases across studies, pairwise genetic correlations were all near 1 (Supplementary Table 2), as determined by linkage disequilibrium (LD) score (LDSC) regression, showing high genetic and phenotypic similarity across the studies,
justifying their meta-analysis. Pairwise LDSC intercepts were all near 0, indicating little or no sample overlap (Supplementary Table 2).

The genomic inflation factor ($\lambda_{G}$) of the fixed-effect meta-analysis results was 1.33 (Supplementary Fig. 1), which is in line with other large meta-analyses\(^31\)–\(^33\) and is as expected for a polygenic trait\(^34\). The univariate LDSC\(^35\) intercept was 1.05 (s.e. 0.01), which, being close to 1.0, suggests that most of the genome-wide elevation of the association statistics comes from true additive polygenic effects rather than from a confounding bias such as population stratification. The LDSC analysis showed a linear trend between the variant’s LD score and its association with migraine, as expected from a highly polygenic phenotype such as migraine (Supplementary Fig. 2). The SNP heritability estimate from LDSC was 11.2% (95% confidence interval (CI) 10.8–11.6%) on a liability scale when assuming a population prevalence of 16%.

We identified 8,117 genome-wide significant (GWS; $P < 5 \times 10^{-8}$) variants represented by 170 LD-independent index variants ($r^2 < 0.1$). We defined the risk loci by including all variants in high LD ($r^2 > 0.6$) with the index variants and merged loci that were closer than 250 kb (Methods). This resulted in 123 independent risk loci (Fig. 1, Supplementary Table 3a and Supplementary Data 1 and 2). Of the 123 loci, 86 are previously unknown, whereas 36 overlap with the previously reported 47 autosomal risk loci (Supplementary Table 8). Of the 123 loci, 86 were previously unknown, whereas 36 overlap with the previously reported 47 autosomal risk loci (Supplementary Table 4) and one with the previously reported X chromosome risk loci (Supplementary Table 5).

In addition, we conducted an approximate stepwise conditional analysis for the 123 risk loci (Methods). Since sample sizes per variant varied considerably, we restricted the conditional analysis to variants with similar effective sample sizes to the lead variant. The conditional analysis returned six single nucleotide polymorphisms (SNPs) within the 123 risk loci that remained GWS after conditioning on the lead variants (Supplementary Table 6a,b).

**Characterization of migraine risk loci.** We mapped the 123 risk loci to genes by their physical location using the Ensembl Variant Effect Predictor (VEP)\(^36\). Of the lead variants, 59% (72/123) were within a transcript of a protein-coding gene, and 80% (99/123) of the loci contained at least one protein-coding gene within 20 kb, and 93% (114/123) within 250 kb (Supplementary Table 3). Of the 123 lead variants, 5 were missense variants (in genes PLC\(\varepsilon\)1, MRG\(\varepsilon\)PRE, SERP\(\varepsilon\)IN\(\varepsilon\), Z\(\varepsilon\)BT\(\varepsilon\)4 and Z\(\varepsilon\)NF\(\varepsilon\)62), and 40 more missense variants were in high LD ($r^2 > 0.6$) with the lead variants (Supplementary Table 7a). Of note, three variants with a predicted high impact consequence on protein function were in high LD with the lead variants: (1) a stop gained variant (rs43458) with lead variant rs42854 ($r^2 = 0.85$) in gene ANK\(\varepsilon\)DD\(\varepsilon\)1B, (2) a splice donor variant (rs66880209) with lead variant rs1472662 ($r^2 = 0.71$) in RP11-420K8.1 and (3) a splice acceptor variant (rs11042902) with lead variant rs4910165 ($r^2 = 0.69$) in MR\(\varepsilon\)VI\(\varepsilon\)1 (Supplementary Table 7b).

We used stratified LDSC (S-LDSC) to partition migraine heritability by 24 functional genomic annotations\(^29\)–\(^30\). We observed enrichment for ten categories (Supplementary Fig. 3 and Supplementary Table 8), with conserved regions showing the highest enrichment (11.2-fold; $P = 1.95 \times 10^{-10}$), followed by coding regions (8.1-fold; $P = 1.36 \times 10^{-3}$) and enhancers (4.2-fold; $P = 3.64 \times 10^{-4}$).

**Prioritization of candidate genes.** We mapped the 123 lead variants to genes via expression quantitative trait locus (eQTL) association using GTEx v.8 (ref.\(^{39}\)) and data repositories included in FUMA\(^40\) at a false discovery rate (FDR) of 5% (Methods). The lead variants were cis-eQTLs for 589 genes (Supplementary Table 9), and variants in high LD with the lead variants were cis-eQTLs for an additional 624 genes (Supplementary Table 10). In total, 84% (103/123) of lead variants were cis-eQTLs for at least one gene. Tibial artery had the highest number (47/123) of lead variants as cis-eQTLs in GTEx v.8, and it was the only tissue type where the enrichment was statistically higher ($P = 6.37 \times 10^{-10}$) than expected based on the overall number of cis-eQTLs per tissue reported by GTEx (Supplementary Fig. 4 and Supplementary Note).

To prioritize candidate genes for the risk loci, we applied two approaches based on GTEx v.8 expression data: fine-mapping of causal gene-sets by FOCUS\(^41\) (Supplementary Table 11a) and a transcriptionome-wide association study (TWAS) by S-PrediXcan\(^42\) combined with colocalization analysis using COLOC\(^43\) (Supplementary Table 11b).

With posterior probability (PP) $> 0.5$, FOCUS found candidate genes for 82 loci and S-PrediXcan + COLOC supported colocalization for 52 loci (Supplementary Table 11c). In total, 73 genes in 46
Two of the new risk loci contain genes (CALCA, CALCB, and HTR1F) whose product proteins are closely related to targets of two migraine-specific drug therapies\(^1\). We observe a convincing association at the chromosome 11 locus that contains the CALCA and CALCB genes encoding CGRP itself (lead SNP rs1003194, \(P=2.43 \times 10^{-16}\); Fig. 2a), while none of the genes encoding CGRP receptor proteins (CALCR, RAMP1 or RCP) show a statistically comparable association (all \(P > 10^{-4}\); Supplementary Fig. 5). Variant rs1003194 is a cis-eQTL for CALCB, and also for three other genes (CGGBP1, PDE3B, and INSC; Supplementary Table 9) and FOCUS prioritizes CALCA, CALCB and INSC (Supplementary Table 11c). In addition, a new locus on chromosome 3 contains HTR1F (lead SNP rs6795209, \(P=1.23 \times 10^{-4}\); Fig. 2b), which encodes the serotonin 5-HT\(_1\)F receptor. Variant rs6795209 is a significant cis-eQTL for HTR1F, as well as for three other genes (CGGBP1, ZNF654, C3orf58) in the same locus (Supplementary Table 9). FOCUS or S-PrediXcan + COLOC did not prioritize HTR1F based on gene expression data (Supplementary Table 11c).

**Migraine subtypes with aura and without aura.** Previously, Gormley et al.\(^{13}\) conducted subtype-specific GWAS with 6,332 MA cases against 144,883 controls and 8,348 MO cases against 139,622 controls, and reported that seven loci were GWS in MO but none were GWS in MA. Here, we added to the previous data 8,292 new MA and 6,707 new MO cases from headache specialist centers in Denmark and the Netherlands as well as from study collections in Iceland and UK Biobank (Table 2), for total sample sizes of 14,624 MA cases and 703,852 controls, and 15,055 MO cases and 682,301 controls. We estimated the effect size for each subtype at the 123 lead variants of the migraine GWAS (Supplementary Table 3b,c and Supplementary Data 4 and 5) and detected four GWS variants in the MA meta-analysis and 15 GWS variants in the MO meta-analysis. We also estimated a probability that the lead variant is either subtype-specific (that is, associated only with MO or with MA but not with both), shared by both subtypes, or not associated with either subtype (Methods; Supplementary Table 12a and Supplementary Data 6). With a probability above 95%, three lead variants (that is, rs12598836 in the HMOX2 locus, rs10405121 in the CACNA1A locus and rs11031122 in the MPPED2 locus) are MA-specific, while two lead variants (that is, rs7684253 in the locus near SPINK2 and rs8087942 in the locus near FECH) are MO-specific at a similar threshold. Nine lead variants were shared by MA and MO with >95% probability (Fig. 3a). In addition to the five subtype-specific lead variants, four other lead variants also showed differences in effect size between the subtypes (\(P<0.05/123\)) (Fig. 3b).

**Phenome-wide association scans with National Human Genome Research Institute GWAS Catalog and FinnGen R4.** Next, we conducted phenome-wide association scans (PheWAS) for the lead variants for 4,314 traits with reported associations in the National Human Genome Research Institute GWAS catalog, and 17 lead variants with 26 defined disease categories in FinnGen at \(P<1 \times 10^{-5}\). The categories with the highest number of reported associations were cardiovascular disease (7 lead variants) and blood pressure (6 lead variants) in the GWAS catalog, and diseases of the circulatory system (11 lead variants) in FinnGen. When we performed PheWAS for all variants in high LD (\(r^2 > 0.6\) with the lead variants, we observed associations for 79 loci with 54 different phenotype categories in the GWAS Catalog, and for 41 loci with 26 disease categories in FinnGen (Supplementary Table 13a and Supplementary Fig. 6).

These findings are consistent with previous results that migraine is a risk factor for several cardiovascular traits\(^{45-47}\), and genetically correlated with blood pressure\(^{48-49}\). However, we did not observe a trend in the direction of the allelic effects between migraine and coronary artery disease (CAD) or migraine and blood pressure traits (Supplementary Table 13d) using the latest meta-analysis of the CARDIoGRAMplusC4D Consortium\(^{50}\) (\(n=336,924\)) and blood pressure GWAS from UK Biobank\(^{51}\) (\(n=422,771\)).

**Enrichment in tissue or cell types and gene sets.** We used LDSC applied to specifically expressed genes (LDSC-SEG)\(^{11}\) (Methods) to...
evaluate whether the polygenic migraine signal was enriched near genes that were particularly active in certain tissue or cell types as determined by gene expression or activating histone marks. Using multi-tissue gene expression data, we found enrichment at FDR 5% in three cardiovascular tissue/cell types, that is, aorta artery (P = 1.78 × 10−4), tibial artery (P = 3.60 × 10−4) and coronary artery (P = 4.29 × 10−4) (Table 3 and Supplementary Table 14a), all of which have previously been reported enriched in migraine without aura11. The fine-scale brain expression data from GTEx, since recently including 13 brain regions, showed enrichment in the caudate nucleus of striatum—a component of basal ganglia (P = 6.02 × 10−4; Table 3 and Supplementary Table 14b). With chromatin-based annotations, we found enrichment in five CNS cell types, three cardiovascular cell types, one cell type of the digestive system, one musculoskeletal/connective cell type and ovary tissue (Table 3 and Supplementary Table 14c). In addition to replicating previous findings13,14, the signal linking to ovary tissue has not been reported before.

Finally, we used DEPICT9 to identify tissues whose eQTLs were enriched for migraine-associated variants. The tissue enrichment analysis replicated three previously reported tissues1: arteries (nominal P = 1.03 × 10−1), stomach (nominal P = 1.04 × 10−1) and upper gastrointestinal tract (nominal P = 1.29 × 10−3) (Supplementary Table 14a). Results of gene-set analyses using DEPICT2 and MAGMA15 are presented in Supplementary Tables 15 and 16.

Discussion

We conducted the largest GWAS meta-analysis on migraine thus far by combining genetic data on 102,084 cases and 771,257 controls. We identified 123 migraine risk loci, of which 86 are previously described13. This shows that we have now reached the statistical power for rapid accumulation of new risk loci for migraine, in line with the progress of GWAS seen with other common diseases14, and as expected for a highly polygenic disorder like migraine51.

Migraine subtypes MO and MA were defined as separate disease entities some 30 years ago, and, since then, the debate has continued as to what extent they are biologically similar. Over the years, arguments in favor and against have been presented, but convincing genetic evidence to support subtype-specific risk alleles has been lacking in genetic studies with smaller sample sizes18,56,57. Here, we increased considerably the evidence for subtype specificity of some risk alleles by including new migraine subtype data at the 123 migraine risk variants. We observed that, with a probability of >95%, three lead variants (in HMOX2, in CACNA1A and in MPPED2) are associated with MA but not MO. Of these variants, CACNA1A is a well-known gene linked to familial hemiplegic migraine, a rare subtype of MA58,59. The observation that CACNA1A seems involved in both monogenic and polygenic forms of migraine provides a gene-based support for the increased sharing of common variants between the two disorders53. We find no evidence that any of the seven loci, previously reported as GWS in MO but not in MA13, would be specific for MO, while four of them (LRP1, FH5L5, near FGF6 and near TRPM8) are among the nine loci shared by both subtypes with a probability over 95%. Loci (for example, LRPI and FH5L5) that are strongly associated with both subtypes provide convincing evidence for a previous hypothesis that the subtypes partly share a genetic background13,60. In accordance with our analysis, effects in both subtypes were suggested before at the TRPM8 and TSPAN2 loci, whereas, in contrast to our results, the LRPI locus was previously reported to be specific for MO14. Finally, we also detected four lead variants (including LRPI) that do not seem specific for MO but do confer a higher risk for MO than for MA.

It has been long debated whether migraine has a vascular or a neuronal origin, or whether it is a combination of both5,61,62. Here, we found genetic evidence for the role of both vascular and central nervous tissue types in migraine from several tissue enrichment analyses, which refined earlier analyses based on smaller sample sizes13,14.

With respect to a vascular involvement in the pathophysiology of migraine, both gene expression and chromatin annotation data from LDSC-SEG showed that migraine signals are enriched for genes and cell-type-specific annotations that are highly expressed...
in aorta and tibial and coronary arteries. The involvement of arteries was also proposed by our DEPICT tissue enrichment analysis. In addition, cardiovascular disease and blood pressure phenotypes were among the top categories in the PheWAS analyses. These results are consistent with previous reports of a shared etiology and some genetic correlation between migraine and cardiovascular and ovary is an interesting finding, although the statistical evidence for menstrual-related migraine (migraine72–74, the involvement of the ovary is an interesting finding, although the statistical evidence for it remains weaker at present compared with that for the vascular system and CNS.

A particularly interesting finding in our GWAS was the identification of risk loci containing genes that encode targets for migraine-specific therapeutics. One new locus contains the CALCA and CALCB genes on chromosome 11 that encode calcitonin gene-related peptide (CGRP). CGRP-related monoclonal antibodies have been successful for the preventive treatment of migraine75, and they are considered as a major breakthrough in migraine-specific treatments since the development of the triptans for acute migraine over two decades ago. Another new locus contains the HTR1F gene that encodes serotonin 5-HT1F receptor, which is the target of another recent migraine drug class called ditans76. Ditans provide a promising acute treatment, especially for those migraine patients that cannot use triptans because of cardiovascular risk factors23.

Table 3 | LDSC-SEG results that are significant at FDR 5%

| Tissue/cell type and histone mark | Tissue category | P value | FDR adjusted P value |
|-----------------------------------|----------------|---------|----------------------|
| **Multitissue gene expression data** |                 |         |                      |
| Aorta                             | Cardiovascular | 1.78 × 10^-4 | 0.029               |
| Tibial artery                     | Cardiovascular | 3.60 × 10^-4 | 0.029               |
| Coronary artery                   | Cardiovascular | 4.29 × 10^-4 | 0.029               |
| **Gene expression data of 13 brain regions from GTEx** |                 |         |                      |
| Caudate (basal ganglia)           | CNS            | 6.00 × 10^-4 | 0.008               |
| **Multitissue chromatin annotation data** |                 |         |                      |
| Fetal brain female, H3K4me3       | CNS            | 2.49 × 10^-4 | 0.012               |
| Brain dorsolateral prefrontal cortex, H3K27ac | CNS | 8.43 × 10^-5 | 0.018               |
| Brain dorsolateral prefrontal cortex, H3K4me3 | CNS | 1.11 × 10^-4 | 0.018               |
| Aorta, H3K4me1                     | Cardiovascular | 2.57 × 10^-4 | 0.031               |
| Stomach mucosa, H3K36me3           | Digestive      | 3.36 × 10^-4 | 0.032               |
| Aorta, H3K27ac                     | Cardiovascular | 4.40 × 10^-4 | 0.032               |
| Artery-tibial ENTEX, H3K4me1       | Cardiovascular | 4.53 × 10^-4 | 0.032               |
| Ganglion eminence derived primary cultured neurospheres, H3K4me3 | CNS | 6.53 × 10^-4 | 0.04               |
| Brain germinal matrix, H3K4me3     | CNS            | 8.42 × 10^-4 | 0.043               |
| Aorta ENTEX, H3K27ac               | Cardiovascular | 1.11 × 10^-3 | 0.043               |
| Artery-coronary ENTEX, H3K4me3     | Cardiovascular | 1.13 × 10^-3 | 0.043               |
| Cortex derived primary cultured neurospheres, H3K36me3 | CNS | 1.14 × 10^-3 | 0.043               |
| Ovary, H3K27ac                     | Other          | 1.15 × 10^-3 | 0.043               |
| Cortex derived primary cultured neurospheres, H3K4me3 | CNS | 1.29 × 10^-3 | 0.045               |
| Aorta ENTEX, H3K4me1               | Cardiovascular | 1.39 × 10^-3 | 0.045               |
| Stomach smooth muscle, H3K4me3     | Musculoskeletal/Connective | 1.55 × 10^-4 | 0.047               |

One-sided P value from testing whether the regression coefficient is positive. FDR, false discovery rate based on Benjamini-Hochberg method. Full results are in Supplementary Table 14a-f.
fine-mapping would require accurate LD information\cite{2}, which is typically lacking in meta-analyses and further distorted from reference panels by variation in effective sample size across variants. Second, computational approaches to gene prioritization require further methodological work\cite{8} and extension to additional sources of functional data to provide more robust and comprehensive gene prioritization results. Another limitation of our study is that a large proportion of migraine diagnoses are self-reported. Therefore, we cannot rule out misdiagnosis, such as, for example, tension headache being reported as migraine, which could overemphasize genetic factors related to general pain mechanisms and not migraine per se. Regardless, the high genetic correlation that we observed supports a strong phenotypic concordance between the study collections that also included deeply phenotyped clinical cohorts from headache specialist centers, which were instrumental for the migraine subtype analyses. While the subtype data provided convincing evidence of both loci with genetic differences and other loci with genetic overlap between subtypes, larger samples are still needed to achieve a more accurate picture of the similarities and differences in genetic architecture behind the subtypes.

To conclude, we report the largest GWAS meta-analysis of migraine so far, detecting 123 risk loci. We demonstrated that both the vascular system and CNS are involved in migraine pathophysiology, supporting the notion that migraine is a neurovascular disease. Our subtype analysis of migraine with aura and migraine without aura shows that these migraine subtypes have both shared risk alleles and risk alleles that seem specific to one subtype. In addition, new loci include two targets of recently developed and effective migraine treatments. Therefore, we expect that these and future GWAS data will reveal more of the heterogeneous biology of migraine and potentially point to new therapies against migraine—a leading burden for population health throughout the world.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00990-0.

Received: 12 January 2021; Accepted: 22 November 2021; Published online: 3 February 2022

References
1. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. Cephalalgia 38, 1–211 (2018).
2. Vois, T. et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396, 1204–1222 (2020).
3. Steiner, T. J. et al. Migraine remains second among the world’s causes of disability, and first among young women: findings from GBD2019. J. Headache Pain 21, 137 (2020).
4. Russell, M. B., Rasmussen, B. K., Thorvaldsen, P. E. R. & Olesen, J. E. S. Prevalence and sex-ratio of the subtypes of migraine. Int. J. Epidemiol. 24, 612–618 (1995).
5. Russell, M. B. & Olesen, J. Increased familial risk and evidence of genetic factor in migraine. BMJ 311, 541–544 (1995).
6. Kallel, M., Wessman, M., Havanka, H., Palotie, A. & Farkkila, M. Familial migraine with and without aura: clinical characteristics and co-occurrence. Eur. J. Neurol. 8, 441–449 (2001).
7. de Boer, I., van den Maagdenberg, A. M. J. M. & Terwindt, G. M. Advance in genetics of migraine. Curr. Opin. Neurol. 32, 413–421 (2019).
8. Tfelt-Hansen, P. C. & Koehl, P. J. One hundred years of migraine research: major clinical and scientific observations from 1910 to 2010. Headache 51, 752–778 (2011).
9. Anttila, V., Wessman, M., Kallel, M. & Palotie, A. in Handbook of Clinical Neurology, Vol. 148 (eds Geschwind, D.H., Paulson, H.L. & Klein, C.) Ch. 31 (Elsevier, 2018).
10. 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).
11. Ulrich, V., Gervil, M., Kyvix, K. O., Olesen, J. & Russell, M. B. Evidence of a genetic factor in migraine with aura: a population-based Danish twin study. Ann. Neurol. 45, 242–246 (1999).
12. Gervil, M., Ulrich, V., Kaprio, J., Olesen, J. & Russell, M. B. The relative role of genetic and environmental factors in migraine without aura. Neurology 53, 995–999 (1999).
13. Gormley, P. et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat. Genet. 48, 856–866 (2016).
14. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. 50, 621–629 (2018).
15. Anttila, V. et al. Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. Nat. Genet. 42, 869–873 (2010).
16. 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).
17. Willenborg, T. et al. Genomic-wide association studies identify novel susceptibility loci for migraine without aura. Nat. Genet. 44, 777–782 (2012).
18. Anttila, V. et al. Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat. Genet. 45, 912–917 (2013).
19. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 48, 799–717 (2016).
20. Chen, S.-P. et al. Genome-wide association study identifies novel susceptibility loci for migraine in Han Chinese residing in Taiwan. Cephalalgia 38, 466–475 (2018).
21. Chang, X. et al. Common variants at 5q33.1 predispose to migraine in African-American children. J. Med. Genet. 55, 831–836 (2018).
22. Tfelt-Hansen, P. & Olesen, J. Taking the negative view of current migraine therapies. Neurology 80, 1421–1427 (2013).
23. 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).
24. Harder, M. J. & Russell, M. B. The relative role of genetic and environmental factors in migraine without aura. Neurology 53, 995–999 (1999).
25. Lipton, R. B. et al. Effect of ubrogepant vs placebo on pain and the most bothersome associated symptom in the acute treatment of migraine: the ACHIEVE II randomized clinical trial. JAMA 322, 1887–1898 (2019).
26. Charles, A. & Pozo-Rosich, P. Targeting calcitonin gene-related peptide: a new era in migraine therapy. Lancet 394, 1774–1777 (2019).
27. Goadsby, P. J. et al. Safety, tolerability, and efficacy of orally administered atogepant for the prevention of episodic migraine in adults: a double-blind, randomised phase 2b/3 trial. Lancet Neurol. 19, 727–737 (2020).
28. Cohen, J. et al. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet. 37, 161–165 (2005).
29. Flannick, J. et al. Loss-of-function mutations in SLCO3A8 protect against type 2 diabetes. Nat. Genet. 46, 357–363 (2014).
30. Bulik-Sullivan, B. et al. Atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).
31. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,884 individuals identifies novel genetic loci and pathways. Nat. Genet. 50, 920–927 (2018).
32. Pardinas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389 (2018).
33. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22, 343–352 (2019).
34. Yang, I. et al. Genomic inflation factors under polygenic inheritance. Eur. J. Human Genet. 19, 807–812 (2011).
35. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
36. McLaren, W. et al. The Ensembl variant effect predictor. Genome Biol. 17, 122 (2016).
37. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1232 (2015).
38. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 49, 1421–1427 (2017).
39. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory architecture behind the subtypes. NATURE GENETICS, Vol. 54 | FEBRUARY 2022 | 152-160 | www.nature.com/naturegenetics
International Headache Genetics Consortium

Heidi Hautakangas1, Bendik S. Winsvold2,3,4, Gyda Bjornsdottir5, Aster V. E. Harder6,7, Lisette J. A. Kogelman8, Padhraig Gormley13, Ville Artto14, Anna Bjornsdottir16, Dorret I. Boomsma17, Mona Ameri Chalmer8, Irene de Boer9, Martin Dichgans21,22, Maria G. Hrafnsdottir27, Espen S. Kristoffersen30,32,33, Tobias Kurth34, Terho Lehtimäki35, Sigurdur H. Magnusson5, Rainer Malik21, Caroline Ran39, Gudrun R. Sigurdardottir16, Olafur A. Sveinsson27, Thorgeir E. Thorgeirsson5, Andrea Carmine Belin39, Tobias Freilinger43,44, M. Arfan Ikram24, Marjo-Riitta Järvelin45,46,47,48, Olli T. Raitakari49,50,51, Gisela M. Terwindt6, Mikko Kallela14, Maija Wessman1,52, Jes Olesen8, Daniel I. Chasman19,20, Dale R. Nyholt53, Hreinn Stefánsson5, Kari Stefansson5, Arn M. J. M. van den Maagdenberg6,7, Thomas Folkmann Hansen8,15, John-Anker Zwart2,3,29, Aarno Palotie1,56,57 and Matti Pirinen1,55,58

Lists of members and their affiliations appears in the Supplementary Information

HUNT All-in Headache

Anne Heidi Skogholt3, Ben M. Brumpton3, Bendik S. Winsvold2,3,4, Espen S. Kristoffersen30,32,33, John-Anker Zwart2,3,29, Knut Hagen25,26, Kristian Hveem3,28, Laurent F. Thomas3,9,10,11, Maiken Elvestad Garbrielsen3 and Marianne Bakke Johnsen3,29,30

Lists of members and their affiliations appears in the Supplementary Information

Danish Blood Donor Study Genomic Cohort

Karina Banasik15, Kristoffer Solvsten Burgdorf18, Thomas Folkmann Hansen8,15, Ole Birger Pedersen36, Hreinn Stefánsson5, Kari Stefansson5 and Henrik Ullum18

Lists of members and their affiliations appears in the Supplementary Information
Methods

**Cohorts and phenotyping.** All participating studies were approved by local research ethics committees, and written informed consent was obtained from all study participants. For all the participating studies, an approval was received to use the data for the present work. Study-specific ethics statements are provided in the Supplementary Note.

First, we performed a genome-wide meta-analysis on migraine including five study collections, as listed in Table 1 and Supplementary Table 1. Second, we performed subtype-specific meta-analyses on MA and on MO, both including five study collections listed in Table 1, for the 123 migraine-related risk variants described in the migraine analysis. A description of the study collections is given in the Supplementary Note. In particular, the migraine phenotype has been self-reported in other cohorts except in IHGC2016, where a subset of patients were phenotyped in specialized headache centers, as previously explained5.

**Quality control.** Before the meta-analysis, a standard quality control (QC) protocol was applied to each individual GWAS. Related individuals were removed from all other cohorts except HUNT (which modeled relatedness via a logistic mixed model) by using an identity by descent cut-off of 0.185 or smaller. Multiallelic variants were excluded from all studies, and only variants that satisfied the following thresholds were kept for further analysis: minor allele frequency (MAF) >0.01, IMPUTE2 info or MACH r2 >0.6 and, when available, Hardy-Weinberg equilibrium (HWE) P-value >1×10−4 and missingness <0.05. Variants were matched by chromosome, position and alleles to the UK Biobank data. Indels were recoded as insertions (I) and deletions (D). For each study, SNPs with an effect allele frequency (EAF) discrepancy of >0.30 and indels with EAF discrepancy of >0.20 to UK Biobank were excluded. MAF and EAF plots of cohorts against the reference cohort are shown in Supplementary Data 7. We conducted a sensitivity analysis on strand-ambiguous SNPs (with alleles A/T or G/C), by counting, for each pair of studies, how often the same allele of A/T or G/C SNP was coded as the minor allele in both cohorts, as a function of MAF threshold (Supplementary Table 17). Minor alleles were same at least in 97.39% of the SNPs without MAF threshold and the corresponding proportions were 99.96% and 79.58% when MAF <0.25 and when MAF >0.4, respectively. The very high concordance for SNPs with MAF <0.25 suggests that the strand-ambiguous SNPs were labeled consistently for almost every SNP. Therefore, we did not exclude any SNPs based on possible labeling mismatches due to strand ambiguity.

**Statistical analysis.** All statistical tests conducted were two-sided unless otherwise indicated. The GWAS for the individual study cohorts were performed by logistic regression with an additive model of imputed dosage of the effect allele on the log odds of migraine. The analyses for IHGC2016 (ref. 18) and 23andMe19 have been described before. For UKBB data and GeneRISK data, we used PLINK v2.0 (ref. 16). For HUNT data, we used a logistic mixed model with the saddlepoint approximation as implemented in SAIGE v0.20 (ref. 18) that accounts for the genetic relatedness. All models were adjusted for sex and at least for the four leading principal components of the genetic population structure (Supplementary Table 15). Age was used as a covariate when available. A detailed description is provided in Supplementary Note. For the chromosome X meta-analysis, male genotypes were coded as [0,2] in all cohorts, and the GWAS were conducted with an X chromosome inactivation model that treats hemizygous males as equivalent to homozygous females18.

We performed an inverse–variance weighted fixed-effect meta-analysis on the five study collections by using GWAMA16. After the meta-analysis, we excluded the variants with effective sample size Neff <5,000 to remove results with very low precision compared with most variants and were left with 10,843,197 variants surpassing the QC thresholds. We estimated the effective sample size for variant i as

\[
N_{eff}(i) = \frac{1}{\left(1 - f_{i}\right) s_i^2},
\]

where f is the effect allele frequency for variant i and s is the s.e. for variant i estimated by the GWAS software. This quantity approximates the value 2N (1−t−t), where N is the total sample size (cases + controls), t is the proportion of cases and i is the imputation info (derivation in Supplementary Note).

**Risk loci.** There were 8,117 GWAS variants with the meta-analysis P-value <5×10−8. For 8,067 of them that were available in UK Biobank, an LD matrix was obtained from UK Biobank using a random sample of 10,000 individuals included in the UKBB GWAS. We defined the index variants as the LD-independent GWAS variants at LD threshold of r2 <0.1 in the following way: First, the GWAS variant with the lowest P-value was chosen and, subsequently, all GWAS variants that were in LD with the chosen variant (r2 >0.1) were excluded. Next, out of the remaining GWS variants, the variant with the lowest P-value was chosen and the GWS variants in LD with that variant were excluded. This procedure was repeated until there were no GWS variants left. Out of the 8,067 variants with LD information, 170 were LD-independent (at r2 <0.1). For 18/30 variants that were not found in UK Biobank, LD information was available from the 23andMe data, and all 18 variants were in LD (r2 >0.1) with some index variant. Of the 18 variants, 2 (rs111404218 and rs12149936) had lower P-value than the original index variant they were in LD with and, hence, they replaced the original index variants. For 32 GWS variants, LD information was unknown. Thus, for these 32 variants, the GWS associations were represented by 202 = 168 + 32 index variants.

Next, to define the risk loci and their lead variants, an LD block around each index variant was formed by the interval spanning all GWS variants that were in high LD (r2 >0.6) with the index variant. Sizes of these regions ranged from 1 bp (only the variant itself) for the variants with unknown LD) to 1,089 kb. Sets of regions that were less than 250 kb away from each other were merged (distance from the end of the first region to the beginning of the second region). This definition resulted in 126 loci. All other GWS variants were included in their closest locus based on their position and the locus boundaries were updated and, finally, loci within 250 kb from each other were merged. This resulted in our final list of 37 risk loci. Each risk locus was represented by its lead variant defined as the variant with the lowest P-value and named by the nearest protein-coding gene to the lead variant or by the nearest noncoding gene if there was no protein-coding gene within 250 kb. The term ‘Near’ was added to the locus name if the lead variant did not overlap with a gene transcript. We note that the nearest gene to the lead variant need not be a causal gene. None of the 32 variants without LD information became a lead variant of a risk locus because all had a variant in the vicinity with a smaller P-value.

We annotated and mapped these loci by their physical position to genes by using the Ensembl Variant Effect Predictor (VEP, GRCh37). We used two different thresholds for annotating genes: a distance of 500 kb and 250 kb to the nearest transcript of a gene. The filtered results including all variants within a gene or a regulatory element are presented in Supplementary Table 7b.

**Stepwise conditional analysis.** We performed a stepwise conditional analysis (CA) on each risk locus by using FINEMAP v.1.4 (ref. 20). FINEMAP uses GWAS summary level statistics together with an LD reference panel and a full fine-mapping reference panel to detect causal SNPs from a large number of individual-level data. When the reference LD does not accurately match the GWAS data, full fine-mapping is prone to false positives34. A simpler stepwise CA is more robust to inaccuracy in reference LD because CA has a much smaller search space than full fine-mapping, and therefore CA is less likely to run into most problematic variant combinations where LD is very inaccurate. Since we did not have the full individual-level data from the present work, we carried out only the CA and not the full fine-mapping. For the CA, we included only the SNPs, but no indels, and we used the same reference LD from the UK Biobank data as we used to define the risk loci. We restricted the CA only to the variants with a similar effective sample size (Neff) by using a threshold of ±10% of the Neff of the lead SNP of the risk locus, because our summary statistics came from the meta-analysis where sample sizes per variant vary greatly. This filter excluded approximately 17% of all GWS variants and was necessary since otherwise CA led to spurious conditional P-values, such as P < 10−26 for some loci. Consequently, for two of the loci where the lead variant was an indel, the lead variant was not included in the CA. For such regions, we checked that the new lead variant from the output was in LD (r2 >0.3) with the original lead variant. For one locus (rs111404218) where the lead variant does not have LD information in the UK Biobank data, there were no GWS variants left in the CA after filtering by Neff. We used the standard GWS (P < 5×10−8) threshold to define the secondary variants that were conditionally independent from the lead variant. The CA results are in Supplementary Tables 6a,b.

eQTL mapping to genes and tissues.** We used two data sources to map the risk variants to genes via eQTL associations. From the GTEx v.8 database (https://gtexportal.org), we downloaded the data of 49 tissues. We first mapped all 123 lead variants to all significant cis-eQTLs across tissues using the FDR cut-off of 5% as provided by the GTEx project21. Next, we also mapped the variants in high LD (r2 >0.6) with the lead variants to all significant cis-eQTLs. Finally, we filtered the results to include only the new significant gene–tissue pairs that were not implicated by the lead variants. Results are shown in Supplementary Tables 9 and 10.

With FUMA v.1.3.6 (ref. 25), we mapped the 123 lead variants, and the variants in high LD (r2 >0.6) with the lead variants, to the other eQTL data repositories provided by FUMA except GTEx, that is, Blood eQTL Browser43, BIOS QTL browser44, BRAINEAC45, MuTHER46, xQTLServer47, CommonMind Consortium48, eQTLGen49, eQTL Catalogue50, DICE51, scRNA-eQTLs52 and PsyENCODEx53.

Results are shown in Supplementary Tables 9 and 10.

To study whether the lead variants were enriched in any of the 49 tissues from GTEx v.8, we fitted a linear regression model where the number of lead variants that are significant cis-eQTLs for a specific tissue was used as the outcome, and the overall number of genes with at least one significant cis-eQTL reported by GTEx for the tissue was the predictor16. We did a separate regression model for each tissue (by leaving out the tissue of interest from the model, and we used the model fitted on the other tissues for predicting the outcome variable for the tissue type of interest. Finally, we checked in which tissues the true observed number of migraine lead variants was outside of the 95% prediction intervals as given by the function ‘predict.lm’(interval=‘prediction’) in R software. Details of the procedure are in the Supplementary Note.
LD-score regression. We estimated both the SNP heritability \((h^2_{SNP})\) of migraine and pairwise genetic correlations \((r_c)\) between each pair of study collections using LDSC v.1.0.0 (refs. 30,35). SNP heritability and genetic correlations were estimated using European LD SCNA from the 1000 Genomes Project, HapMap3 SNPs, downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE. We reformatted the meta-analysis association statistics to LDSC format with munge-tool, which excluded variants that did not match with the HapMap3 SNPs, had strand ambiguity (that is, A/T or G/C SNPs), MAF <0.01 or missingness more than two-thirds of the 90th percentile of the total sample size, or resided in long-range LD regions\(^{37,38}\), in centromere regions or in the major histocompatibility locus (MHC) of chromosome 6, leaving 1,165,201 SNPs for the LDSC analyses. We used a migraine population prevalence of 16% and a sample proportion of cases of 11.7% = 102,084/(102,084 + 771,257) to turn the LDSC slope into the estimate of \(h^2_{SNP}\) on the liability scale\(^{38}\). Pairwise genetic correlation residuals were used to identify the significance threshold migraine meta-analysis\(^{37,38}\), LDSC reported \(h^2_{SNP}\) value of 14.6% (13.8–15.5%)%, which was considerably larger than the value 11.2% (10.8–11.6%) that we report in our analysis. When we ran our LDSC pipeline on the data of Gormley et al.\(^ {37}\), we estimated \(h^2_{SNP}\) value of 10.6% (10.1–11.1%). Thus, it seems that our liability transformation estimates lower values of heritability than the transformation used by Gormley et al.\(^ {37}\).

Stratified LD-score regression. We used S-LDSC to partition the SNP heritability by functional genomic annotations\(^{37}\). We used the baseline-LD model\(^ {38}\) that contained MAF- and LD-related annotations, such as recombination rate and predicted allele frequencies, and 75 annotations, including conserved, coding and regulatory regions of the genome and different histone modifications. The baseline-LD model adjusts for MAF- and LD-related annotations, such as recombination rate and predicted allele frequency, which decreases the risk of model misspecification\(^ {37,38}\). We used the same QC as with the univariate LDSC, and the baseline LDv.1.1 European LD scores estimated from the 1000 Genomes Project Phase 3, downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE, from predictdb.org and the European 1000 Genomes v.3 LD reference panel (hg38; https://data.broadinstitute.org/alkesgroup/LDSCORE/LDSCORE_v1_0_0 tổng 102,084/(102,084 + 771,257) = 54.3% leading to 612 clumps formed from 7,672 variants (Supplementary Table 15d–e). We tested whether the SNPs were in LD by calculating the LD R2 statistic for each pair of SNPs. We also divided similar endpoints into broader categories are listed in Supplementary Table 19. The same approach was used for the PheWAS of FinnGen R4. We first downloaded all the available summary statistics (2,263 endpoints) and, next, obtained all the associations for the 123 risk loci with all the high LD variants included using Pvalue thresholds of \(P < 1\times 10^{-4}\) and \(P < 1\times 10^{-8}\) (Supplementary Table 13a-c). We also divided similar endpoints into broader categories, which are listed in Supplementary Table 20.

We tested the direction of allelic effects between migraine and the following three traits that shared multiple associated variants with migraine: CAD\(^ {98}\), diastolic blood pressure\(^ {99}\) and systolic blood pressure\(^ {98}\). We first used all migraine lead variants that were available also in the summary statistics of the other trait without any Pvalue threshold and used a binomial test to test whether the proportion of variants with same direction of effects was 0.5. Next, we used a Pvalue threshold of \(1\times 10^{-3}\) for the association with the other trait. Results are in Supplementary Table 13d.

LD-score regression applied to specifically expressed genes. We used LDSC-SEG\(^ {100}\) to identify tissues and cell types implicated by the migraine GWAS results. LDSC-SEG uses gene expression data and GWAS results from all variants together with an LD reference panel. For our analyses, we used the same QC as for the other LDSC analyses and six different sets of readily constructed annotation-specific LD scores downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/LDSCORE_v1_0_0. We used DEPICT\(^ {52}\) is an integrative tool to identify the most likely causal genes at associated loci, and enriched pathways and tissues or cell types in which the genes from the associated loci are highly expressed. As an input, DEPICT takes a set of trait-associated SNPs. First, DEPICT uses coregulation data from 77,840 microarrays to predict biological functions of genes and to construct 14,461 reconstituted gene sets. Next, information of similar predicted gene functions is used to identify and prioritize gene sets that are enriched for genes in the associated loci. For the tissue- and cell-type–enrichment analysis, DEPICT uses a set of 37,427 human gene expression microarrays. We used DEPICT v.1.194 and ran the analyses twice for each of the Pvalue thresholds for clumping, as recommended\(^ {52}\), and using the default settings of 500 permutations for bias adjustment and 50 replications for the FDR estimation and for the Pvalue calculation. As an input, we used not only the autosomal SNPs and the same UK Biobank LD reference data as for the other analyses. We applied a Bonferroni correction (\(\alpha = 0.05\times 18,985\)) to identify significantly associated genes for migraine with the same list of Supplementary Table 16a. Finally, we used the results from the gene-based analysis to perform a gene-set analysis by using two different gene-set collections from the Molecular Signature Database v.7.0 (refs. 102,103); the curated genes sets containing 5,500 genes and the GO sets containing 9,988 genes. We performed the gene-set analysis using a model and one-sided LDSC-SEG with a value threshold of 5\times 10^{-8} leading to 612 clumps formed from 22,480 variants (Supplementary Table 15a–d).

DEPICT. DEPICT\(^ {15}\) is an integrative tool to identify the most likely causal genes at associated loci, and enriched pathways and tissues or cell types in which the genes from the associated loci are highly expressed. As an input, DEPICT takes a set of trait–associated SNPs. First, DEPICT uses coregulation data from 77,840 microarrays to predict biological functions of genes and to construct 14,461 reconstituted gene sets. Next, information of similar predicted gene functions is used to identify and prioritize gene sets that are enriched for genes in the associated loci. For the tissue- and cell-type–enrichment analysis, DEPICT uses a set of 37,427 human gene expression microarrays. We used DEPICT v.1.194 and ran the analyses twice for each of the Pvalue thresholds for clumping, as recommended\(^ {15}\), and using the default settings of 500 permutations for bias adjustment and 50 replications for the FDR estimation and for the Pvalue calculation. As an input, we used not only the autosomal SNPs and the same UK Biobank LD reference data as for the other analyses. We ran the analysis using a clumping Pvalue threshold of 5\times 10^{-8} leading to 615 clumps formed from 7,672 variants (Supplementary Table 15d–f). Second, we used a Pvalue threshold of 1\times 10^{-3} leading to 612 clumps formed from 22,480 variants (Supplementary Table 15a–c).

Transcriptome-wide association study and colocalization. We performed a transcriptome-wide association study (TWAS) by S-PrediXcan\(^ {16}\) v.0.7.5 using GTEx v.8 multivariate adaptive shrinkage models (MASHR-M) for 49 tissues downloaded from predictdb.org and the European 1000 Genomes v.3 LD reference panel (hg38; https://data.broadinstitute.org/alkesgroup/LDSCORE/LDSCORE_v1_0_0).
Fine-mapping of causal gene sets. To prioritize genes for the migraine loci, we applied a gene-based fine-mapping approach using fine-mapping of causal gene sets (FOCUS) v0.7 (ref. 4). FOCUS is a Bayesian approach that models predicted expression correlations among TWAS signals to estimate posterior probabilities for all genes within a tested region.

We used the European 1000 Genomes v3 LD reference panel and same GTEx v8 predicted expression weights for the 49 tissues as with S-PrediXcan. First, we mapped the migraine summary statistics from hg37 to hg38 with UCSC liftOver\(^4\). Next, we followed the suggested QC protocol and applied the modified munge-tool to obtain cleaned summary statistics. After the QC steps, we had 6,237,177 variants left for the analysis. We performed tissue-prioritized fine-mapping of gene-sets for the 49 tissues with otherwise default settings except that we increased the P value threshold to \(1 \times 10^{-6}\) so that the fine-mapping would cover most of the same regions that contained at least one significant gene-tissue pair by S-PrediXcan. Posterior inclusion probability (PIP) from FOCUS is reported for all available significant S-PrediXcan gene-tissue pairs in Supplementary Table 1b, and all prioritized genes by FOCUS with PIP > 0.9 are reported in Supplementary Table 1a.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Results for 8,117 genome-wide significant SNP associations (\(P < 5 \times 10^{-8}\)) from the meta-analysis including 23andMe data are available on the International Headache Genetics Consortium website (http://www.headache-genetics.org/content/datasets-and-cohorts/). Genome-wide summary statistics for the other study collections except 23andMe are available for bona fide researchers (contact Dale Nyholt, d.nyholt@qut.edu.au) within 2 weeks from the request. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/collaborate/#publication for more information and to apply to access the data.

Code availability

R code for the subtype specificity analysis: https://github.com/mjpirinen/migraine-meta.

References

81. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, s1374–s13784 (2015).
82. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample randomness in large-scale genetic association studies. Nat. Genet. 50, 1335–1341 (2018).
83. Tukiaen, T. et al. Chromosome X-wide association study identifies loci for fasting insulin and height and evidence for incomplete dosage compensation. PLoS Genet. 10, e1004127 (2014).
84. Mági, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. PLoS Biol. 11, 288 (2013).
85. Benner, C. et al. FinemAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32, 1493–1501 (2016).
86. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1288–1243 (2013).
87. Zinmanakova, D. V. et al. Identification of context-dependent expression-quantitative trait loci in whole blood. Nat. Genet. 49, 139–145 (2017).
88. Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. 17, 1418–1428 (2014).
89. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012).
90. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain’s transcriptional and epigenome. Nat. Neurosci. 20, 1418–1426 (2017).
91. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1453 (2016).
92. Võsa, U. et al. Large-scale cis and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. 53, 1296–1301 (2021).
93. Kerimov, N. et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat. Genet. 50, 493–497 (2018).
94. Wäng, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaaf4644 (2018).
95. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. Am. J. Human Genet. 83, 132–139 (2008).
96. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. Am. J. Human Genet. 88, 294–305 (2011).
97. Huttunen, M. L. A., Gaziano, M. A., Communaldi, F., van de Geijn, B. & Price, A. L. Disease heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function across species. Am. J. Human Genet. 104, 611–624 (2019).
98. Trochet, H. et al. Bayesian meta-analysis across genome-wide association studies of diverse phenotypes. Genet. Epidemiol. 43, 532–547 (2019).
99. Chichoukova, A. et al. metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. Bioinformatics 32, 1981–1989 (2016).
100. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740 (2011).
101. Schrammanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).
102. Hinrichs, A. S. et al. The UCSC genome browser database: update 2006. Nucleic Acids Res. 34, D590–D598 (2006).

Acknowledgements

We thank the study participants for their contribution to this research. We also thank the numerous individuals who contributed to sample collection, storage, handling, phenotyping and genotyping for each of the individual cohorts. We acknowledge the participants and investigators of the Finngen study. This research has been conducted using the UK Biobank Resource under Application Number 22627. We are supported by following grants: the US National Institute of Neurological Disorders and Stroke (NINDS) of the US National Institutes of Health (NINDS) (grant numbers R21NS09296 and R21NS104398 (D.I.C.)), the Finnish innovation fund Sitra and Finska Lakaresällskapet (E.W.), the Academy of Finland (grant nos. 288509, 310276, 336825 (M.P.)), the Sigrid Juselius Foundation (M.P. and S.R.), the Academy of Finland Center of Excellence in Complex Disease Genetics (grant no. 310262 (S.R.)), the Finnish Foundation for Cardiovascular Research (S.R.), University of Helsinki HILDE Fellow and Grand Challenge grants (S.R.), The Novo Nordisk Foundation (NNF14CC0001 and NNF17CC0027594 (TEH. and K.B.)), CANDY foundation (CEHEAD (TEH.) and the South-Eastern Norway Regional Health Authority (grant no. 2020034 (B.S.W.))). A list of study-specific acknowledgements and funding information can be found in the Supplementary Note.

Author contributions

H.S., A.M.J.M.v.d.M., TEH. and M.P. conceived the study. H.H., B.S.W., S.E.R., G.B., A.V.E.H., I.J.A.K., I.F.T., R.N. and L.S.V. performed analyses in their respective cohorts. H.H., B.S.W., D.I.C., D.R.N., A.M.J.M.v.d.M., T.F.H., J.-A.Z., J.L.M., A.V.E.H., L.J.A.K., L.F.T., R.N. and L.S.V. performed analyses in their respective cohorts. H.H., B.S.W., D.I.C., D.R.N., A.M.J.M.v.d.M., T.F.H., J.-A.Z., A.P. and M.P. contributed to writing the manuscript. H.H. performed meta- analysis, and created figures and tables. H.H. and M.P. performed downstream analyses and drafted the manuscript. M.P. supervised project. All authors interpreted the results and reviewed and commented on the manuscript.

Competing interests

G.B., T.E.T., S.H.M. and H.S. and K.S. are employees of deCODE genetics/Amgen. PG. is a current employee and stockholder of GlaxoSmithKline but work was conducted while employed by Massachusetts General Hospital, Boston, MA, USA. TF reports possible competing interest for Electrocore (participation in clinical studies), Novartis (speaker’s honoraria, participation in advisory boards, participation in clinical studies), Teva (speaker’s honoraria, participation in advisory boards), Lilly (speaker’s honoraria, participation in clinical studies), and Bayer
M. Kallela has served on Advisory Boards for MSD and Allergan; has received funding for travel and/or speaker honoraria from MSD, Allergan, Teva, Novartis and Genzyme; has received compensation for producing educational material from Teva and Allergan; has received research support from Helsinki University Central Hospital; and holds stock/stock options and/or has received Board of Directors compensation from Helsinki Headache Center. T.K. reports having received honoraria from Eli Lilly, Newsenselab and Total for providing methodological advice and from the BMJ for editorial services. A.P. is the Scientific Director of the public–private partnership project FinnGen that has 12 industry partners that provide funding for the FinnGen project. V.A. has served on advisory board for Allergan, Lundbeck, Teva and Lilly. G.M.T. reports consultancy support from Novartis, Allergan, Lilly and Teva, and independent support from Dutch Organization for Scientific Research, the Dutch Heart & Brain Foundations, IRRF and Dioraphte. Other authors report no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00990-0.
Correspondence and requests for materials should be addressed to Matti Pirinen.
Peer review information Nature Genetics thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at www.nature.com/reprints.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of any covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: No software was used for data collection.

Data analysis: Individual GWAS were conducted using either PLINK v2.0 or SAIGE v0.20. Meta-analyses were conducted by GWAMA v2.1 and v.2.2.0. Conditional analysis was conducted by FINEMAP v.1.4. Variants were mapped to gene transcripts by Ensembl Variant Effect Predictor (VEP version 99, GRCh37), and to genes and tissues by FUMA v1.3.6. LDSC v1.0.0 was used to estimate and partition heritability, to estimate genetic correlations, and for LDSC-SE analyses. MAGMA v1.09 was used to identify genes and gene sets. DIPECt v1.194 was used for tissue and gene set enrichment analyses. R (versions 3.4.0, 3.5.0 and 3.5.1) were used for PhenoWAS, for tissue enrichment analysis, for gene expression analyses and for subtype-specificity analyses (https://github.com/mpirinen/migraine-meta). FOCUS v0.7 was used for finding mapping causal gene sets. S-PredXcan (from MetaXcan v0.7.5) was used for transcriptome-wide association study. COLOC v4.0.4 was used for colocalization analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Results for 8,117 genome-wide significant SNP associations (\(P < 5 \times 10^{-8}\)) from the meta-analysis including 23andMe data are available on the International Headache Genetics Consortium website (http://www.headache-genetics.org/content/datasets-and-cohorts). Genome-wide summary statistics for the other study collections except 23andMe are available for bona fide researchers (contact Dr. Dale Nyholt, d.nyholt@qut.edu.au) within two weeks from the request. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit research.23andme.com/collaborate/#publication for more information and to apply to access the data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☑ Life sciences
- ☐ Behavioural & social sciences
- ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
- Our total sample size is 102,084 migraine cases and 771,257 controls, and sample sizes for subtypes are 14,624 MA cases and 703,852 controls, and 15,055 MO cases and 682,301 controls. These sample sizes resulted when we included all available samples to maximize statistical power for GWAS discovery. Exact sample size was not predetermined by any other criterion except the availability of samples.

Data exclusions
- We followed standard quality control procedures of GWAS to exclude individuals and genetic variants. Further details are described in the Methods section and in Supplementary Note.

Replication
- High genetic correlations showed that the genetic architecture of migraine phenotype in different study collections was highly similar. We assessed the consistency across the study cohorts in subtype-specific analyses by sign tests.
- We do not report a separate replication because we included all available data in the analyses to maximize the statistical power.

Randomization
- Our study is a case-control study, and randomization is not applicable.

Blinding
- Our study is a case-control study, and blinding is not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a
- ☑ Antibodies
- ☑ Eukaryotic cell lines
- ☑ Palaeontology and archaeology
- ☑ Animals and other organisms
- □ Human research participants
- ☑ Clinical data
- □ Dual use research of concern

Methods

- n/a
- □ ChiP-seq
- □ Flow cytometry
- □ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics
- All study participants are adult females or males of European descent. The migraine sample prevalence was 11.7% for the
| Population characteristics | main meta-analysis (14.4% [HGC2016], 18.7% [23andMe], 3.2% [UK Biobank], 18.2% [GeneRISK], 19.4% [HUNT]); MA sample prevalence was 2.0% and MO sample prevalence was 2.2%. Age distribution varied between studies. More detailed description of each study collections are provided in Supplementary Note, and for the UK Biobank in (https://www.ukbiobank.ac.uk). |
|-----------------------------|------------------------------------------------------------------------------------------------------------|
| Recruitment                 | The migraine GWAS meta-analysis consists of 5 study collections, and the subtype analyses included also 3 other study collections. Participants were recruited through population-based cohort studies, case-control studies, a biobank, a direct-to-consumer study, and through hospitals and clinics. For a majority of the cases, migraine phenotype was self-reported, but a subset of the patients were phenotyped in specialized headache centers. Further details of each study’s recruitment are provided in the Supplementary Note. We note that a large proportion of migraine diagnoses is self-reported. Therefore, it is possible that there are some cases among the controls. The consequence of this is that the observed differences in frequencies of migraine risk alleles between cases and controls are smaller, and we would have less statistical power, compared to more accurate control definition. However, in this scenario, the bias would be towards zero at the migraine risk variants, but null variants would not be biased. Further, we cannot rule out misdiagnosis, such as, e.g., tension headache being reported as migraine. The consequence of this would be that some of the risk loci could overemphasize genetic factors related to some other migraine-associated traits such as general pain mechanisms rather than genetic factors of migraine itself. However, the high genetic correlation that we observed supports a strong phenotypic concordance between the study collections that include also deeply phenotyped clinical cohorts from headache specialist centers. |
| Ethics oversight            | All participating studies were approved by local research ethics committees and written informed consent was obtained from all study participants. Further details are described in Supplementary Note. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.