Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles

4

Heidi Hautakangas¹, Bendik S. Winsvold^{2,3,4}, Sanni E. Ruotsalainen¹, Gyda Bjornsdottir⁵, Aster V. E.
Harder^{6,7}, Lisette J. A. Kogelman⁸, Laurent F. Thomas^{3,9,10,11}, Raymond Noordam¹², Christian Benner¹,
Padhraig Gormley¹³, Ville Artto¹⁴, Karina Banasik¹⁵, Anna Bjornsdottir¹⁶, Dorret I. Boomsma¹⁷, Ben
M. Brumpton³, Kristoffer Sølvesten Burgdorf¹⁸, Julie E. Buring^{19,20}, Mona Ameri Chalmer⁸, Irene de
Boer⁶, Martin Dichgans^{21,22}, Christian Erikstrup²³, Markus Färkkilä¹⁴, Maiken Elvestad Garbrielsen³,
Mohsen Ghanbari²⁴, Knut Hagen^{25,26}, Paavo Häppölä¹, Jouke-Jan Hottenga¹⁷, Maria G. Hrafnisdottir²⁷,
Kristian Hveem^{3,28}, Marianne Bakke Johnsen^{3,29,30}, Mika Kähönen³¹, Espen S. Kristoffersen^{30,32,33},
Tobias Kurth³⁴, Terho Lehtimäki³⁵, Lannie Lighart¹⁷, Sigurdur H. Magnusson⁵, Rainer Malik²¹, Ole
Birger Pedersen³⁶, Nadine Pelzer⁶, Brenda W. J. H. Penninx^{37,38}, Caroline Ran³⁹, Paul M. Ridker^{19,20},
Frits R. Rosendaal⁴⁰, Gudrun R. Sigurdardottir¹⁶, Anne Heidi Skogholt³, Olafur A. Sveinsson²⁷,
Thorgeir E. Thorgeirsson⁵, Henrik Ullum¹⁸, Lisanne S. Vijfhuizen⁷, Elisabeth Widén¹, Ko Willems van
Dijk^{7,41}, International Headache Genetics Consortium⁴², HUNT All-in Headache⁴², Danish Blood
Donor Study Genomic Cohort⁴², Arpo Aromaa⁴³, Andrea Carmine Belin³⁹, Tobias Freilinger^{44,45}, M.
Arfan Ikram²⁴, Marjo-Riitta Järvelin^{46,47,48,49}, Olli T. Raitakari^{50,51,52}, Gisela M. Terwindt⁶, Mikko

19 Kallela¹⁴, Maija Wessman^{1,53}, Jes Olesen⁵⁴, Daniel I. Chasman^{19,20}, Dale R. Nyholt⁵⁵, Hreinn
20 Stefánsson⁵, Kari Stefansson⁵, Arn M. J. M. van den Maagdenberg^{6,7}, Thomas Folkmann Hansen^{8,15},
21 Samuli Ripatti^{1,56,57}, John-Anker Zwart^{2,3,29}, Aarno Palotie^{1,58,59}, Matti Pirinen^{1,57,60*}

22 1. Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of
23 Helsinki, Helsinki, Finland 2. Department of Research, Innovation and Education, Division of Clinical
24 Neuroscience, Oslo University Hospital, Oslo, Norway 3. K. G. Jebsen Center for Genetic Epidemiology,
25 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of
26 Science and Technology, Trondheim, Norway 4. Department of Neurology, Oslo University Hospital, Oslo,
27 Norway 5. deCODE genetics/Amgen Inc., Reykjavik, Iceland 6. Department of Neurology, Leiden University
28 Medical Center, Leiden, the Netherlands 7. Department of Human Genetics, Leiden University Medical Center,
29 Leiden, the Netherlands 8. Danish Headache Center, Department of Neurology, Copenhagen University
30 Hospital, Copenhagen, Denmark 9. Department of Clinical and Molecular Medicine, Norwegian University of
31 Science and Technology, Trondheim, Norway 10. BioCore - Bioinformatics Core Facility, Norwegian
32 University of Science and Technology, Trondheim, Norway 11. Clinic of Laboratory Medicine, St.Olavs
33 Hospital, Trondheim University Hospital, Trondheim, Norway 12. Department of Internal Medicine, Section of
34 Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands 13. Genetics and
35 Pharmacogenomics, Merck & Co., Inc., Boston, MA, USA 14. Department of Neurology, Helsinki University
36 Central Hospital, Helsinki, Finland 15. Novo Nordic Foundation Center for Protein Research, Copenhagen
37 University, Copenhagen, Denmark 16. Neurology private practice, Laeknasetrid, Reykjavik, Iceland 17.
38 Netherlands Twin Register, Dept Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands 18.
39 Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
40 19. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA 20. Harvard Medical
41 School, Boston, MA, USA 21. Institute for Stroke and Dementia Research, University Hospital, LMU Munich,
42 Munich, Germany 22. Munich Cluster for Systems Neurology (Synergy), Munich, Germany 23. Department of
43 Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark 24. Department of Epidemiology, Erasmus

44 University Medical Center, Rotterdam, the Netherlands 25. Department of Neuromedicine and Movement
 45 Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU),
 46 Trondheim, Norway 26. Clinical Research Unit Central Norway, St. Olavs University Hospital, Trondheim,
 47 Norway 27. Landspítali University Hospital, Reykjavik, Iceland 28. HUNT Research Center, Department of
 48 Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and
 49 Technology, Trondheim, Norway 29. Institute of Clinical Medicine, Faculty of Medicine, University of Oslo,
 50 Oslo, Norway 30. Research and Communication Unit for Musculoskeletal Health (FORMI), Department of
 51 Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
 52 31. Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research
 53 Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland 32.
 54 Department of General Practice, Institute of Health and Society, University of Oslo, Oslo, Norway 33.
 55 Department of Neurology, Akershus University Hospital, Lørenskog, Norway 34. Institute of Public Health,
 56 Charité – Universitätsmedizin Berlin, Berlin, Germany 35. Department of Clinical Chemistry, Fimlab
 57 Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health
 58 Technology, Tampere University, Tampere, Finland 36. Department of Clinical Immunology, Zealand
 59 University Hospital, Køge, Denmark 37. Department of Psychiatry, Amsterdam UMC, Vrije Universiteit,
 60 Amsterdam Public Health research institute, Amsterdam, The Netherlands 38. GGZ inGeest Specialized Mental
 61 Health Care, Amsterdam, The Netherlands 39. Department of neuroscience, Karolinska Institutet, Stockholm,
 62 Sweden 40. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
 63 41. Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the
 64 Netherlands 42. A full list of members and affiliations appears in the Supplementary Note. 43. National Public
 65 Health Institute (present THL), Helsinki, Finland 44. Klinikum Passau, Department of Neurology, Passau,
 66 Germany 45. Centre of Neurology, Hertie Institute for Clinical Brain Research, Tuebingen, Germany 46.
 67 Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public
 68 Health, Imperial College London, London, United Kingdom 47. Center for Life Course Health Research, Faculty
 69 of Medicine, University of Oulu, Oulu, Finland 48. Unit of Primary Health Care, Oulu University Hospital,

70 OYS, Oulu, Finland 49. Department of Life Sciences, College of Health and Life Sciences, Brunel University
 71 London, London, United Kingdom 50. Centre for Population Health Research, University of Turku and Turku
 72 University Hospital, Turku, Finland 51. Research Centre of Applied and Preventive Cardiovascular Medicine,
 73 University of Turku, Turku, Finland 52. Department of Clinical Physiology and Nuclear Medicine, Turku
 74 University Hospital, Turku, Finland 53. Folkhälsan Research Center, Helsinki, Finland 54. Department of
 75 neurology, Danish Headache Center, Copenhagen University Hospital, Copenhagen, Denmark 55. School of
 76 Biomedical Sciences and Centre for Genomics and Personalised Health, Faculty of Health, Queensland
 77 University of Technology, Brisbane, QLD, Australia 56. Broad Institute of MIT and Harvard, Cambridge, MA,
 78 USA 57. Department of Public Health, University of Helsinki, Helsinki, Finland 58. Analytic and Translational
 79 Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts
 80 General Hospital, Boston, MA, USA 59. The Stanley Center for Psychiatric Research and Program in Medical
 81 and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA 60. Department of
 82 Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

83 * Correspondence to: matti.pirinen@helsinki.fi

84

85

Abstract

Migraine affects over a billion individuals worldwide but its genetic underpinning remains largely unknown. This genome-wide association study (GWAS) of 102,084 migraine cases and 771,257 controls identified 123 loci of which 86 are novel. The loci provide an opportunity to evaluate shared and distinct genetic components in the two main migraine subtypes: migraine with aura and migraine without aura. A stratification of the risk loci using 29,679 cases with subtype information, of which approximately half have never been used in a GWAS before, indicated three risk variants that appear specific for migraine with aura (in *HMOX2*, *CACNA1A* and *MPPED2*), two that appear specific for migraine without aura (near *SPINK2* and near *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 (*CALCA/CALCB*) and serotonin 1F receptor (*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.

Introduction

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 (Headache Classification Committee of the International Headache Society (IHS), 2018). Migraine has a lifetime prevalence of 15-20% and is ranked as the second most disabling condition in terms of years

lived with disability (Vos et al., 2020), (Steiner et al., 2020). Migraine is three times more prevalent in females than in males. For about one-third of patients, migraine attacks often include an aura phase (Russell et al., 1995) characterized by transient neurological symptoms such as scintillations. Hence, the two main migraine subtypes are defined as migraine with aura (MA) and migraine without aura (MO).

It has been debated for decades whether or not the migraine subtypes are in fact two separate disorders (Russell and Olesen, 1995), (Kallela et al., 2001), (de Boer et al., 2019) and if so, what the underlying causes are. Prevailing theories about migraine pathophysiology emphasize neuronal and/or vascular dysfunction (Tfelt-Hansen and Koehler, 2011), (Anttila et al., 2018). Current knowledge on disease mechanisms largely comes from studies of a rare monogenic sub-form of MA, familial hemiplegic migraine, for which three ion transporter genes (*CACNA1A*, *ATP1A2* and *SCN1A*) have been identified (Ferrari et al., 2015). The common forms of migraine, MA and MO, instead have a complex polygenic architecture with an increased familial relative risk (Russell and Olesen, 1995), increased concordance in monozygotic twins (Ulrich et al., 1999) and a heritability of 40-60% (Gervil et al., 1999). The largest GWAS thus far with 59,674 cases and 316,078 controls reported 38 genomic loci that confer migraine risk (Gormley et al., 2016). Subsequent analyses of these GWAS data (Finucane et al., 2018) showed enrichment of migraine signals near activating histone marks specific to cardiovascular and central nervous system tissues, as well as for genes expressed in vascular and smooth muscle tissues (Gormley et al., 2016). Other smaller GWAS (Anttila et al., 2010), (Chasman et al., 2011), (Freilinger et al., 2012), (Anttila et al., 2013), (Pickrell et al., 2016), (Chen et al., 2018), (Chang et al., 2018) have suggested 10 additional loci. Of note, the previous datasets were too small to perform a meaningful comparison of the genetic background between migraine subtypes.

128 As migraine is globally the second largest contributor to years lived with disability (Vos et al., 2020),
 129 (Steiner et al., 2020) there clearly is a large need for new treatments. Triptans, i.e., serotonin 5-HT_{1B/1D}
 130 receptor agonists, are migraine-specific acute treatments for the headache phase but are not effective in
 131 every patient, whereas preventive medication is far from satisfactory all together (Tfelt-Hansen and
 132 Olesen, 2012). Recent promising alternatives for acute treatment are serotonin 5-HT_{1F} receptor
 133 agonists ('ditans') (Kuca et al., 2018) and small-molecule calcitonin-gene related peptide (CGRP)
 134 receptor antagonists ('gepants') (Lipton et al., 2019), (Dodick, 2018). For preventive treatment,
 135 monoclonal antibodies (mAbs) targeting CGRP or its receptor have recently proven effective (Charles
 136 and Pozo-Rosich, 2019) and new gepants are under development for migraine prevention (Goadsby et
 137 al., 2020). Still, there remains an urgent need for treatment options for patients who do not respond to
 138 the existing treatments. Genetics has proven a promising way to develop novel therapeutic hypotheses
 139 in other prevalent complex diseases, such as cardiovascular disease (Cohen et al., 2005) and type 2
 140 diabetes (Flannick et al., 2014), and we anticipate that large genetic studies of migraine could also yield
 141 to similar insights.

142 We conducted a GWAS meta-analysis of migraine by adding to the previous meta-analysis of Gormley
 143 et al. (2016) 42,410 new migraine cases from four study collections (Table 1). This increased the
 144 number of migraine cases by 71% for a total sample of 102,084 cases and 771,257 controls.
 145 Furthermore, we assessed the subtype specificity of the risk loci in 8,292 new MA and 6,707 new MO
 146 cases in addition to the 6,332 MA and 8,348 MO cases used previously (Gormley et al. 2016) (Table
 147 2). Here we report 123 genomic loci, of which 86 are novel, and include the first four loci that reach
 148 genome-wide significance ($P < 5 \times 10^{-8}$) in MA. Our subtype data compellingly show that migraine
 149 risk is conferred both by risk loci that appear specific for only one subtype as well as by loci that are
 150 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-HT_{1F} receptor. Finally, our data support the concept that migraine is brought about by both neuronal and vascular genetic factors, strengthening the view that migraine truly is 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 5 study collections (Table 1 and Supplementary Table 1) and analyzed 10,843,197 common variants (Methods). In spite of different approaches to the ascertainment of migraine cases across the studies, the pairwise genetic correlations were all near 1 (Supplementary Table 2), as determined by LD Score (LDSC) regression (Bulik-Sullivan et al., 2015a), 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 (λ_{GC}) of the fixed-effect meta-analysis results was 1.33 (Supplementary Fig. 1), which is in line with other large meta-analyses (Nagel et al., 2018), (Pardiñas et al., 2018), (Howard et al., 2019) and is as expected for a polygenic trait (Yang et al., 2011). The univariate LDSC (Bulik-Sullivan et al., 2015b) 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%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, Supplementary Fig. 3 and 4). Of the 123 loci, 86 are novel whereas 36 overlap with the previously reported 47 autosomal risk loci (Supplementary Table 4) and one with the previously reported X chromosome risk locus. When we represented each risk locus by its lead variant, i.e., the variant with the smallest GWAS P -value, 47 GWS variants were LD-independent ($r^2 < 0.1$) of the 123 lead variants, and with a more stringent threshold ($r^2 < 0.01$), 15 GWS variants remained LD independent of the 123 lead variants (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, which excluded approximately 17% of all GWS variants. The conditional analysis returned 6 SNPs within the 123 risk loci that remained GWS after conditioning on the lead variants (Supplementary Tables 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) (McLaren et al., 2016). In total, transcripts of 390 protein coding and 329 non-protein coding genes were identified within the genomic region of 20 kb around each locus. The number of

transcripts increased to 1,036 protein coding and 1,236 non-protein coding genes when the genomic region was expanded to 250 kb around each locus (Supplementary Table 3). Of the lead variants, 59% (72 of 123) were within a transcript of a protein coding gene, and 80% (99 of 123) of the loci contained at least one protein coding gene within 20 kb; and 93% (114 of 123) within 250 kb. Five of the 123 lead variants were missense variants (in genes *PLCE1*, *MRGPRE*, *SERPINA1*, *ZBTB4* and *ZNF462*), 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 (rs34358) with lead variant rs42854 ($r^2 = 0.85$) in gene *ANKDD1B*, (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 *MRVII* (Supplementary Table 7b).

We further mapped the 123 lead variants to genes via expression quantitative trait locus (eQTL) association using the GTEx V8 data (Aguet et al., 2019) and data repositories included in FUMA (Watanabe et al., 2017) at a false discovery rate (FDR) of 5% (Methods). The lead variants were *cis*-eQTLs for 589 genes (Supplementary Table 8). Variants in high LD with the lead variants were *cis*-eQTLs for an additional 624 genes (Supplementary Table 9). In total, 84% (103/123) of lead variants were *cis*-eQTLs for at least one gene, and 67% (84/123) were *cis*-eQTLs in the GTEx data. All 49 GTEx tissues had at least one lead variant as a significant *cis*-eQTL. Tibial artery had the highest number (47/123) of lead variants as *cis*-eQTLs, and it was the only GTEx tissue type where the enrichment was statistically higher ($P = 6.37 \times 10^{-6}$) than expected based on the overall number of *cis*-eQTLs per tissue reported by GTEx (Supplementary Fig. 5, Supplementary Note 1).

213 We used the stratified LDSC (S-LDSC) to partition migraine heritability by 24 functional genomic
214 annotations (Finucane et al., 2015). Enrichment was observed for 14 functional categories at an FDR of
215 5% (Supplementary Table 10a), of which 11 remained statistically significant after Bonferroni
216 correction for the 24 hypotheses tested (Supplementary Fig. 6). Conserved regions showed the highest
217 enrichment, namely 15.43x (s.e. 1.66; $P = 3.32 \times 10^{-16}$), followed by coding regions at 9.97x (s.e. 2.53;
218 $P = 6.74 \times 10^{-4}$), and transcription start sites (TSS) at 6.70x (s.e. 1.83; $P = 1.18 \times 10^{-3}$) (Supplementary
219 Table 10b).

220 Notably, two of the new risk loci contain genes (*CALCA/CALCB* and *HTR1F*) whose protein products
221 are closely related to targets of two migraine-specific drug therapies (Do et al., 2019). We observe a
222 convincing association at the chromosome 11 locus that contains the *CALCA* and *CALCB* genes
223 encoding CGRP itself (lead SNP rs1003194, $P = 2.43 \times 10^{-10}$; Fig. 2a), while none of the genes
224 encoding CGRP receptor proteins (*CALCRL*, *RAMP1* or *RCP*) show a statistically comparable
225 association (all $P > 10^{-4}$; Supplementary Fig. 7). Variant rs1003194 is a *cis*-eQTL for *CALCB*, but also
226 for *COPB1*, *PDE3B* and *INSC* (Supplementary Table 8). In addition, a new locus on chromosome 3
227 contains *HTR1F* (lead SNP rs6795209, $P = 1.23 \times 10^{-8}$; Fig. 2b), which encodes the serotonin 5-HT_{1F}
228 receptor. Variant rs6795209 is a significant *cis*-eQTL for *HTR1F*, as well as for three other genes
229 (*CGGBP1*, *ZNF654*, *C3orf38*) in the same locus (Supplementary Table 8).

230 Migraine subtypes with aura and without aura

231 Previously, Gormley et al. (2016) conducted subtype-specific GWAS with 6,332 MA cases against
232 144,883 controls and 8,348 MO cases against 139,622 controls, and reported that 7 loci were GWS in
233 MO but none was GWS in MA. Here, we added to the previous data 8,292 new MA and 6,707 new
234 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 a 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 Tables 3b,c, Supplementary Fig. 8 and 9) 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 (i.e., associated only with MO or with MA but not with both), shared by both subtypes, or not associated with either subtype (Methods; Supplementary Table 11a, Supplementary Fig. 10). With a probability above 95%, three lead variants, i.e., rs12598836 in the *HMOX2* locus, rs10405121 in the *CACNA1A* locus and rs11031122 in the *MPPED2* locus, are MA-specific, while two lead variants, i.e., rs7684253 in the locus near *SPINK2*, 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, also four other lead variants showed differences in effect size between the subtypes ($P < 0.05/123$) (Fig. 3b).

PheWAS with NHGRI 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 NHGRI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and for the GWAS summary statistics of 2,263 disease traits in the FinnGen release 4 data. We identified 25 lead variants that were reported to be associated with 23 different phenotype categories (Methods) in the 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 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 12a and Supplementary Fig. 11).

These findings are consistent with previous results that migraine is a risk factor for multiple cardiovascular traits, (Kurth et al., 2016), (Hippisley-Cox et al., 2017), (Adelborg et al., 2018), and genetically correlated with blood pressure (Siewert et al., 2020), (Guo et al., 2020). 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 12d), using the latest meta-analysis of CARDIoGRAMplusC4D Consortium (Nelson et al., 2017), ($N = 336,924$), and blood pressure GWAS from UK Biobank (Loh et al., 2018), ($N = 422,771$).

Enrichment in tissue or cell types and gene sets

We used LDSC applied to specifically expressed genes (LDSC-SEG) (Finucane et al., 2018) (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, i.e., aorta artery ($P = 1.78 \times 10^{-4}$), tibial artery ($P = 3.60 \times 10^{-4}$) and coronary artery ($P = 4.29 \times 10^{-4}$) (Table 3 and Supplementary Table 13a), all of which have previously been reported enriched in migraine without aura (Finucane et al., 2018). 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 ($P = 6.02 \times 10^{-4}$; Table 3 and Supplementary Table 13b). With chromatin-

based annotations, we found enrichment at FDR 5% in five central nervous system (CNS) cell types, in three cardiovascular cell types, one cell type of the digestive system, one musculoskeletal/connective cell type, and in ovary tissue (Table 3 and Supplementary Table 13c). In addition to replicating previous findings (Gormley et al., 2016), (Finucane et al., 2018), the signal linking to ovary tissue has not been reported before.

Next, we used DEPICT (Pers et al., 2015) to identify tissues whose eQTLs were enriched for migraine-associated variants. The tissue enrichment analysis replicated three previously reported tissues (Gormley et al., 2016): arteries (nominal $P = 1.03 \times 10^{-3}$), stomach (nominal $P = 1.04 \times 10^{-3}$) and upper gastrointestinal tract (nominal $P = 1.29 \times 10^{-3}$) (Supplementary Table 14a). Results of gene set analyses using DEPICT (Pers et al., 2015) and MAGMA (de Leeuw et al., 2015) are presented in Supplementary Tables 14 and 15.

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 novel since the previous migraine meta-analysis that yielded 38 loci (Gormley et al., 2016). 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 diseases (Need and Goldstein, 2014), and as expected for a highly polygenic disorder like migraine (Gormley et al., 2018).

298 Migraine subtypes MO and MA were defined as separate disease entities some 30 years ago, and since
 299 then, there has been a continuous debate to what extent they are biologically similar. Over the years
 300 arguments in favour (Kallela et al., 2001) and against (Russell and Olesen, 1995) have been presented
 301 but convincing genetic evidence to support subtype-specific risk alleles has been lacking in previous
 302 genetic studies with smaller sample sizes (Anttila et al., 2013), (Chasman et al., 2014), (Nyholt et al.,
 303 2014). Here we increased considerably the evidence for subtype specificity of some risk alleles by
 304 including new migraine subtype data at the 123 migraine risk variants. We observed that, with a
 305 probability of > 95%, three lead variants (in *HMOX2*, in *CACNA1A* and in *MPPED2*) are associated
 306 with MA but not MO. Of them, *CACNA1A* is a well-known gene linked to familial hemiplegic
 307 migraine, a rare subform of MA (Ophoff et al., 1996), (de Vries et al., 2009). The observation that
 308 *CACNA1A* seems involved in both monogenic and polygenic forms of migraine provides first gene-
 309 based support for the increased sharing of common variants between the two disorders (Gormley et al.,
 310 2018). We do not find evidence that any of the seven loci, that were earlier reported GWS in MO but
 311 not in MA by Gormley et al. (2016), would be specific for MO, while four of them (*LRP1*, *FHL5*, near
 312 *FGF6* and near *TRPM8*) are among the nine loci shared by both subtypes with a probability over 95%.
 313 Loci (e.g., *LRP1* and *FHL5*) that are strongly associated with both subtypes provide convincing
 314 evidence for a previous hypothesis that the subtypes partly share a genetic background (Gormley et al.,
 315 2016), (Zhao et al., 2016). In accordance with our analysis, effects in both subtypes were suggested
 316 before at the *TRPM8* and *TSPAN2* loci while, in contrast to our results, the *LRP1* locus was earlier
 317 reported specific for MO (Chasman et al., 2014). Finally, we also detected four lead variants (including
 318 *LRP1*) that do not appear specific to MO but do confer a higher risk for MO than for MA.

319 It has been long debated whether migraine has a vascular or a neuronal origin, or whether it is a
 320 combination of both (Tfelt-Hansen and Koehler, 2011), (Jacobs and Dussor, 2016), (Anttila et al.,

2018), (Hoffmann et al., 2019). 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 on smaller sample sizes (Gormley et al., 2016), (Finucane et al., 2018).

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 the 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 studies that have reported a shared etiology and some genetic correlation between migraine and cardiovascular and cerebrovascular endpoints (Bigal et al., 2009), (Malik et al., 2015), (Winsvold et al., 2015), (Adelborg et al., 2018), (Mahmoud et al., 2018), (Siewert et al., 2020), (Guo et al., 2020), (Daghlal et al., 2020). However, in our analysis, the migraine risk alleles did neither consistently increase nor consistently decrease the risk of coronary artery disease or the risk of hypertension.

A key role of the central nervous system (CNS) in migraine pathophysiology has emerged from animal models, human imaging, and neurophysiological studies (Ferrari et al., 2015), (Charles, 2018) while support for CNS involvement from genetic studies has been more difficult to obtain. A likely reason is the paucity of gene expression data from CNS tissue types, but recently more data has become available making such studies feasible. Our LDSC-SEG analysis using gene expression data from 13 brain regions showed an enrichment for caudate nucleus in the basal ganglia, and with chromatin-based annotations for five CNS tissue types: dorsolateral prefrontal cortex, neurospheres derived from cortex,

344 fetal brain, germinal matrix and neurospheres derived from ganglion eminence. Genes related to
345 neurogenesis were also enriched for migraine-associated genes in the gene set enrichment analysis
346 performed by MAGMA. Alterations in the structure and/or function of several brain regions (Burstein
347 et al., 2015), (Charles, 2018), (Andreou and Edvinsson, 2019) including basal ganglia, cortex,
348 hypothalamus, thalamus, brainstem, amygdala and cerebellum have been reported for individuals who
349 suffer from migraine, but the cause of these changes is not known.

350

351 In addition to the support for the hypothesis that both vascular and CNS are important in migraine
352 pathogenesis (Olesen et al., 2009), (Tfelt-Hansen and Koehler, 2011), (Charles, 2018), the tissue
353 enrichment analyses also reported some tissue types of the digestive system as well as ovary at FDR
354 5%. Given the female preponderance and suggested influence of sex hormones (e.g. menstrual related
355 migraine) in migraine (Brandes, 2006), (Borsook et al., 2014), (Delaruelle et al., 2018), the
356 involvement of the ovary is an interesting finding, although the statistical evidence for it currently
357 remains weaker compared to that for the vascular and central nervous systems.

358 A particularly interesting finding in our GWAS was the identification of risk loci containing genes that
359 encode targets for migraine-specific therapeutics. One new locus contains *CALCA* and *CALCB* genes
360 on chromosome 11 that encode calcitonin gene-related peptide (CGRP). CGRP is a highly potent
361 vasodilator with a potential role in migraine headache pain (Edvinsson et al., 2018), (Hargreaves and
362 Olesen, 2019). Several biological migraine therapeutics have recently been developed against CGRP or
363 its receptor. In particular, CGRP-related monoclonal antibodies have been successful for the preventive
364 treatment of migraine (Diener, 2019), and they are considered a major breakthrough in migraine-
365 specific treatments since the development of the triptans for acute migraine over two decades ago.

366 Another new locus contains the *HTR1F* gene that encodes serotonin 5-HT_{1F} receptor that is the target of

367 another recent migraine drug class called ditans (de Vries et al., 2020). The 5-HT_{1F} receptors are
 368 expressed in the trigeminal ganglion, the trigeminal nucleus caudalis and cephalic blood vessels
 369 (Mitsikostas and Tfelt-Hansen, 2012). Ditans provide a promising acute treatment especially for those
 370 migraine patients that cannot use triptans because of cardiovascular risk factors (Kuca et al., 2018).
 371 These two new GWAS associations near genes that are already targeted by effective migraine drugs
 372 suggest that there could be other potential drug targets among the new loci and provide a clear rationale
 373 for future GWAS efforts to increase the number of loci by increasing sample sizes further. In addition,
 374 GWAS data with migraine subtype information can help prioritize treatment targets for particular
 375 migraine symptomatology, such as aura symptoms, that currently lack treatment options. More
 376 generally, utilizing genetic evidence when selecting new drug targets is estimated to double the success
 377 rate in clinical development (Nelson et al., 2015), (King et al., 2019).

378 A limitation of our study is that a large proportion of migraine diagnoses is self-reported. Therefore, we
 379 cannot rule out misdiagnosis, such as, e.g., tension headache being reported as migraine, which could
 380 overemphasize genetic factors related to general pain mechanisms and not migraine *per se*. Regardless,
 381 the high genetic correlation that we observed supports a strong phenotypic concordance between the
 382 study collections that include also deeply phenotypes clinical cohorts from headache specialist centers,
 383 which were instrumental for the migraine subtype analyses. While the subtype data provided
 384 convincing evidence of both loci with genetic differences and other loci with genetic overlap between
 385 subtype, larger samples are still needed to achieve a more accurate picture of the similarities and
 386 differences in genetic architecture behind the subtypes.

387

388 To conclude, we report the largest GWAS meta-analysis on migraine to date detecting 123 risk loci.
 389 We demonstrated that both vascular and central nervous systems are involved in migraine
 390 pathophysiology, supporting the notion that migraine is a neurovascular disease. Our subtype analysis
 391 of migraine with aura and migraine without aura shows that these migraine subtypes have both shared
 392 risk alleles and risk alleles that appear specific to one subtype. In addition, new loci include two targets
 393 of recently developed and effective migraine treatments. Therefore, we expect that these and future
 394 GWAS data will reveal more of the heterogeneous biology of migraine and potentially point to new
 395 therapies against migraine that currently is a leading burden for population health throughout the world.

396

397

398 Methods

399 Cohorts and phenotyping

400 First, we performed a genome-wide meta-analysis on migraine including 5 study collections listed in
 401 Table 1 and Supplementary Table 1. Second, we performed subtype-specific meta-analyses on MA and
 402 on MO, both including 5 study collections listed in Table 2, for the 123 independent risk variants
 403 identified in the migraine analysis. A description of IHGC2016 (including 21 cohorts in the migraine
 404 meta-analysis, 12 cohorts in the MA meta-analysis and 11 cohorts in the MO meta-analysis) have been
 405 previously provided (Gormley et al., 2016). Details of the four other cohorts in the migraine meta-
 406 analysis as well as the four new cohorts that were only used in the migraine subtype analyses are shown
 407 in Supplementary Note 1. 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 explained (Gormley et al., 2016).

Quality control

Before conducting the meta-analysis, a standard quality control protocol was applied to each individual GWAS. Related individuals were removed from all other cohorts except HUNT by using an IBD 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 $r^2 > 0.6$, and, when available, Hardy-Weinberg equilibrium (HWE) P -value $> 1 \times 10^{-6}$ 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, the 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 Fig. 12. We conducted a sensitivity analysis on strand-ambiguous SNPs, i.e. A/T and G/C variants, by counting, for each pair of studies, how often the same allele of A/T or G/C variant was coded as the minor allele in both cohorts, as a function of MAF thresholds (Supplementary Table 16). Minor alleles were same at least in 97.39% of the variants 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 variants with $MAF < 0.25$ suggests that the strand-ambiguous variants were consistently labelled for almost every variant. Therefore, we did not exclude any variants based on possible labelling mismatches due to strand ambiguity.

429 Statistical analysis

430 The GWAS for the individual study cohorts were performed by logistic regression with an additive
 431 model of imputed dosage of the effect allele on the log-odds of migraine. The analyses for IHGC2016
 432 (Gormley et al., 2016) and 23andMe (Pickrell et al., 2016) have been described before. For UKBB data
 433 and GeneRISK data, we used PLINK v2.0 (Chang et al., 2015). For HUNT data, we used a generalized
 434 logistic mixed model with the saddlepoint approximation as implemented in SAIGE v0.20 (Zhou et al.,
 435 2018) that accounts for the genetic relatedness. All models were adjusted for sex and at least for the
 436 four leading principal components of the genetic population structure (Supplementary Table 17). Age
 437 was used as a covariate when available. A detailed description is provided in Supplementary Note 1.
 438 For the chromosome X meta-analysis, male genotypes were coded as {0,2} in all cohorts, and the
 439 GWAS were conducted with an X chromosome inactivation model that treats hemizygous males as
 440 equivalent to homozygous females (Tukiainen et al., 2014).

441 We performed an inverse-variance weighted fixed-effect meta-analysis on the five study collections by
 442 using GWAMA (Mägi and Morris, 2010). After the meta-analysis, we excluded the variants with
 443 effective sample size $N_{\text{eff}} < 2500$ to remove results with very low precision compared to the majority of
 444 variants and were left with 10,843,197 variants surpassing the QC thresholds. We estimated the
 445 effective sample size for variant i as $N_{\text{eff}(i)} = \frac{1}{2f_i(1-f_i)se_i^2}$, where f_i is the effect allele frequency for
 446 variant i and se_i is the standard error estimated by the GWAS software. This quantity approximates the
 447 value $N t (1-t) I$, where N is the total sample size (cases + controls), t is the proportion of cases and I is
 448 the imputation info (derivation in Supplementary Note 1).

449 Risk loci

450

451 There were 8,117 genome-wide significant (GWS) variants with the meta-analysis P -value $< 5 \times 10^{-8}$.
 452 For 8,067 of them, that were available in UK Biobank, an LD matrix was obtained from UK Biobank
 453 using a random sample of 10,000 individuals included in the UKBB GWAS. We defined the index
 454 variants as the LD-independent GWS variants at LD threshold of $r^2 < 0.1$ in the following way. First,
 455 the GWS variant with the lowest P -value was chosen, and subsequently all GWS variants that were in
 456 LD with the chosen variant ($r^2 > 0.1$) were excluded. Next, out of the remaining GWS variants, the
 457 variant with the lowest P -value was chosen and the GWS variants in LD with that variant were
 458 excluded. This procedure was repeated until there were no GWS variants left. Out of the 8,067 variants
 459 with LD information, 170 were LD-independent (at $r^2 < 0.1$). For 18/50 variants that were not found in
 460 UK Biobank, LD information was available from the 23andMe data, and all 18 variants were in LD
 461 ($r^2 > 0.1$) with some index variant. Two of the 18 variants (rs111404218 and rs12149936) had lower P -
 462 value than the original index variant they were in LD with and hence they replaced the original index
 463 variants. For 32 GWS variants, LD remained unknown. Thus, at this stage, the GWS associations were
 464 represented by $202 = 168 + 2 + 32$ index variants.

465

466 Next, to define the risk loci and their lead variants, an LD block around each index variant was formed
 467 by the interval spanning all GWS variants that were in high LD ($r^2 > 0.6$) with the index variant. Sizes
 468 of these regions ranged from 1 base pair (only the variant itself, e.g., the variants with unknown LD) to
 469 1,089 kb. Sets of regions that were less than 250 kb away from each other were merged (distance from
 470 the end of the first region to the beginning of the second region). This definition resulted in 126 loci.
 471 All other GWS variants were included in their nearest locus based on their position and the locus
 472 boundaries were updated, and finally loci within 250 kb from each other were merged. This resulted in

our final list of 123 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 non-coding 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) (McLaren et al., 2016). We used two different thresholds for annotating the nearest genes: a distance of 20 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 in Supplementary Table 7b.

Stepwise conditional analysis

We performed a stepwise conditional analysis (CA) on each risk locus by using FINEMAP v1.4. (Benner et al., 2016). FINEMAP uses GWAS summary statistics together with an LD reference panel and does not require individual-level data. The reference LD needs to accurately match the GWAS data for reliable fine-mapping (Benner et al., 2017) but a simpler stepwise CA is more robust to mismatches between the reference LD and the true in-sample LD of the GWAS data than the full fine-mapping. Since we did not have the in-sample LD from our GWAS data, we only carried out the CA and not the full fine-mapping. For the CA, we included only the SNPs, so 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 (N_{eff}) by using a threshold of $\pm 10\%$ of the N_{eff} 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 was necessary since without this restriction, CA led to spurious conditional P -values, such as $P < 10^{-250}$, for some loci. As a consequence, 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 CA output was in LD ($r^2 > 0.3$) 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 N_{eff} . We used the standard GWS ($P < 5 \times 10^{-8}$) 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 GTEx v8 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 project (Aguet et al., 2019). Next, we also mapped the variants in high LD ($r^2 > 0.6$) 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 8 and 9.

518 With FUMA v1.3.6 (Watanabe et al., 2017), we mapped the 123 lead variants, and the variants in high
 519 LD ($r^2 > 0.6$) with the lead variants, to the other eQTL data repositories provided by FUMA except
 520 GTEx, i.e., Blood eQTL Browser (Westra et al., 2013), BIOS QTL browser (Zhernakova et al., 2017),
 521 BRAINEAC (Ramasamy et al., 2014), MuTHER (Grundberg et al., 2012), xQTLServer (Ng et al.,
 522 2017), CommonMind Consortium (Fromer et al., 2016), eQTLGen (Võsa et al., 2018), eQTL
 523 Catalogue (Kerimov et al., 2020), DICE (Schmiedel et al., 2018), scRNA eQTLs (van der Wijst et al.,
 524 2018) and PsychENCODE (Wang et al., 2018). Results are shown in Supplementary Tables 8 and 9.
 525
 526 To study whether the lead variants were enriched in any of the 49 tissues from GTEx v8, we fitted a
 527 linear regression model where the number of lead variants that are significant *cis*-eQTLs for a specific
 528 tissue was used as the outcome, and the overall number of genes with at least one significant *cis*-eQTL
 529 reported by GTEx for the tissue was the predictor (Aguet et al. 2019). We did a separate regression
 530 model for each tissue type by leaving the tissue of interest out from the model, and we used the model
 531 fitted on the other tissues for predicting the outcome variable for the tissue type of interest. Finally, we
 532 checked in which tissues the true observed number of migraine lead variants was outside of the 95%
 533 prediction intervals as given by the function ‘predict.lm(, interval=”prediction”)’ in R software. Details
 534 of the procedure are in Supplementary Note 1.

535 LD Score regression

536 We estimated both the SNP-heritability (h^2_{SNP}) of migraine and pairwise genetic correlations
 537 (r_G) between each pair of study collections using LDSC v1.0.0 (Bulik-Sullivan et al., 2015b) (Bulik-
 538 Sullivan et al., 2015a). SNP-heritability and genetic correlations were estimated using European LD
 539 scores from the 1000 Genomes Project Phase 3 data for the HapMap3 SNPs, downloaded from

540 <https://data.broadinstitute.org/alkesgroup/LDSCORE/>. From the meta-analysis association statistics, we
541 excluded variants that did not match with the HapMap3 SNPs, that had strand ambiguity (i.e., A/T or
542 G/C SNPs), MAF < 0.01, INFO < 0.6 or missingness more than two-thirds of the 90th percentile of the
543 total sample size, that resided in long-range LD regions (Price et al., 2008), in centromere regions or in
544 the major histocompatibility locus (MHC) of chromosome 6, leaving 1,165,201 SNPs for the LDSC
545 analyses. We used a migraine population prevalence of 16% and a sample proportion of cases of 11.7%
546 = 102,084/(102,084 + 771,257) to turn the LDSC slope into the estimate of h^2_{SNP} on the liability scale
547 (Lee et al., 2011). Pairwise genetic correlation results are listed in Supplementary Table 2. We note that
548 in the previous migraine meta-analysis (Gormley et al. 2016), LDSC reported h^2_{SNP} value of 14.6%
549 (13.8 – 15.5%), which was considerably larger than the value 11.2% (10.8 – 11.6%) that we report in
550 our analysis. When we ran our LDSC pipeline on the data of Gormley et al. (2016), we estimated
551 h^2_{SNP} value of 10.6% (10.1 – 11.1%). Thus, it seems that our liability transformation estimates lower
552 values of heritability than the transformation used by Gormley et al. (2016).

553 Stratified LD Score regression

554 We used stratified LD Score regression (S-LDSC) to partition the SNP heritability by functional
555 genomic annotations (Finucane et al., 2015). We used the baseline model that contains 24 main
556 functional categories including conserved, coding and regulatory regions of the genome and different
557 histone modifications. We used the same QC as with the univariate LDSC, and the baseline European
558 LD scores estimated from the 1000 Genomes Project Phase 3, downloaded from
559 <https://data.broadinstitute.org/alkesgroup/LDSCORE/>. Results are listed in Supplementary Tables
560 10a,b.

561 Subtype analyses of migraine with and without aura

First, we combined new MA and MO data (Table 2) with the previously used migraine subtype-specific meta-analysis data (Gormley et al., 2016), and estimated migraine subtype-specific effect sizes for the 123 lead variants from the migraine meta-analysis. We tested how often the direction of allelic effects was similar between the IHGC MA/MO and the new cohorts using a binomial test (Supplementary Table 11b). Next, we stratified the lead variants by using the information from the migraine subtype-specific analyses. For each of the variants, we estimated probabilities between four possible explanations of the observed data that we call ‘NULL’, ‘MO’, ‘MA’ and ‘BOTH’. Under model NULL the effect is not present in either of the migraine subtypes (i.e., the effect is zero); under model MO or MA the effect is present only in MO or only in MA but not in both; and under model BOTH, a non-zero effect is shared by both MO and MA. We used a Bayesian approach for model comparison that combines a bivariate Gaussian prior distribution on the two effect sizes with a bivariate Gaussian approximation to the likelihood using GWAS summary statistics (Trochet et al., 2019). Across all models, the prior standard deviation for the effect is 0.2 on the log-odds scale for non-zero effects and 0 for a zero effect. The bivariate priors for the four models are as follows: NULL assumes a zero effect in both migraine subtypes, MO and MA assume a non-zero effect for one subtype and a zero effect for the other subtype, and BOTH combines the fixed-effect model (exactly the same effect in both subtypes) with the independent-effects model (the two effect sizes are non-zero but uncorrelated with each other) with equal weights. Finally, we assumed that each of the four models (NULL, MO, MA, BOTH) is equally probable *a priori*, which we considered an appropriate assumption since all these variants show a convincing association to overall migraine ($P < 5 \times 10^{-8}$). Then we used the Bayes formula to work out the posterior probability on each model. The results are shown in Fig. 3a, thresholded by a probability cut-off of 95% and in Supplementary Table 11a. The correlation parameter between MO and MA GWAS statistics needed in the bivariate likelihood approximation was estimated to be 0.148

585 using the empirical Pearson correlation of the effect size estimates of the common variants (MAF >
586 0.05) that did not show a strong association to either of the migraine subtypes ($P > 1 \times 10^{-4}$)
587 (Cichonska et al., 2016). We tested whether the effect sizes between MA and MO were equal at a
588 Bonferroni corrected significance threshold of $\alpha = 0.05/123$ by using a normal approximation and
589 accounting for the correlation in effect size estimators.

590 We note that the amount of information in the data (“statistical power”) is taken automatically into
591 account in this model comparison, which we consider an advantage compared to a comparison of the
592 raw P -values between the subtype analyses that does not automatically account for statistical power.
593 We want to point out that the inference in the model comparison approach is conditional on the
594 particular set of models being included in the comparison as well as on the particular choice of the prior
595 distributions.

596 PheWAS with NHGRI GWAS Catalog and FinnGen R4

597

598 We performed phenome-wide association studies (PheWAS) for the 123 lead variants using the
599 NHGRI GWAS Catalog and the FinnGen R4 GWAS summary statistics. In addition, we performed the
600 same lookups for the 123 risk loci including all variants in high LD ($r^2 > 0.6$) with the lead variants.
601 With the GWAS Catalog, we first downloaded all the available results (4,314 traits) from the GWAS
602 Catalog webpage (accessed 6.4.2020). Next, we obtained all the associations for the 123 risk loci with
603 all the high LD variants included using P -value thresholds of $P < 1 \times 10^{-5}$, $P < 1 \times 10^{-6}$ and $P < 1 \times$
604 10^{-4} (Supplementary Tables 12a-c). Because the GWAS Catalog includes results from several different
605 GWAS for the same phenotype or for a very similar phenotype with a different name, we divided the
606 phenotype associations into broader categories. The new categories are listed in Supplementary Table

18. 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 using the same three *P*-value thresholds as with the GWAS Catalog (Supplementary Tables 12a-c). We also divided similar endpoints into broader categories that are listed in Supplementary Table 19.

We tested the direction of allelic effects between migraine and the following three traits that shared multiple associated variants with migraine: coronary artery disease (CAD) (Nelson et al., 2017), diastolic blood pressure (Loh et al., 2018) and systolic blood pressure (Loh et al., 2018). We first took all migraine lead variants that were available also in the summary statistics of the other trait without any *P*-value 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 *P*-value threshold of 1×10^{-5} for the association with the other trait. Results are in the Supplementary Table 12d.

LD Score regression applied to specifically expressed genes (LDSC-SEG)

We used LD Score regression applied to specifically expressed genes (LDSC- SEG) (Finucane et al., 2018) 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/LDSC_SEG_ldscores/: multi-tissue gene expression, multi-tissue chromatin, GTEx brain, Cahoy, Corces ATAC and ImmGen LD Scores. FDR was controlled by the Benjamini-Hochberg method. The results are in Supplementary Tables 13a-f.

There were no significant results with the Cahoy, Corces ATAC and ImmGen data at FDR 5%.

630

631 MAGMA (Multi-marker Analysis of GenoMic Annotation)

632 We applied MAGMA v1.07 (de Leeuw et al., 2015) to identify genes and gene sets associated with the
 633 migraine meta-analysis results. First, we mapped the meta-analysis SNPs to 18,985 protein-coding
 634 genes based on their physical position in the NCBI 37 build by using default settings of MAGMA.
 635 Next, we performed a gene-based analysis using the default SNPwise-mean model and the same UK
 636 Biobank LD reference as for the other analyses. We applied a Bonferroni correction ($\alpha = 0.05/18,985$)
 637 to identify significantly associated genes for migraine with the results listed in Supplementary Table
 638 15a. Finally, we used the results from the gene-based analysis to perform a gene-set analysis by using
 639 two different gene-set collections from the Molecular Signature Database v.7.0 (Liberzon et al., 2011),
 640 (Subramanian et al., 2005): the curated gene sets containing 5,500 gene sets and the GO gene sets
 641 containing 9,988 gene sets. The gene-set analysis was performed using the competitive gene set model
 642 that tests whether the genes in the gene-set are more strongly associated with the phenotype compared
 643 to the other genes. To correct for multiple testing, we used a Bonferroni correction ($\alpha =$
 644 $0.05/(5,500+9,988)$). Results are in Supplementary Tables 15b,c and in Supplementary Fig. 13.

645 DEPICT (Data-driven Expression Prioritized Integration for Complex Traits)

646 DEPICT (Pers et al., 2015) is an integrative tool to identify the most likely causal genes at associated
 647 loci, and enriched pathways and tissues or cell types in which the genes from the associated loci are
 648 highly expressed. As an input, DEPICT takes a set of trait-associated SNPs. First, DEPICT uses co-
 649 regulation data from 77,840 microarrays to predict biological functions of genes and to construct
 650 14,461 reconstituted gene sets. Next, information of similar predicted gene functions is used to identify
 651 and prioritize gene sets that are enriched for genes in the associated loci. For the tissue and cell type

652 enrichment analysis, DEPICT uses a set of 37,427 human gene expression microarrays. We used
653 DEPICT v1.194, and ran the analyses twice for each of the P -value thresholds for clumping, as
654 recommended in (Pers et al., 2015), and using the default settings of 500 permutations for bias
655 adjustment and 50 replications for the FDR estimation and for the P -value calculation. As an input, we
656 used only the autosomal SNPs and the same UK Biobank LD reference data as for the other analyses.
657 First, we ran the analysis using a clumping P -value threshold of 5×10^{-8} that resulted in 165 clumps
658 formed from 7,672 variants (Supplementary Tables 14d-f). Second, we used a P -value threshold of $1 \times$
659 10^{-5} leading to 612 clumps formed from 22,480 variants (Supplementary Tables 14a-c).

660

661

Table 1 Five migraine study collections included in the meta-analysis.

Abbreviation	Full Name	Ethnicity	Cases	Controls	Case %	Migraine Definition
IHGC2016*	Gormley et al. 2016 (no 23andMe)	European descent	29,209	172,931	14.4	Self-reported and ICHD-II
23andMe	23andMe, Inc. (23andMe.com)	European descent	53,109	230,876	18.7	Self-reported
UKBB	UK Biobank (ukbiobank.ac.uk)	European, British	10,881	330,170	3.2	Self-reported
GeneRISK	GeneRISK (generisk.fi)	European, Finnish	1,084	4,857	18.2	Self-reported
HUNT	Nord-Trøndelag Health Study (ntnu.edu/hunt)	European, Norwegian	7,801	32,423	19.4	Self-reported migraine or fulfilling modified ICHD-II criteria

*IHGC2016 is a meta-analysis of 21 studies listed in Supplementary Table 1. Some studies of IHGC2016 determined migraine status through clinical phenotyping while migraine status in other studies is based on self-reported information. ICHD-II = the International Classification of Headache Disorders 2nd edition.

Table 2 Study collections included in MO and MA subtype analyses.

Abbreviation	Full Name	Ethnicity	Subtype	Cases	Controls
IHGC2016*	Gormley et al. 2016	European descent	MO	8,348	139,622
			MA	6,332	144,883
UKBB	UK Biobank (ukbiobank.ac.uk)	European, British	MO	187	320,139
			MA	1,333	320,139
deCODE	deCODE Genetics Inc.	European, Icelandic	MO	1,648	193,050
			MA	2,297	209,338
DBDS	Danish Blood Donor Study	European, Danish	MO	3,756	28,045
			MA	3,938	28,045
LUMINA	LUMINA migraine without aura or with aura	European, Dutch	MO	1,116	1,445
			MA	724	1,447

*IHGC2016 MO is a meta-analysis of 11 studies and IHGC2016 MA is a meta-analysis of 12 studies listed in Gormley et al. 2016. MO = migraine without aura, MA = migraine with aura.

688 **Table 3 LDSC-SEG results that are significant at FDR 5%.**

Tissue/Cell type and histone mark	Tissue category	P-value	FDR
Multi-tissue gene expression data			
Aorta	Cardiovascular	1.78E-04	0.029
Tibial Artery	Cardiovascular	3.60E-04	0.029
Coronary Artery	Cardiovascular	4.29E-04	0.029
Gene expression data of 13 brain regions from GTEx			
Caudate (basal ganglia)	Central nervous system	6.00E-04	0.008
Multi-tissue chromatin annotation data			
Fetal Brain Female, H3K4me3	Central nervous system	2.49E-05	0.012
Brain Dorsolateral Prefrontal Cortex, H3K27ac	Central nervous system	8.43E-05	0.018
Brain Dorsolateral Prefrontal Cortex, H3K4me3	Central nervous system	1.11E-04	0.018
Aorta, H3K4me1	Cardiovascular	2.57E-04	0.031
Stomach Mucosa, H3K36me3	Digestive	3.36E-04	0.032
Aorta, H3K27ac	Cardiovascular	4.40E-04	0.032
Artery-Tibial ENTEX, H3K4me1	Cardiovascular	4.53E-04	0.032
Ganglion Eminence derived primary cultured neurospheres, H3K4me3	Central nervous system	6.53E-04	0.04
Brain Germinal Matrix, H3K4me3	Central nervous system	8.42E-04	0.043
Aorta ENTEX, H3K27ac	Cardiovascular	1.11E-03	0.043
Artery-Coronary ENTEX, H3K4me3	Cardiovascular	1.13E-03	0.043
Cortex derived primary cultured neurospheres, H3K36me3	Central nervous system	1.14E-03	0.043
Ovary, H3K27ac	Other	1.15E-03	0.043
Cortex derived primary cultured neurospheres, H3K4me3	Central nervous system	1.29E-03	0.045
Aorta ENTEX, H3K4me1	Cardiovascular	1.39E-03	0.045
Stomach Smooth Muscle, H3K4me3	Musculoskeletal/Connective	1.55E-03	0.047

689 *P*-value from testing whether the regression coefficient is positive. FDR = False discovery rate based
690 on Benjamini-Hochberg method. Full results are in Supplementary Tables 12a-f.

691
692
693 **Supplementary Tables 1-19.** MANU_SUPP_TABLES_all6f.xlsx

Fig. 1: Manhattan plot of migraine GWAS meta-analysis (N = 873,341; 102,084 cases and 771,257 controls). On the X-axis, variants are plotted along the 22 autosomes and the X chromosome. Y-axis shows the statistical strength of the association from the fixed-effect meta-analysis as the negative log₁₀ of the *P*-value. Horizontal line is the genome-wide significance threshold ($P = 5 \times 10^{-8}$). The 123 risk loci passing the threshold are divided into 86 new loci (purple) and 37 previously known loci (red).

Fig. 2: Locuszoom-plots of two novel migraine loci with genes that are targets of recent migraine specific drugs.

a, Locus containing *CALCA* and *CALCB* genes which encode CGRP, that is the target of preventive and acute therapies via monoclonal antibodies and gepants. **b**, Locus containing *HTR1F* gene that encodes a serotonin 5-HT_{1F} receptor that is the target of acute therapies via ditans. The squared correlation to the lead variant is shown by colors based on the UK Biobank data for variants that have an effective sample size $\pm 20\%$ of the lead variant's effective sample size. Horizontal line corresponds to $P = 5 \times 10^{-8}$. Blue graph shows the recombination rate.

Fig. 3: Lead variants stratified by migraine subtype for risk loci with minor allele frequency > 5%.

a, Axes show the negative log10 *P*-value of MO (X-axis) and MA (Y-axis) analyses. Symbols that are colored and annotated indicate > 95% posterior probability that a non-zero effect is present in both MO and MA (model BOTH), or that the effect is present only in MO or only in MA but not both (models MO and MA, respectively). Variants with a probability less than 95% for each of the three models are shown as gray. **b**, Axes show logarithm of odds ratios for MO (X-axis) and MA (Y-axis) calculated for the migraine risk allele. The effects at variants that have been colored and annotated differ between the subtypes at significance level of $0.0004 = 0.05/123$. The 95% confidence intervals are shown for the annotated variants.

MO = migraine without aura, MA = migraine with aura.

739 **Supplementary Figures:** Supplementary_Figures_a4.docx

740

741 Data access

742 Results for 8,117 genome-wide significant SNP associations ($P < 5 \times 10^{-8}$) from the meta-analysis
 743 including 23andMe data are available on the International Headache Genetics Consortium (IHGC)
 744 website (<http://www.headachegenetics.org/>). Genome-wide summary statistics for the four other study
 745 collections except 23andMe are available as described on the IHGC website. The full GWAS summary
 746 statistics for the 23andMe discovery data set will be made available through 23andMe to qualified
 747 researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants.
 748 Please visit research.23andme.com/collaborate/#publication for more information and to apply to
 749 access the data.

750

751 Acknowledgements

752 We thank the study participants for their contribution to this research. We also thank the numerous
 753 individuals who contributed to sample collection, storage, handling, phenotyping, and genotyping for
 754 each of the individual cohorts. We acknowledge the participants and investigators of the FinnGen
 755 study. This research has been conducted using the UK Biobank Resource under Application Number
 756 22627. A list of study-specific acknowledgements and funding information can be found in the
 757 Supplementary Note 1.

758

759 Competing interests

760 GB, TET, SHM, HS and KS are employees of deCODE genetics/Amgen.

761

762 PG was an employee of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth,
763 NJ, USA.

764

765 TF reports possible competing interest for Electrocore (participation in clinical studies), Novartis
766 (speakers' honoraria, participation in advisory boards, participation in clinical studies), Teva (speakers'
767 honoraria, participation in advisory boards), Lilly (speakers' honoraria, participation in clinical
768 studies), Bayer (speakers' honoraria).

769

770 MK has served on Advisory Boards for MSD and Allergan; has received funding for travel and/or
771 speaker honoraria from MSD, Allergan, TEVA, Novartis, and Genzyme; has received compensation
772 for producing educational material from TEVA and Allergan; has received research support from
773 Helsinki University Central Hospital; and holds stock/stock options and/or has received Board of
774 Directors compensation from Helsinki Headache Center.

775

776 TK reports having received honoraria from Eli Lilly, Newsenselab, and Total for providing
777 methodological advice and from the BMJ for editorial services.

778

779 AP is the Scientific Director of the public-private partnership project FinnGen that has 12 industry
780 partners that provide funding for the FinnGen project.

781

782 VA has served on advisory board for Allergan and Lilly.

783

784 Other authors report no conflicts of interests.

References

- Adelborg K, Szépligeti SK, Holland-Bill L, et al. (2018) Migraine and risk of cardiovascular diseases: Danish population based matched cohort study. *BMJ* 360: k96.
- Aguet F, Barbeira AN, Bonazzola R, et al. (2019) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*. DOI: 10.1101/787903. 787903.
- Andreou AP and Edvinsson L (2019) Mechanisms of migraine as a chronic evolutive condition. *The Journal of Headache and Pain* 20(1): 117.
- Anttila V, Stefansson H, Kallela M, et al. (2010) Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. 2010/08/29 ed., 869-873.
- Anttila V, Wessman M, Kallela M, et al. (2018) Chapter 31 - Genetics of migraine. In: Geschwind DH, Paulson HL and Klein C (eds) *Handbook of Clinical Neurology*. Elsevier, pp.493-503.
- Anttila V, Winsvold BS, Gormley P, et al. (2013) Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nature Genetics* 45(8): 912-917.
- Benner C, Havulinna AS, Järvelin M-R, et al. (2017) Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *American Journal of Human Genetics* 101(4): 539-551.
- Benner C, Spencer CCA, Havulinna AS, et al. (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics (Oxford, England)* 32(10): 1493-1501.
- Bigal ME, Kurth T, Hu H, et al. (2009) Migraine and cardiovascular disease: possible mechanisms of interaction. *Neurology* 72(21): 1864-1871.
- Borsook D, Erpelding N, Lebel A, et al. (2014) Sex and the migraine brain. *Neurobiology of Disease* 68: 200-214.
- Brandes JL (2006) The Influence of Estrogen on MigraineA Systematic Review. *JAMA* 295(15): 1824-1830.
- Bulik-Sullivan B, Finucane HK, Anttila V, et al. (2015a) An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 47(11): 1236-1241.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. (2015b) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* 47(3): 291-295.
- Burstein R, Nosedà R and Borsook D (2015) Migraine: multiple processes, complex pathophysiology. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35(17): 6619-6629.
- Chang CC, Chow CC, Tellier LCAM, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4(1): s13742-13748.
- Chang X, Pellegrino R, Garifallou J, et al. (2018) Common variants at 5q33.1 predispose to migraine in African-American children. *J Med Genet* 55(12): 831.
- Charles A (2018) The pathophysiology of migraine: implications for clinical management. *The Lancet Neurology* 17(2): 174-182.

- Charles A and Pozo-Rosich P (2019) Targeting calcitonin gene-related peptide: a new era in migraine therapy. *The Lancet* 394(10210): 1765-1774.
- Chasman DI, Anttila V, Buring JE, et al. (2014) Selectivity in Genetic Association with Sub-classified Migraine in Women. *PLoS genetics* 10(5): e1004366.
- Chasman DI, Schürks M, Anttila V, et al. (2011) Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nature Genetics* 43: 695.
- Chen S-P, Fuh J-L, Chung M-Y, et al. (2018) Genome-wide association study identifies novel susceptibility loci for migraine in Han Chinese resided in Taiwan. *Cephalalgia* 38(3): 466-475.
- Cichonska A, Rousu J, Marttinen P, et al. (2016) metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics* 32(13): 1981-1989.
- Cohen J, Pertsemlidis A, Kotowski IK, et al. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. 161-165.
- Daghlal I, Guo Y and Chasman DI (2020) Effect of genetic liability to migraine on coronary artery disease and atrial fibrillation: a Mendelian randomization study. *European Journal of Neurology* 27(3): 550-556.
- de Boer I, van den Maagdenberg AMJM and Terwindt GM (2019) Advance in genetics of migraine. *Current opinion in neurology* 32(3): 413-421.
- de Leeuw CA, Mooij JM, Heskes T, et al. (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* 11(4): e1004219.
- de Vries B, Frants RR, Ferrari MD, et al. (2009) Molecular genetics of migraine. *Human Genetics* 126(1): 115.
- de Vries T, Villalón CM and MaassenVanDenBrink A (2020) Pharmacological treatment of migraine: CGRP and 5-HT beyond the triptans. *Pharmacology & Therapeutics*. DOI: <https://doi.org/10.1016/j.pharmthera.2020.107528>. 107528.
- Delaruelle Z, Ivanova TA, Khan S, et al. (2018) Male and female sex hormones in primary headaches. *J Headache Pain* 19(1): 117.
- Diener H-C (2019) CGRP antibodies for migraine prevention — new kids on the block. *Nature Reviews Neurology* 15(3): 129-130.
- Do TP, Guo S and Ashina M (2019) Therapeutic novelties in migraine: new drugs, new hope? *J Headache Pain* 20(1): 37.
- Dodick DW (2018) Migraine. *The Lancet* 391(10127): 1315-1330.
- Edvinsson L, Haanes KA, Warfvinge K, et al. (2018) CGRP as the target of new migraine therapies — successful translation from bench to clinic. *Nature reviews. Neurology* 14(6): 338-350.
- Ferrari MDP, Klever RRM, Terwindt GMM, et al. (2015) Migraine pathophysiology: lessons from mouse models and human genetics. *Lancet Neurology, The* 14(1): 65-80.
- Finucane HK, Bulik-Sullivan B, Gusev A, et al. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* 47(11): 1228-1235.
- Finucane HK, Reshef YA, Anttila V, et al. (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* 50(4): 621-629.
- Flannick J, Thorleifsson G, Beer NL, et al. (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. 357-363.
- Freilinger T, Anttila V, de Vries B, et al. (2012) Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nature Genetics* 44(7): 777-782.
- Fromer M, Roussos P, Sieberts SK, et al. (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* 19(11): 1442-1453.

874 Gervil M, Ulrich V, Kaprio J, et al. (1999) The relative role of genetic and environmental factors in
875 migraine without aura. *Neurology* 53(5): 995.

876 Goadsby PJ, Dodick DW, Ailani J, et al. (2020) Safety, tolerability, and efficacy of orally administered
877 atogepant for the prevention of episodic migraine in adults: a double-blind, randomised phase
878 2b/3 trial. *The Lancet Neurology* 19(9): 727-737.

879 Gormley P, Anttila V, Winsvold BS, et al. (2016) Meta-analysis of 375,000 individuals identifies 38
880 susceptibility loci for migraine. *Nature Genetics* 48(8): 856-+.

881 Gormley P, Kurki M, Kurki MI, et al. (2018) Common Variant Burden Contributes to the Familial
882 Aggregation of Migraine in 1,589 Families. *Neuron* 98(4): 743-753.e744.

883 Grundberg E, Small KS, Hedman ÅK, et al. (2012) Mapping cis- and trans-regulatory effects across
884 multiple tissues in twins. *Nature Genetics* 44(10): 1084-1089.

885 Guo Y, Rist PM, Daghlal I, et al. (2020) A genome-wide cross-phenotype meta-analysis of the
886 association of blood pressure with migraine. *Nature Communications* 11(1): 3368.

887 Hargreaves R and Olesen J (2019) Calcitonin Gene-Related Peptide Modulators – The History and
888 Renaissance of a New Migraine Drug Class. *Headache: The Journal of Head and Face Pain*
889 59(6): 951-970.

890 Headache Classification Committee of the International Headache Society (IHS) (2018) Headache
891 Classification Committee of the International Headache Society (IHS) The International
892 Classification of Headache Disorders, 3rd edition. *Cephalalgia* 38(1): 1-211.

893 Hippisley-Cox J, Coupland C and Brindle P (2017) Development and validation of QRISK3 risk
894 prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study.
895 *BMJ* 357: j2099.

896 Hoffmann J, Baca SM and Akerman S (2019) Neurovascular mechanisms of migraine and cluster
897 headache. *Journal of cerebral blood flow and metabolism : official journal of the International*
898 *Society of Cerebral Blood Flow and Metabolism* 39(4): 573-594.

899 Howard DM, Adams MJ, Clarke T-K, et al. (2019) Genome-wide meta-analysis of depression
900 identifies 102 independent variants and highlights the importance of the prefrontal brain
901 regions. *Nature neuroscience* 22(3): 343-352.

902 Jacobs B and Dussor G (2016) Neurovascular contributions to migraine: Moving beyond vasodilation.
903 *Neuroscience* 338: 130-144.

904 Kallela M, Wessman M, Havanka H, et al. (2001) Familial migraine with and without aura: clinical
905 characteristics and co-occurrence. *Eur J Neurol* 8(5): 441-449.

906 Kerimov N, Hayhurst JD, Manning JR, et al. (2020) eQTL Catalogue: a compendium of uniformly
907 processed human gene expression and splicing QTLs. *bioRxiv*. DOI:
908 10.1101/2020.01.29.924266. 2020.2001.2029.924266.

909 King EA, Davis JW and Degner JF (2019) Are drug targets with genetic support twice as likely to be
910 approved? Revised estimates of the impact of genetic support for drug mechanisms on the
911 probability of drug approval. *PLoS genetics* 15(12): e1008489.

912 Kuca B, Silberstein SD, Wietecha L, et al. (2018) Lasmiditan is an effective acute treatment for
913 migraine. *Neurology* 91(24): e2222-e2232.

914 Kurth T, Winter AC, Eliassen AH, et al. (2016) Migraine and risk of cardiovascular disease in women:
915 prospective cohort study. *BMJ* 353: i2610.

916 Lee SH, Wray NR, Goddard ME, et al. (2011) Estimating missing heritability for disease from
917 genome-wide association studies. *American Journal of Human Genetics* 88(3): 294-305.

918 Liberzon A, Subramanian A, Pinchback R, et al. (2011) Molecular signatures database (MSigDB) 3.0.
919 *Bioinformatics* 27(12): 1739-1740.

920 Lipton RB, Dodick DW, Ailani J, et al. (2019) Effect of Ubrogepant vs Placebo on Pain and the Most
921 Bothersome Associated Symptom in the Acute Treatment of Migraine: The ACHIEVE II
922 Randomized Clinical Trial. *JAMA* 322(19): 1887-1898.

923 Loh P-R, Kichaev G, Gazal S, et al. (2018) Mixed-model association for biobank-scale datasets. *Nature*
924 *Genetics* 50(7): 906-908.

925 Mägi R and Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. *BMC*
926 *bioinformatics* 11(1): 288.

927 Mahmoud AN, Mentias A, Elgendy AY, et al. (2018) Migraine and the risk of cardiovascular and
928 cerebrovascular events: a meta-analysis of 16 cohort studies including 1 152 407 subjects. *BMJ*
929 *open* 8(3): e020498.

930 Malik R, Freilinger T, Winsvold BS, et al. (2015) Shared genetic basis for migraine and ischemic
931 stroke: A genome-wide analysis of common variants. *Neurology* 84(21): 2132-2145.

932 McLaren W, Gil L, Hunt SE, et al. (2016) The Ensembl Variant Effect Predictor. *Genome biology*
933 17(1): 122.

934 Mitsikostas DD and Tfelt-Hansen P (2012) Targeting to 5-HT_{1F} receptor subtype for migraine
935 treatment: lessons from the past, implications for the future. *Cent Nerv Syst Agents Med Chem*
936 12(4): 241-249.

937 Nagel M, Jansen PR, Stringer S, et al. (2018) Meta-analysis of genome-wide association studies for
938 neuroticism in 449,484 individuals identifies novel genetic loci and pathways.

939 Need AC and Goldstein DB (2014) Schizophrenia Genetics Comes of Age. *Neuron* 83(4): 760-763.

940 Nelson CP, Goel A, Butterworth AS, et al. (2017) Association analyses based on false discovery rate
941 implicate new loci for coronary artery disease. *Nature Genetics* 49(9): 1385-1391.

942 Nelson MR, Tipney H, Painter JL, et al. (2015) The support of human genetic evidence for approved
943 drug indications. *Nature Genetics* 47(8): 856-860.

944 Ng B, White CC, Klein H-U, et al. (2017) An xQTL map integrates the genetic architecture of the
945 human brain's transcriptome and epigenome. *Nature neuroscience* 20(10): 1418-1426.

946 Nyholt D, Anttila V, Winsvold B, et al. (2014) Concordance of genetic risk across migraine subgroups:
947 Impact on current and future genetic association studies. *Cephalalgia : an international journal*
948 *of headache* 35.

949 Olesen J, Burstein R, Ashina M, et al. (2009) Origin of pain in migraine: evidence for peripheral
950 sensitisation. *The Lancet Neurology* 8(7): 679-690.

951 Ophoff RA, Terwindt GM, Vergouwe MN, et al. (1996) Familial Hemiplegic Migraine and Episodic
952 Ataxia Type-2 Are Caused by Mutations in the Ca²⁺ Channel Gene CACNL1A4. *Cell* 87(3):
953 543-552.

954 Pardiñas AF, Holmans P, Pocklington AJ, et al. (2018) Common schizophrenia alleles are enriched in
955 mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*
956 50(3): 381-389.

957 Pers TH, Karjalainen JM, Chan Y, et al. (2015) Biological interpretation of genome-wide association
958 studies using predicted gene functions. *Nature Communications* 6.

959 Pickrell JK, Berisa T, Liu JZ, et al. (2016) Detection and interpretation of shared genetic influences on
960 42 human traits. *Nature Genetics* 48: 709.

961 Price AL, Weale ME, Patterson N, et al. (2008) Long-range LD can confound genome scans in
962 admixed populations. *Elsevier*, 132-139.

963 Ramasamy A, Trabzuni D, Guelfi S, et al. (2014) Genetic variability in the regulation of gene
964 expression in ten regions of the human brain. *Nature neuroscience* 17(10): 1418-1428.

965 Russell MB and Olesen J (1995) Increased familial risk and evidence of genetic factor in migraine.
966 *BMJ (Clinical research ed.)* 311(7004): 541-544.

967 Russell MB, Rasmussen BK, Thorvaldsen PER, et al. (1995) Prevalence and Sex-Ratio of the Subtypes
968 of Migraine. *International journal of epidemiology* 24(3): 612-618.

969 Schmiedel BJ, Singh D, Madrigal A, et al. (2018) Impact of Genetic Polymorphisms on Human
970 Immune Cell Gene Expression. *Cell* 175(6): 1701-1715.e1716.

971 Siewert KM, Klarin D, Damrauer SM, et al. (2020) Cross-trait analyses with migraine reveal
972 widespread pleiotropy and suggest a vascular component to migraine headache. *International*
973 *journal of epidemiology*. DOI: 10.1093/ije/dyaa050. dyaa050.

974 Steiner TJ, Stovner LJ, Jensen R, et al. (2020) Migraine remains second among the world's causes of
975 disability, and first among young women: findings from GBD2019. *The Journal of Headache*
976 *and Pain* 21(1): 137.

977 Subramanian A, Tamayo P, Mootha VK, et al. (2005) Gene set enrichment analysis: A knowledge-
978 based approach for interpreting genome-wide expression profiles. *Proceedings of the National*
979 *Academy of Sciences* 102(43): 15545-15550.

980 Tfelt-Hansen P and Olesen J (2012) Taking the Negative View of Current Migraine Treatments. *CNS*
981 *Drugs* 26(5): 375-382.

982 Tfelt-Hansen PC and Koehler PJ (2011) One Hundred Years of Migraine Research: Major Clinical and
983 Scientific Observations From 1910 to 2010. *Headache: The Journal of Head and Face Pain*
984 51(5): 752-778.

985 Trochet H, Pirinen M, Band G, et al. (2019) Bayesian meta-analysis across genome-wide association
986 studies of diverse phenotypes. *Genetic epidemiology* 43(5): 532-547.

987 Tukiainen T, Pirinen M, Sarin A-P, et al. (2014) Chromosome X-Wide Association Study Identifies
988 Loci for Fasting Insulin and Height and Evidence for Incomplete Dosage Compensation. *PLoS*
989 *genetics* 10(2): e1004127.

990 Ulrich V, Gervil M, Kyvik KO, et al. (1999) Evidence of a genetic factor in migraine with aura: A
991 population-based Danish twin study. *Annals of Neurology* 45(2): 242-246.

992 van der Wijst MGP, Brugge H, de Vries DH, et al. (2018) Single-cell RNA sequencing identifies
993 celltype-specific cis-eQTLs and co-expression QTLs. *Nature Genetics* 50(4): 493-497.

994 Vos T, Lim SS, Abbafati C, et al. (2020) Global burden of 369 diseases and injuries in 204 countries
995 and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019.
996 *The Lancet* 396(10258): 1204-1222.

997 Vösa U, Claringbould A, Westra H-J, et al. (2018) Unraveling the polygenic architecture of complex
998 traits using blood eQTL metaanalysis. *bioRxiv*. DOI: 10.1101/447367. 447367.

999 Wang D, Liu S, Warrell J, et al. (2018) Comprehensive functional genomic resource and integrative
1000 model for the human brain. *Science (New York, N.Y.)* 362(6420): eaat8464.

1001 Watanabe K, Taskesen E, van Bochoven A, et al. (2017) Functional mapping and annotation of genetic
1002 associations with FUMA. *Nature Communications* 8(1): 1826.

1003 Westra H-J, Peters MJ, Esko T, et al. (2013) Systematic identification of trans eQTLs as putative
1004 drivers of known disease associations. *Nature Genetics* 45(10): 1238-1243.

1005 Winsvold BS, Nelson CP, Malik R, et al. (2015) Genetic analysis for a shared biological basis between
1006 migraine and coronary artery disease. *Neurology Genetics* 1(1): e10.

1007 Yang J, Weedon MN, Purcell S, et al. (2011) Genomic inflation factors under polygenic inheritance.
1008 *European Journal of Human Genetics* 19(7): 807-812.

1009 Zhao H, Eising E, de Vries B, et al. (2016) Gene-based pleiotropy across migraine with aura and
1010 migraine without aura patient groups. *Cephalalgia : an international journal of headache* 36(7):
1011 648-657.

1012 Zhernakova DV, Deelen P, Vermaat M, et al. (2017) Identification of context-dependent expression
1013 quantitative trait loci in whole blood. *Nature Genetics* 49(1): 139-145.

1014 Zhou W, Nielsen JB, Fritsche LG, et al. (2018) Efficiently controlling for case-control imbalance and
1015 sample relatedness in large-scale genetic association studies. *Nature Genetics* 50(9): 1335-1341.
1016