

1 Fine-mapping a genome-wide meta-analysis of 98,374 migraine cases identifies 181
2 sets of candidate causal variants

3

4 Heidi Hautakangas^{1*}, FinnGen^{**}, International Headache Genetics Consortium^{**},
5 HUNT All-in Headache^{**}, Aarno Palotie^{1,2,3}, Matti Pirinen^{1,4,5*}

6

7 1. Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science
8 (HiLIFE), University of Helsinki, Helsinki, Finland. 2. Analytic and Translational Genetics
9 Unit, Department of Medicine, Department of Neurology and Department of Psychiatry
10 Massachusetts General Hospital, Boston, MA, USA. 3. The Stanley Center for
11 Psychiatric Research and Program in Medical and Population Genetics, The Broad
12 Institute of MIT and Harvard, Cambridge, MA, USA. 4. Department of Public Health,
13 University of Helsinki, Helsinki, Finland. 5. Department of Mathematics and Statistics,
14 University of Helsinki, Helsinki, Finland.

15

16 * Correspondence to: heidi.hautakangas@helsinki.fi and matti.pirinen@helsinki.fi

17 ** A list of authors and their affiliations appears at the end of the paper.

18

19

20 Abstract

21

22 Migraine is a highly prevalent neurovascular disorder for which genome-wide
23 association studies (GWAS) have identified over one hundred risk loci, yet the
24 causal variants and genes remain mostly unknown. Here, we meta-analyzed three
25 migraine GWAS including 98,374 cases and 869,160 controls and identified 122
26 independent risk loci of which 35 were new. Fine-mapping of a meta-analysis is
27 challenging because some variants may be missing from some participating studies
28 and accurate linkage disequilibrium (LD) information of the variants is often not
29 available. Here, using the exact in-sample LD, we first investigated which statistics
30 could reliably capture the quality of fine-mapping when only reference LD was
31 available. We observed that the posterior expected number of causal variants best
32 distinguished between the high- and low-quality results. Next, we performed fine-
33 mapping for 102 autosomal risk regions using FINEMAP. We produced high-quality
34 fine-mapping for 93 regions and defined 181 distinct credible sets. Among the high-
35 quality credible sets were 7 variants with very high posterior inclusion probability
36 (PIP > 0.9) and 2 missense variants with PIP > 0.5 (rs6330 in *NGF* and rs1133400 in
37 *INPP5A*). For 35 association signals, we managed to narrow down the set of
38 potential risk variants to at most 5 variants.

39

40

41 Introduction

42

43 Migraine is a common neurological disorder characterized by recurrent disabling
44 episodes of severe headache that are typically one-sided, pulsating in nature, and

45 accompanied by other symptoms such as nausea, and hypersensitivity to light
46 and/or sound. It has two main subtypes, migraine without aura and migraine with
47 aura. The aura is a reversible visual, sensory or speech disturbance, that typically
48 occurs before the headache phase. Migraine attacks last usually from 4 to 72 hours,
49 and can significantly harm daily life of patients¹. Migraine was ranked as the second
50 most disabling disease worldwide in terms of years lived with disability by Global
51 Burden of Diseases Study in 2019². Its lifetime prevalence has been estimated to be
52 about 15 to 20 % worldwide, and it is three times more common in females than in
53 males². Family and twin studies estimate the heritability to be about 40%³. To date,
54 over 100 migraine associated loci have been reported by GWAS^{4,5,6,7,8,9,10,11,12,13,14}.
55 The genetic association of migraine has shown a general enrichment in genes highly
56 expressed in vascular and central nervous system related tissues^{15,13} but we lack
57 detailed information on specific genetic variants that affect the migraine risk.

58

59 Identification of causal genes and variants that have a biological effect on migraine is
60 crucial for understanding the biology of migraine, and for developing new effective
61 treatments for the disorder. Here, we aim to narrow down correlated genetic variation
62 in migraine associated regions to a smaller number of candidate causal variants by
63 applying statistical fine-mapping¹⁶. Fine-mapping methods evaluate how plausibly
64 each variant in the region is among the causal variants by utilizing the observed
65 association statistics and the LD structure of the region¹⁶. Multiple methods that can
66 utilize GWAS summary statistics have been developed, including PAINTOR¹⁷,
67 CAVIAR¹⁸, FINEMAP¹⁹, JAM²⁰ and SuSIE²¹. The optimal way to apply fine-mapping
68 is to compute the LD information from the original GWAS data (in-sample LD), but
69 when the original genotype data are unavailable, approximate LD information is often

70 obtained from a reference genotype panel (reference LD). However, when reference
71 LD is used, the discrepancy from the in-sample LD can cause errors in fine-mapping
72 and this problem becomes more severe as the GWAS sample size grows²².

73
74 Even though large meta-analyses have become a successful way to increase
75 statistical power of GWAS, they remain difficult to fine-map reliably for several
76 reasons²³. First, meta-analyses are combinations of multiple studies and typically no
77 single analyst has access to the exact in-sample LD of the whole meta-analysis,
78 which means that reference LD must be used. Second, differences in genotyping
79 platforms and genotype imputation pipelines between the meta-analyzed studies can
80 bias the fine-mapping results. Third, some variants included in the meta-analysis
81 may be present in only a subset of the studies, which leads to variation in information
82 content of the association statistics of different variants. In a landmark fine-mapping
83 study on schizophrenia, Trubetskoy et al. (2022)²⁴ avoided these problems by
84 collecting all genotype-phenotype data into a single analysis site. Unfortunately, to
85 our knowledge, no other international disease consortium has been able to create a
86 comparable analysis environment that would allow an in-sample fine-mapping of a
87 large meta-analysis. Given that fine-mapping of meta-analysis results typically relies
88 on reference LD, a crucial question is how we can assess when the results of fine-
89 mapping based on reference LD are reliable.

90
91 So far, the largest GWAS meta-analysis on migraine contained 102,084 cases and
92 771,257 controls from 25 study collections¹³. Unfortunately, we cannot perform
93 reliable fine-mapping for that meta-analysis, since the in-sample LD is not available.
94 Instead, we conducted a migraine meta-analysis with 98,374 migraine cases and

95 869,160 controls by combining data from three sources: 23andMe, Inc., FinnGen,
96 and UK Biobank (UKB). Of these data sets, 23andMe and UKB were included in the
97 earlier meta-analysis of Hautakangas et al. (2022) while FinnGen was not. Statistical
98 power of our meta-analysis was comparable to the previous migraine meta-analysis
99 of Hautakangas et al. (2022), with effective sample sizes of 339,000 and 326,000,
100 respectively. Importantly, we have the full in-sample LD available for 26 risk loci and
101 for the remaining risk loci we have the in-sample LD for FinnGen and UKB but not for
102 23andMe (Table 1). This set-up allowed us to investigate how different LD reference
103 panels perform compared to the in-sample LD. In particular, we evaluated different
104 statistics that could be used to assess fine-mapping quality when only reference LD
105 is available. Finally, we utilized our results to fine-map 102 migraine risk loci to
106 narrow down the putative causal variants behind the associations. We were able to
107 get reliable fine-mapping results for 93 out of 102 regions and identified 7 variants
108 with a high probability (>90%) of being causal and two missense variants, rs6330 in
109 *NGF* and rs1133400 in *INPP5A*, with a probability > 50% of being causal.

110

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

Study	Ancestry	Cases	Controls	N	Case %	Migraine definition	LD availability
UK Biobank	European, British	10,881	330,169	341,050	0.03	Self-reported	In-sample
23andMe, Inc	European descent	53,109	230,876	283,985	0.19	Self-reported	In-sample for 26/102 fine-map regions
FinnGen R8	European, Finnish	34,385	308,114	342,499	0.10	Medication purchases	In-sample
Meta-analysis	European descent	98,374	869,160	967,534	0.10	Self-reported, medication purchases	In-sample 26/102, reference LD 76/102

112

113

114 Results

115

116 We conducted an inverse-variance weighted meta-analysis on migraine by
117 combining results from the three GWAS (Table 1): UK Biobank (UKB; 10,881 cases
118 and 330,169 controls), 23andMe, Inc. (53,109 cases and 230,876 controls), and
119 FinnGen Release 8 (34,385 cases and 308,114 controls). The total sample size is
120 98,374 migraine cases and 869,160 controls. Before meta-analyzing the data, we
121 estimated pairwise genetic correlations between the study collections by LD Score
122 regression (LDSC)²⁵. The estimated genetic correlations were 1.00 (s.e. 0.04)
123 between UKB and 23andMe, 0.84 (s.e. 0.05) between UKB and FinnGen, and 0.87
124 (s.e. 0.03) between 23andMe and FinnGen. The lower genetic correlation between
125 FinnGen and the other two studies could be due to differences in the case definitions
126 (triptan purchases in FinnGen vs. self-reporting in UKB and 23andMe). A
127 comparable level of genetic correlation (0.81) has been reported before between
128 primary care and self-reported migraine cases within UKB²⁶. Another source of
129 possible heterogeneity in effect sizes is the difference in genetic ancestry (Finnish in
130 FinnGen vs. Non-Finnish European in the other two).

131

132 The genomic inflation factor (λ_{GC}) of the migraine meta-analysis was 1.38. There was
133 a linear relationship between the association statistic and the LD-score
134 (Supplementary Fig 1) indicating that the polygenic background of migraine was the
135 main source of the genomic inflation. However, as the intercept from LDSC was
136 elevated to 1.09 (s.e. 0.01) from its null value of 1.0, some inflation could also be due
137 to confounding factors such as cryptic relatedness, population stratification or other
138 model misspecification. Consequently, we further checked the LDSC intercepts for

139 the individual studies: 1.03 (s.e. 0.01) for 23andMe, 1.00 (s.e. 0.01) for UKB and
140 1.10 (s.e. 0.01) for FinnGen. The higher intercept for FinnGen could be due to a
141 different GWAS analysis method (whole genome-regression by REGENIE²⁷
142 including related samples) compared to UKB and 23andMe (logistic regression
143 excluding related samples). Estimated SNP-heritability was 11.49% (s.e. 0.47%)
144 from LDSC when population prevalence was assumed to be 16%.

145

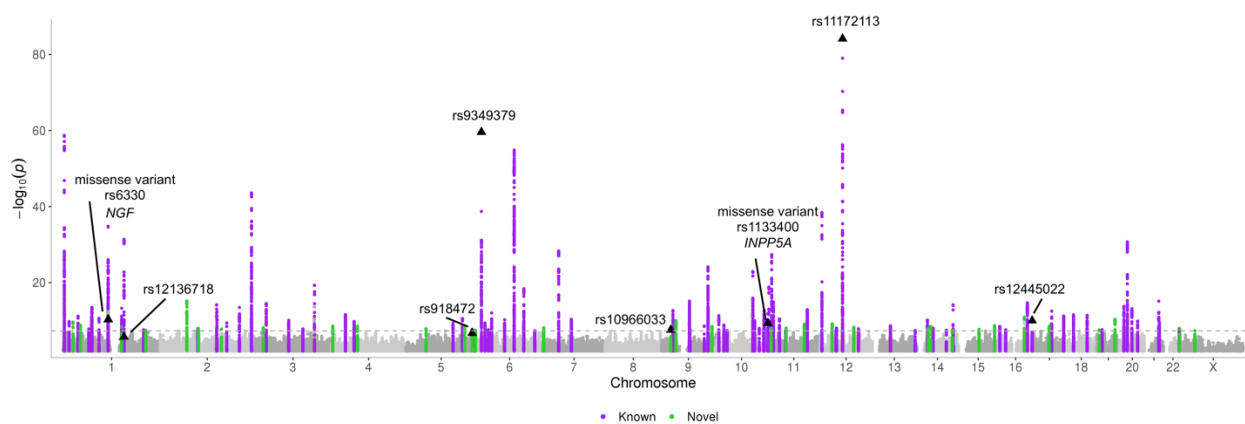
146 We followed the locus definition of Hautakangas et al. (2022) and defined the LD-
147 independent genome-wide significant (GWS; $P < 5 \times 10^{-8}$) risk loci from the meta-
148 analysis iteratively by choosing the variant with the smallest P-value as an index
149 variant and excluding all other GWS variants with LD $r^2 > 0.1$ to that index variant
150 from further considerations until no GWS variants remained. Next, we formed a high
151 LD region around each index variant extending to the level of $r^2 > 0.6$, and merged
152 regions that were closer than 250 kb. Lastly, all other GWS variants were included in
153 their closest region, and the region boundaries were updated, and once again
154 regions closer than 250 kb were merged (see further details in Methods). Based on
155 this locus definition, we identified 122 LD-independent risk loci, of which 35 were
156 new (Table 2), and 87 overlapped with the previously known risk loci (Fig 1,
157 Supplementary Table 1, Supplementary Figs 2-4)^{4,5,6,7,8,9,10,11,12,13,14}.

158 We observed statistically significant heterogeneity ($P < 0.05/122$) in effect sizes
159 between the study collections only for two lead variants, both of which resided in the
160 previously known migraine loci (PRDM16 and near ZCCHC14)(Supplementary Table
161 1, Supplementary Fig 3). As external replication data of 34,807 cases and 193,475
162 controls, we meta-analyzed data from the Trøndelag Health Study (HUNT)²⁸ and
163 IHGC16 migraine meta-analysis excluding the Finnish cohorts and the 23andMe

164 data⁹. Of the 35 lead variants of our new loci, 32 were consistent in direction ($P = 2.1$
165 $\times 10^{-7}$, one-sided binomial test) and 17 replicated with $P < 0.05$ (one-sided test;
166 Supplementary Table 2) in the replication data. When we meta-analyzed the
167 discovery and the replication data, 28 out of the 35 novel loci remained GWS
168 (Supplementary Table 2).

169
170 To define the fine-map regions, we merged together the risk loci that were closer
171 than 1.5 Mb. This resulted in 102 fine-map regions. To avoid problems due to
172 varying sample sizes across the variants, we included in fine-mapping only
173 autosomal SNPs that were available in all three cohorts. This criterion reduced the
174 number of common variants ($MAF > 0.05$) per regions on average by 19%.

175
176 Figure 1. A Manhattan plot of the inverse-variance weighted fixed effects migraine
177 meta-analysis including 98,374 cases and 869,160 controls. X-axis presents the
178 chromosomal location and y-axis the $-\log_{10}(P\text{-value})$. Known loci are highlighted in
179 purple and new loci in green. Variants with posterior inclusion probability (PIP) > 0.9
180 and missense variants with PIP > 0.5 in high-quality fine-mapping regions are
181 annotated.



182

183

184 Table 2. New 35 migraine risk loci identified from the meta-analysis of 98,374
185 migraine cases and 869,160 controls.

Locus name	RSID	Chromosome	Position GRCh37	Effect allele	Other allele	Effect allele frequency	Log-odds ratio	S.e.	P-value
near RUNX3	rs71014329	1	25348950	I	D	0.604	0.034	0.005	2.57E-10
ST3GAL3	rs783302	1	44366341	G	A	0.878	0.047	0.008	1.68E-09
SF3B4	rs7544531	1	149897217	T	C	0.084	0.072	0.012	5.08E-09
near DTL	rs61830764	1	212289976	A	G	0.382	0.031	0.006	3.71E-08
near APLF	rs112706954	2	68819969	G	A	0.023	0.137	0.017	7.88E-16
TMEM131	rs2305142	2	98375722	G	A	0.322	0.031	0.005	1.18E-08
near GPD2	rs74482068	2	157560108	D	I	0.039	0.076	0.014	1.76E-08
near RANP7	rs11386839	3	22929430	D	I	0.500	0.029	0.005	7.68E-09
ADD1	rs10026792	4	2862190	G	A	0.687	0.032	0.005	2.79E-09
EPHA5	rs147908403	4	66362482	C	T	0.054	0.069	0.012	2.80E-09
ITGA1	rs4865540	5	52184268	C	A	0.820	0.037	0.007	1.41E-08
near GLRA1	rs372257780	5	151200938	I	D	0.599	0.033	0.006	2.27E-09
KCNIP1	rs78151838	5	170108683	A	G	0.905	0.054	0.010	1.82E-08
MAML1	rs10794701	5	179181061	A	G	0.119	0.043	0.008	3.57E-08
near COX19	rs117303395	7	1001963	A	G	0.019	0.122	0.022	4.40E-08
MAD1L1	rs10479762	7	2045351	T	C	0.419	0.029	0.005	8.01E-09
ELAVL2	rs10966033	9	23705736	G	T	0.617	0.029	0.005	2.70E-08
near ZCCHC7	rs10973207	9	37100525	T	G	0.187	0.042	0.007	1.04E-10
near LMX1B	rs4358894	9	129464802	C	G	0.513	0.030	0.005	3.33E-09
near DENND5A	rs34494849	11	9287030	C	T	0.768	0.034	0.006	1.17E-08
near MTCH2	rs11039324	11	47665686	G	A	0.601	0.030	0.005	9.76E-09
MRE11A	rs639311	11	94205747	C	T	0.681	0.033	0.005	9.02E-10
IPO8	rs12369125	12	30807195	A	C	0.251	0.036	0.006	7.08E-10
MGAT4C	rs73187675	12	86409247	T	A	0.193	0.037	0.006	6.08E-09
RP11-562L8.1	rs1957110	14	29777492	T	C	0.409	0.029	0.005	1.59E-08
INSM2	rs2296919	14	36005659	T	C	0.807	0.038	0.006	3.44E-09
RPS6KA5	rs117151272	14	91415550	A	T	0.026	0.097	0.018	3.59E-08
near ONECUT1	rs1899730	15	53166138	T	G	0.707	0.032	0.006	2.11E-08
FAM174B	rs12910861	15	93218540	C	T	0.227	0.037	0.006	2.15E-09
FAM65A	rs9934328	16	67573367	C	G	0.137	0.049	0.007	1.32E-11
TUBG2	rs2292750	17	40811781	C	T	0.452	0.030	0.005	3.53E-09
near NRTN	rs76899991	19	5822370	G	T	0.963	0.077	0.014	2.89E-08
SYMPK	rs74821481	19	46320041	G	T	0.678	0.036	0.005	4.59E-11
near SERHL2	rs141478056	22	42939927	G	A	0.120	0.046	0.008	2.23E-08
near FTHL17	rs149675702	23	31063624	C	T	0.945	0.079	0.014	4.56E-08

186 RSID = reference SNP ID, GRCh37 = Genome Reference Consortium Human Build
187 37, s.e. = standard error. Alleles D and I refer to deletion and insertion, respectively.

188

189 Comparison of different LD panels in fine-mapping

190

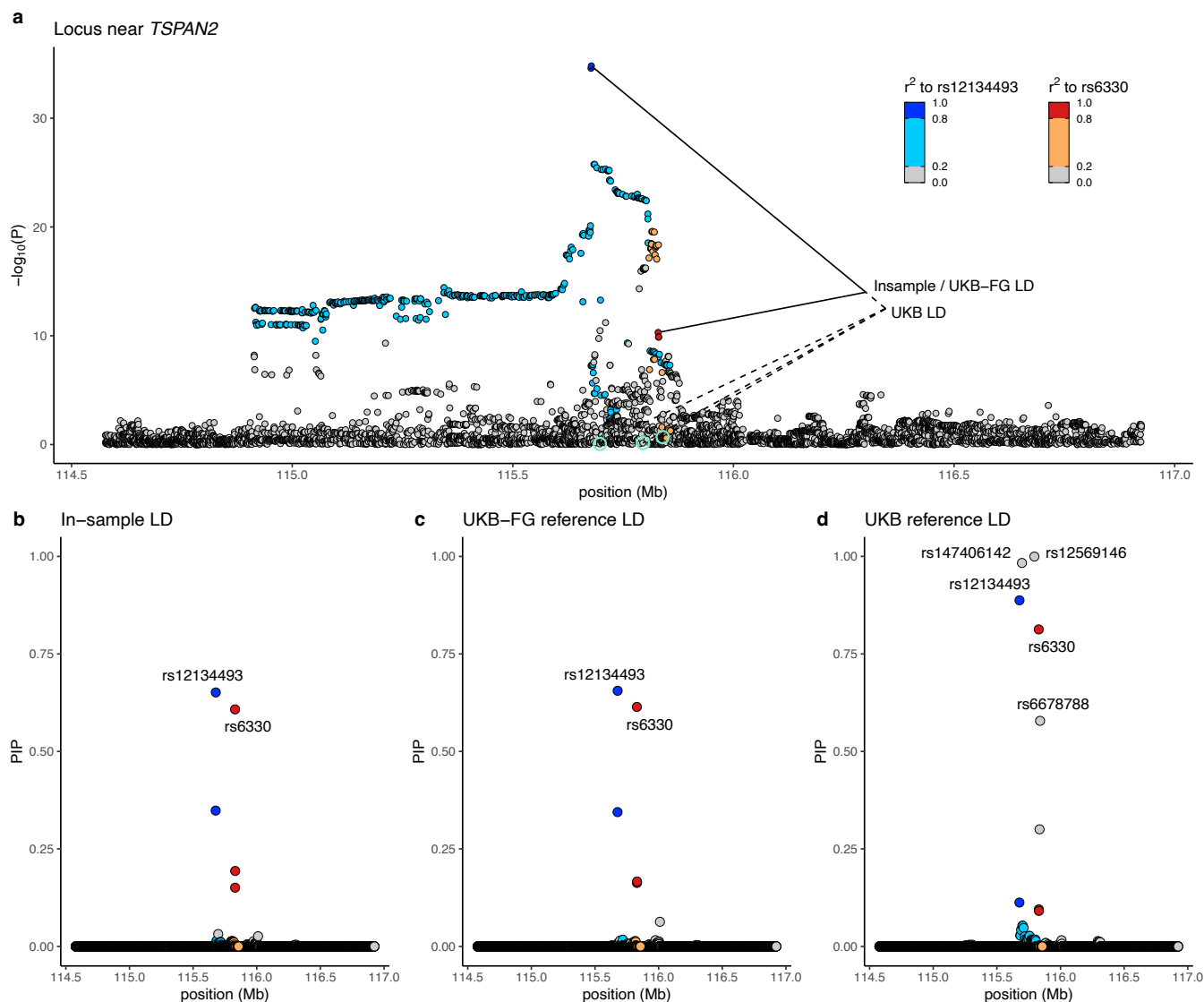
191 A common problem in meta-analyses is that the in-sample LD is not available, and
192 use of reference LD may lead to biased results. Figure 2 demonstrates this problem
193 at the locus around *TSPAN2* where fine-mapping using the in-sample LD disagrees
194 strongly with the UKB reference LD but agrees well with a more accurate UKB-FG
195 reference LD. This shows that, in our setting, fine-mapping based on the UKB-FG
196 reference LD has a potential to yield reliable results but that we need some way to
197 assess, for each region, whether the reference LD has provided reliable results.

198 Therefore, we evaluated whether some statistics, either derived from the GWAS
199 results or from the fine-mapping results, could flag the regions where the reference
200 LD produced unreliable fine-mapping results compared to the in-sample LD. We did
201 this comparison in the 26 regions where the in-sample LD was available. As
202 candidate statistics, we considered: (1) posterior expectation of the number of causal
203 variants (PENC), and, from the top variant(s) of the credible sets, (2) maximum
204 pairwise r^2 , (3) maximum marginal P -value, and (4) minimum INFO value. We used
205 the maximum difference of the variant-specific posterior inclusion probabilities
206 ($\max\Delta$) between the reference LD and the in-sample LD to assess the quality of the
207 reference LD results. A small $\max\Delta$ value (close to 0) indicates high quality (the
208 reference LD produces similar results to the in-sample LD), and a large value (close
209 to 1) indicates low quality (the reference LD produces different results from the in-
210 sample LD).

211 In general, both LD reference panels performed well in most of the 26 regions
212 available for this comparison, but, as expected²², the more accurate UKB-FG panel
213 performed clearly better than the UKB panel alone. For example, $\max\Delta$ was above
214 0.1 only in 2/26 regions with the UKB-FG panel but in 8/26 regions with the UKB
215 panel (Fig 3a).

216

217 Figure 2. Fine-mapping a region near *TSPAN2* at chromosome 1 using three
218 different LD sources. a) Plot of the GWAS results with the chromosomal location on
219 x-axis and the strength of the association as $-\log_{10} P$ -values from the inverse-
220 variance weighted fixed-effect meta-analysis with 98,374 migraine cases and
221 869,160 controls on y-axis. Variants are colored based on the squared correlation
222 (r^2) to the two variants in the top configuration suggested by FINEMAP with the in-
223 sample LD. The suggested top configurations based on three LD panels are marked
224 by lines with the in-sample LD and the UKB-FG reference LD giving the same top
225 configuration and the UKB reference LD including three additional variants
226 (highlighted in green). Posterior inclusion probabilities (PIPs) for the variants based
227 on b) in-sample LD, c) UKB-FG reference LD and d) UKB reference LD.



228

229

230 We then investigated how well the four different statistics could separate the regions

231 with low-quality fine-mapping results from those with high-quality results for the two

232 LD reference panels (Supplementary Fig 5). First, when PENC was used, both LD

233 reference panels performed similarly for the regions where FINEMAP suggested only

234 one or two causal variants (Supplementary Fig 5a). Those results were also close to

235 the in-sample results ($\max\Delta < 0.07$). All low-quality regions (with $\max\Delta > 0.1$) had

236 $\text{PENC} > 2$ with the UKB panel and $\text{PENC} > 3$ with the UKB-FG panel. Thus, we used

237 these PENC thresholds to define low-quality regions when the in-sample LD was not

238 available. We expect that these thresholds have a high sensitivity for low-quality

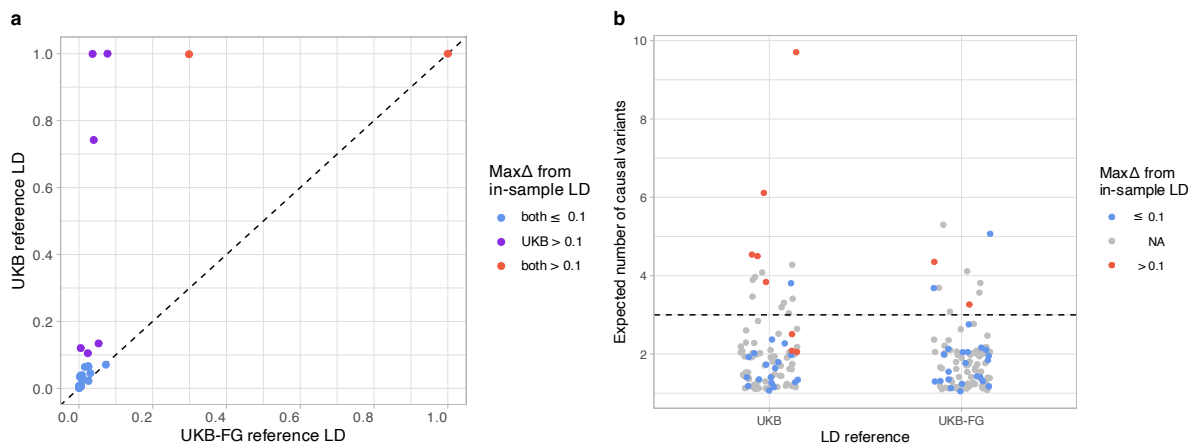
239 results but will simultaneously exclude some of the regions that truly have many
240 causal variants. The other three statistics are not able to distinguish the low-quality
241 regions as clearly as PENC (Supplementary Figs 5b-d). First, the maximum r^2
242 among the top configuration variants does not distinguish both of the low-quality
243 regions with the UKB-FG panel (Supplementary Fig 5b). Additionally, neither the
244 maximum P nor the minimum INFO within the top credible set variants separates
245 well the low-quality regions from the good-quality regions (Supplementary Figs 5c,d).
246 We conclude that PENC gives the best separation among the statistics investigated.
247 Previously, PENC has been used to filter FINEMAP results in the schizophrenia fine-
248 mapping study²⁴.

249

250 Next, we evaluated how PENC classifies the 76 fine-map regions where only
251 reference LD was available to us. The 76 grey points in Figure 3b show that the fine-
252 map regions without the in-sample LD are typically having $PENC < 2.5$ and, with the
253 UKB-FG LD, only 6 of the 76 regions have $PENC > 3$.

254

255 Figure 3. a) Scatter plot comparing the maximum PIP differences ($\max\Delta$) between
256 the in-sample and reference LD for 26 fine-map regions. X-axis shows the UKB-FG
257 reference LD and y-axis the UKB reference LD. b) Strip chart shows the posterior
258 expected number of causal variants (PENC) from fine-mapping for the two LD
259 reference panels for the 102 fine-map regions. Red dots indicate large differences
260 from the in-sample LD ($\max\Delta > 0.1$), and grey color indicates regions for which only
261 reference LD is available and therefore $\max\Delta$ is not known. Horizontal line shows
262 $PENC = 3$ that we use as a threshold to define reliable results with the UKB-FG
263 panel.



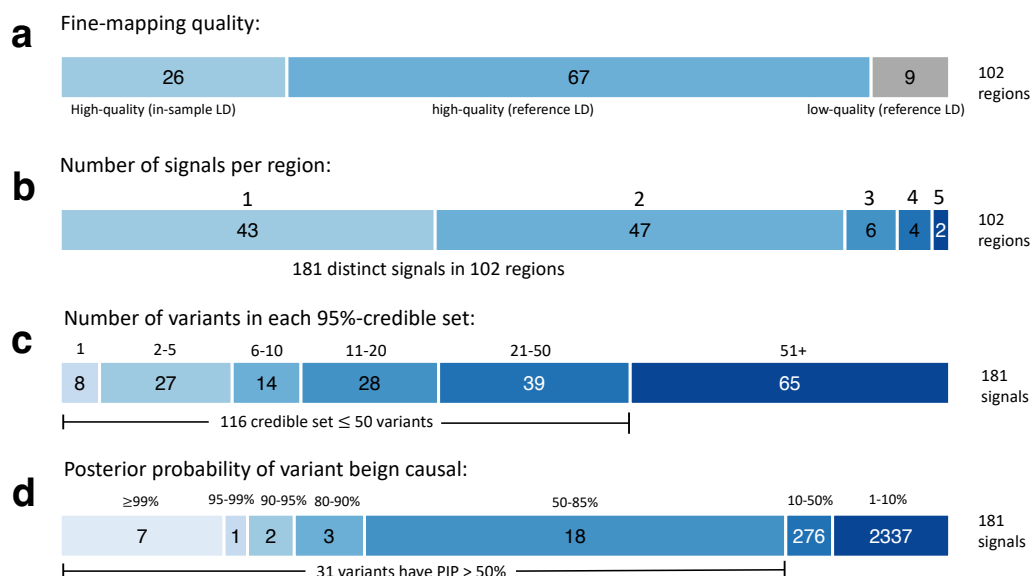
264

265

266 FINEMAP results overview

267

268 Figure 4. Summary of the fine-mapping results across the 102 migraine risk regions.



269

270 Overall, for a majority of the fine-map regions, FINEMAP suggested one (42%) or

271 two (46%) causal variants (Supplementary Table 3, Fig 4.). The 102 fine-map

272 regions together had 181 distinct signals when the signals were defined by the

273 number of causal variants per region with the highest posterior probability. Among

274 the 76 regions without the in-sample LD, 6 had PENC above 3. We flagged these

275 regions to be of low-quality, and their interpretation requires extra caution. The
276 largest PENC observed was 5 and it occurred for two fine-map regions: *PRDM16*
277 (index variant rs10218452) and *HOXB3* (index variant rs2555111). Of these, *HOXB3*
278 region is flagged as low-quality because there is no in-sample LD available.
279 The sizes of 95%-credible sets ranged from 1 to 2,787 variants, and 49 credible sets
280 had 10 variants or less. A very high PIP (≥ 0.9) was observed for 10 variants
281 (Supplementary Table 4), of which seven were in the high-quality fine-map regions
282 (Table 3). We conducted a look-up from Variant Effect Predictor (VEP) database for
283 all credible sets to search for variants that could have an impact on the gene
284 transcript. In total, 149 unique missense variants were found of which 3 had PIP >
285 0.5: rs6330 (PIP=0.59) in *NGF* located at chromosome 1, rs1133400 (PIP=0.93) in
286 *INPP5A* located at chromosome 10 and rs28929474 (PIP=0.64) in *SERPINA1*
287 located in a low-quality fine-map region at chromosome 14 (Table 3, Supplementary
288 Table 5). Of these, rs6330 is a significant *cis*-eQTL for *NGF-AS1* expressed in atrial
289 appendage of heart and rs28929474 for *IFI27L2* expressed in tibial artery and in left
290 ventricle of heart in GTEx v.08 data.
291 *NGF* encodes protein nerve growth factor beta ($\text{NGH}\beta$) that is important in the
292 development and survival of neurons, and involved in transmission of pain,
293 temperature, and touch sensations via sensory neurons. It binds to two receptors,
294 NTRK1 encoded by *NTRK1* and NGFR/p75^{NTR} encoded by *NGFR*. Of note, two
295 additional missense variants among the credible sets, rs6339 (PIP= 0.48) and
296 rs6336 (PIP=0.39), are located in *NTRK1* in a separate locus. The missense variant
297 rs6330 shows association with multiple diseases of the musculoskeletal system and
298 connective tissue including spinal stenosis, spondylosis, spondylopathies and hallux

299 valgus in FinnGen R10 PheWAS scan, all to the opposite direction compared to the
300 migraine risk (Supplementary Table 6).

301 *INPP5A* encodes a membrane-associated type I inositol 1,4,5-trisphosphate 5-
302 phosphate protein, which hydrolyzes Ins(1,4,5)P₃ leading to the mobilization of
303 intracellular calcium. It has a central role in various cellular signaling processes
304 including neurotransmission, hormone secretion, cell proliferation and muscle
305 contraction. *INPP5A* is highly expressed in Purkinje cells of cerebellum, and in mice
306 studies its deletion have been shown to cause ataxia and cerebellar
307 degeneration^{29,30}.

308 *SERPINA1* encodes an alpha-1 antitrypsin, a serine protease inhibitor protein, that
309 belongs to the serpin superfamily. Its primary target is elastase, and other targets are
310 plasmin and thrombin. Several mutations, including our high-PIP variant
311 rs28929474C>T, in *SERPINA1* can cause an autosomal co-dominant genetic
312 disorder alpha-1 antitrypsin (AAT) deficiency, which can lead to lung or liver disease
313 due to reduced alpha-1 antitrypsin levels³¹. A missense variant rs28929474 is highly
314 pleiotropic and shows associations to multiple disease categories in PheWAS of
315 FinnGen R10 data including, for example, diseases of the respiratory system,
316 diseases of the circulatory system, diseases of digestive system, pregnancy related
317 diseases, diseases of the nervous system, and diseases of musculoskeletal system
318 and connective tissue (Supplementary Tables 6-8).

319 Five additional high-impact variants on protein function (1 stop gained, 2 start lost,
320 and 2 splice acceptor variants) were among the credible sets, but only with modest
321 PIPs below 0.01 (Supplementary Table 5), and another 5 variants with high-impact
322 on something else than protein coding function (long non-coding RNA, antisense or
323 nonsense mediated decay) with PIPs below 0.02.

324 Our results provided new information on two of the strongest known migraine risk
325 loci by estimating PIPs of 1.00 for the intronic variants rs9349379 in *PHACTR1* and
326 rs11172113 in *LRP1*. We were able to fine-map both of these loci by using the in-
327 sample LD. The candidate variant in *PHACTR1* is also associated with many
328 vascular diseases and its effects on gene expression of the genes in the locus have
329 been studied in detail but with contradicting results^{32,33}. Also, the candidate variant in
330 *LRP1* is associated with several vascular diseases, such as sporadic thoracic aortic
331 dissection, fibromuscular dysplasia and spontaneous coronary artery
332 dissection^{34,35,36}. The LDL receptor-related protein 1 (LRP1) is a cell surface receptor
333 and has an important role in vascular and blood brain barrier integrity^{37,38,39}. It is
334 expressed in almost every tissue, and most studied in liver and brain. LRP1 is also
335 involved in vascular calcium signaling by regulating smooth muscle cell
336 contractility³⁸. A recent study suggested that *LRP1* expression is regulated by allele-
337 specific mechanism of intronic rs11172113 located in an enhancer region through
338 two transcription factors (MECP2 and SNAIL)⁴⁰.
339
340

341 Table 3. Variants with high (>0.9) posterior inclusion probability (PIP) and missense
 342 variants with PIP > 0.5 among the 93 high-confidence fine-map regions.

Gene (VEP)	Predicted consequence (VEP)	RSID	Chromosome	Position GRCh37	Effect allele	Other allele	PIP	Minor allele frequency	Log-odds ratio	S.E.	P-value	LDsource
PHACTR1	Intron variant	rs9349379	6	12903957	G	A	1.000	0.422	-0.084	0.005	2.59E-60	in-sample
LRP1	Intron variant	rs11172113	12	57527283	C	T	1.000	0.404	-0.101	0.005	7.27E-85	in-sample
-	Intergenic variant	rs12445022	16	87575332	A	G	1.000	0.333	-0.035	0.005	1.04E-10	in-sample
-	Intergenic variant	rs12136718	1	156409585	A	G	0.999	0.072	0.046	0.010	1.95E-06	in-sample
ELAVL2	Intron variant	rs10966033	9	23705736	T	G	0.954	0.383	-0.029	0.005	2.70E-08	UKB-FG
TLX3	3' UTR variant	rs918472	5	170738836	G	A	0.932	0.292	-0.029	0.006	1.95E-07	UKB-FG
INPP5A	missense variant	rs1133400	10	134459388	G	A	0.926	0.198	0.039	0.006	5.06E-10	UKB-FG
NGF	missense variant	rs6330	1	115829313	A	G	0.593	0.461	0.033	0.005	4.97E-11	in-sample

343
 344 Due to the restriction of including in fine-mapping only the variants that are available
 345 in all three data sets, the original lead variant was missing in 17/102 fine-map
 346 regions (Supplementary Table 3b). In 14/17 of these regions, the original lead variant
 347 was represented by one of the top credible set variants (defined as being in LD with
 348 $r^2 > 0.1$ in the UKB data). For the remaining 3 regions, the signal related to the
 349 original lead variant may be missing from the fine-mapping results, and we flagged
 350 these regions to be of low-quality. Among the fine-map regions for which the lead
 351 variant was included in the analysis, the lead variant was within the 95% credible
 352 sets in 83/85 fine-map regions and within the top configuration in 73/85 of the
 353 regions.

354
 355 Phenome-wide association scans for the credible set variants
 356
 357 We conducted three separate phenome-wide association studies (PheWAS) by
 358 using data from FinnGen Data Freeze 10 including 429,209 individuals. First, by a

359 PheWAS for the 181 credible set top variants and the list of 2,399 FinnGen
360 endpoints excluding the migraine endpoints, we identified 404 variant-disease
361 associations with $P < 1 \times 10^{-5}$ (Supplementary Table 6, phewas_app). Of these, 108
362 variant-disease associations belonged to diseases of the circulatory system,
363 including, for example, hypertension and ischemic heart disease, followed by 39
364 variant-trait associations in a category of quantitative endpoints, including, e.g.,
365 height and BMI, 34 in diseases of the musculoskeletal system and connective tissue
366 category, including, e.g., spinal stenosis and rheumatoid arthritis, and 28
367 associations in diseases of the respiratory system, including, e.g., asthma and
368 COPD.

369 Second, for the 159 functional variants among the credible sets, we conducted a
370 targeted PheWAS scan within neurological and cardiovascular endpoints, and
371 identified 122 variant-disease associations with $P < 1 \times 10^{-4}$ (Supplementary Table
372 7, phewas_app), including traits such as sleep apnea and stroke. Third, for the 307
373 variants with PIP > 0.1, with a similar targeted PheWAS scan within the neurological
374 and cardiovascular endpoints, we identified 330 variant-disease associations with P
375 < 1×10^{-4} (Supplementary Table 8, phewas_app), including, e.g., focal epilepsy and
376 hydrocephalus.

377

378

379 Discussion

380

381 Well over one hundred risk loci for migraine have been reported from GWAS, but the
382 causal variants and genes are still mostly unknown^{4,5,6,7,8,9,10,11,12,13,14}. Statistical fine-
383 mapping of the GWAS results at the risk loci is a natural next step but reliable fine-

384 mapping of large meta-analysis data has turned out to be very difficult. Our recent
385 migraine meta-analysis of 25 studies¹³ illustrated these difficulties as the accurate LD
386 information was not available and the sample size varied considerably across
387 variants. In this study, our goal was to provide reliable fine-mapping for migraine by
388 creating a new migraine meta-analysis for which accurate LD information was
389 available and sample size across variants was more stable. Despite the more
390 stringent selection criteria, the effective sample size of our new meta-analysis
391 (339,000) turned out to be comparable to that of the earlier meta-analysis (326,000).

392

393 A key question in fine-mapping a GWAS meta-analysis is how to assess the
394 reliability of the results. We were able to study this question by directly comparing
395 results between accurate in-sample LD and approximate reference panel LD. We
396 observed that the posterior expected number of causal variants (PENC) as reported
397 by FINEMAP distinguished well the regions with high-quality fine-mapping results
398 from those with low-quality results. We also observed that an appropriate PENC
399 threshold depends on the quality of the reference panel. In our case, we were able to
400 use an upper limit of 3.0 for PENC. While this upper limit restricts our ability to fine-
401 map the migraine risk regions that truly have more than 3 causal signals, we expect
402 that the proportion of such regions is small, as only 3/26 (12%) of the migraine loci
403 with the in-sample LD had PENC over 3 in our analysis.

404

405 Here, we performed the first systematic fine-mapping of a migraine meta-analysis
406 and provided high-quality fine-mapping results for 91% of the migraine risk regions
407 identified by the meta-analysis. Our high-quality results highlight two missense

408 variants with high PIPs: rs6330 (PIP=0.59) in *NGF* and rs1133400 (PIP=0.93) in
409 *INPP5A*.

410 The variant rs6330 is only in weak LD ($r^2 = 0.04$) with the lead variant (rs12134493)
411 of its locus and was identified as a secondary signal in our fine-mapping. A recent
412 study¹⁴ has also reported that the migraine association of rs6330 remained
413 statistically significant in a conditional analysis after adjusting for the stronger signal
414 (rs2078371) within the same risk locus. *NGF* has been reported to be highly
415 expressed in hippocampus and cortex^{41,42} although according to the GTEx v8 data,
416 *NGF* does not show statistically significant expression in any brain tissue but shows
417 high expression in multiple other tissues, including, for example, ovary, tibial nerve,
418 arteries, visceral adipose, and heart. NGF levels have been reported to be elevated
419 in cerebrospinal fluid in chronic migraine patients compared to controls⁴³, and
420 decreased in blood serum of episodic migraine patients compared to controls and
421 chronic migraine patients⁴⁴. In addition, we observed two additional missense
422 variants with considerable PIPs, rs6339 (PIP=0.48) and rs6336 (PIP=0.39), located
423 in *NTRK1* which encodes one of the two receptors for NGF. NGF and its receptors
424 have a central role in the pain perception, and elevated NGF levels have been
425 observed also in many other chronic pain conditions, such as osteoarthritis and low
426 back pain^{45,46,47}. Multiple antibodies of NGF or small molecular inhibitors of the NGF
427 receptors have been developed and tested in clinical studies to treat chronic pain
428 conditions, including low back pain and osteoarthritis^{48,49,50,51,52}. Even though some
429 candidate drugs have shown potential benefit relating to pain relief, an increased risk
430 of progressive osteoarthritis has been observed in a small group of the treated
431 patients⁵², and therefore none of the drugs have yet received FDA approval.
432 Currently, other type of drug classes (p75 neurotrophin receptor fusion protein, LEVI-

433 04 ([ClinicalTrials.gov ID: NCT05618782](#)) and anti-NGF PEGylated Fab' antibody⁵³),
434 are being developed and in pre-clinical or clinical testing. In adults, after pain stimuli,
435 NGF activates overexpression of other neuronal molecules, including calcitonin
436 gene-related peptide (CGRP) and substance P⁵². CGRP is involved in migraine pain,
437 and several effective monoclonal antibodies targeting either CGRP or its receptors
438 have been developed to treat migraine^{54,55,56}.

439 Gene *INPP5A* is highly expressed in Purkinje cells of cerebellum⁵⁷ and involved in
440 multiple cellular signaling processes including neurotransmission, hormone
441 secretion, cell proliferation and muscle contraction through its role in the pathway
442 regulating intracellular calcium levels. The missense variant rs1133400 is in modest
443 LD ($r^2 = 0.36$) with the lead variant of the locus (rs200314499) that was filtered out
444 from fine-mapping due to QC. For this locus, FINEMAP suggested two causal
445 variants (PENC = 1.65). PheWAS showed no other significant associations with this
446 missense variant.

447
448 Another important finding is in the *PHACTR1* locus, which is one of the strongest
449 known migraine risk loci. There our fine-mapping suggested one causal variant
450 (PENC = 1.29), with the lead variant rs9349379 being a clear candidate for being
451 causal with PIP of 1.00. In our FinnGen PheWAS, we detected also strong
452 associations between the variant and, for example, major coronary disease events
453 ($P = 8.22 \times 10^{-52}$), ischemic heart disease ($P = 1.18 \times 10^{-38}$) and angina pectoris (P
454 $= 7.71 \times 10^{-26}$), all to the opposite directions compared to migraine risk. Because of
455 these well-known associations with multiple vascular diseases, this locus has been
456 previously studied in detail but with contradicting results. Gupta et al. (2017)³²
457 reported that rs9349379 regulates upstream gene *EDN1*, whereas Wang et al.

458 (2018)³³ reported that they failed to replicate this endothelial rs9349379-EDN1 eQTL,
459 but instead showed that rs9349379 regulates the closest gene *PHACTR1*,
460 confirming previous vascular rs9349379-PHACTR1 eQTLs. Further, Rubin et al
461 (2022)⁵⁸ observed that a loss of *PHACTR1* gene does not seem to have any effect
462 on the endothelial or smooth muscle cells of the transgenic mice, and suggested that
463 *PHACTR1* has no contribution to pathological vascular phenotype in mice through
464 cells involved in vascular physiology. Our fine-mapping has provided strong
465 evidence that the lead variant rs9349379 is causal for migraine, but given that the
466 variant is intronic, our fine-mapping results alone do not provide direct evidence
467 through which gene or mechanism this association affects the disease risk.

468

469 Our study has some limitations. First, since reliable fine-mapping requires that we
470 exclude variants that are not present in all three component studies of our meta-
471 analysis, it is possible that we exclude also some of the true causal variants. This is
472 a potential problem especially when some of the top variants of the fine-map region
473 have been filtered out from fine-mapping. To identify the regions that are likely to be
474 affected by this problem, we studied the LD patterns between the fine-mapped
475 variants and those top variants from the fine-map regions that were not included in
476 the fine-mapping analysis. For most (14/17) regions where the top variants were
477 missing from fine-mapping, the signal of the top variant was at least partly
478 represented by another variant in LD with the top variant. Additionally, since very
479 rare variants were not included in our analysis, we miss the true causal variants that
480 are rare. Since our variant set is not comprehensive, we must keep in mind that also
481 variants that have a very high probability of being causal in our analysis may still
482 have such variants in high LD that were not included in our analysis. A valid

483 calibration of the PIPs would require that all potential causal variants were included
484 in the analysis. In practice, for common variants, this would require comprehensively
485 imputed data sets with no missing variants in any of the meta-analyzed studies, and,
486 for rare variants, availability of high coverage sequencing data. Currently, we do not
487 yet have such resources available in typical GWAS meta-analyses of common
488 diseases such as migraine.

489 Another limitation of our study relates to the phenotype definitions of different
490 substudies. First, both the UKB and 23andMe GWAS are based on self-reported
491 migraine status, and therefore some other conditions, such as tension headache,
492 may have been wrongly reported as migraine for some cases. Second, the FinnGen
493 GWAS is based on triptan purchase data, which may represent a specific subset of
494 migraine patients. Triptans are not suitable for all migraineurs and, especially, they
495 are contraindicated in patients with cardiovascular diseases. Overrepresentation of
496 migraineurs without any cardiovascular diseases could lead some FinnGen PheWAS
497 associations where migraine risk alleles seem to have protective effect on
498 cardiovascular phenotypes. Observational studies have reported that both migraine
499 and cardiovascular disease risk in women are positively associated⁵⁹.

500
501 To conclude, we performed a migraine GWAS meta-analysis with 98,375 migraine
502 cases and 869,159 controls and identified 122 risk loci of which 35 were new. We
503 followed up the meta-analysis by the first systematic fine-mapping analysis of
504 migraine risk loci and identified 7 variants with a high probability of being causal. In
505 addition to providing new information about genetic risk of migraine, we also
506 proposed how one could, in general, evaluate whether the fine-mapping results of
507 each risk loci seem reliable based only on the output from the fine-mapping software

508 FINEMAP. While a definitive fine-mapping analyses will require more comprehensive
509 data than are currently available for the GWAS meta-analyses of common diseases,
510 our study shows how reliable and novel fine-mapping results can be extracted
511 already from the currently available data sets by a suitable analysis approach.

512 Methods

513

514 Data

515

516 We performed a new migraine meta-analysis by combining summary statistics from
517 three migraine GWAS: UK Biobank (N= 341,050, 10,881 cases and 330,169
518 controls), 23andMe (N=283,985, 53,109 cases and 230,876 controls), and FinnGen
519 R8 (N= 342,499, 34,385 cases and 308,114 controls). By meta-analyzing the three
520 studies, the total sample size was 967,534 including 98,375 migraine cases and
521 869,159 controls.

522

523 UK Biobank: The UK Biobank project is a population-based prospective cohort study
524 that consists of over 500,000 participants aged 40-69 at recruitment collected from
525 several regions across the United Kingdom. The participants completed
526 questionnaires and attended interviews and clinical examinations by a trained staff
527 member. A detailed description of UK Biobank is provided elsewhere⁶⁰, and detailed
528 genotyping, quality control and imputation procedures are described at the UK
529 Biobank website (<https://www.ukbiobank.ac.uk/>). We used the migraine GWAS data
530 described in¹³ with self-reported migraine as the phenotype. UK Biobank received
531 ethical approval from the North West Multi-centre Research Ethics Committee
532 (MREC) and informed consent has been obtained from all participants.

533

534 23andMe: 23andMe migraine GWAS was performed by a personal genomics
535 company 23andMe, Inc. (<https://www.23andme.com/>) and detailed description of the
536 migraine GWAS is provided elsewhere⁸. All participants have provided informed

537 consent and filled an online survey according to 23andMe's human subjects
538 protocol, which was reviewed and approved by Ethical & Independent Review
539 Services, a private institutional review board. Briefly, migraine cases were assessed
540 from the participants that had reported migraine or answered "Yes" to any of the
541 questions related to migraine, and controls from participants that did not report
542 having migraine or answered "No" to all of the questions related to migraine,
543 excluding participants with discordant answers.

544

545 FinnGen: FinnGen (<https://www.finngen.fi/en>) is a large biobank study that has
546 collected and genotyped 500,000 Finns and combined these data with longitudinal
547 registry data including The National Hospital Discharge Registry, Causes of Death
548 Registry and medication reimbursement registries, all of these linked by unique
549 national personal identification codes. FinnGen includes prospective and
550 retrospective epidemiological and disease-based cohorts and hospital biobank
551 samples. A detailed description of FinnGen is provided in⁶¹. We used the 8th Data
552 Freeze for the migraine GWAS. The migraine cases were defined as the individuals
553 who had at least one triptan purchase and the remaining individuals without any
554 triptan purchases were defined as controls from the social insurance institution of
555 Finland (KELA) registry including medication reimbursement and drug purchases
556 (https://r8.risteys.finngen.fi/phenocode/MIGRAINE_TRIPTAN).

557 FinnGen participants provided informed consent under the Finnish Biobank Act.
558 Older cohorts with study-specific consents were transferred to the Finnish biobanks
559 after approval by Fimea, the National Supervisory Authority for Welfare and Health.
560 Recruitment protocols followed the biobank protocols approved by Fimea. The

561 Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa
562 (HUS) approved the FinnGen study protocol (Nr HUS/990/2017).
563 The FinnGen study is approved by Finnish Institute for Health and Welfare (permit
564 numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018,
565 THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and
566 THL/1524/5.05.00/2020), Digital and population data service agency (permit
567 numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social
568 Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018,
569 KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019,
570 KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers
571 THL/2364/14.02/2020, THL/4055/14.06.00/2020,,THL/3433/14.06.00/2020,
572 THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020,
573 THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021,
574 THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020,
575 THL/2658/14.06.00/2021, THL/4235/14.06.00/2021 and Statistics Finland (permit
576 numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20)
577 TK/1735/07.03.00/2021).
578 The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen
579 Data Freeze 8 include: THL Biobank BB2017_55, BB2017_111, BB2018_19,
580 BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26,
581 BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank
582 HUS/359/2017, Auria Biobank AB17-5154 and amendment #1 (August 17 2020),
583 AB20-5926 and amendment #1 (April 23 2020), Biobank Borealis of Northern
584 Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 §
585 /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 &

586 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB
587 2018001.

588

589 We have access to the complete in-sample LD information for the UK Biobank and
590 FinnGen samples via the individual-level genotype data. Additionally, we have
591 access to the in-sample LD-matrices in 23andMe data for 26 of our fine-map regions.
592 Thus, for the 26 fine-map regions, we are able to do a high-quality fine-mapping
593 based on the in-sample LD while, for the remaining 76 regions, we need to apply an
594 LD reference panel that does not perfectly match the LD information corresponding
595 to our GWAS summary statistics. To assess the effect of the LD reference panel, we
596 formed two reference panels from the available LD information: one including data
597 only from the UK Biobank (UKB), and the other combining the LD matrices from UK
598 Biobank and FinnGen (UKB-FG), as explained in section “Fine-mapping”.

599

600 Genetic association analyses

601

602 The UK Biobank and 23andMe GWAS had been conducted by logistic regression on
603 migraine (using PLINK2⁶² or custom software of the 23andMe Research Team,
604 respectively), and the FinnGen GWAS by a whole-genome regression model for a
605 binary trait with REGENIE²⁷.

606 All the samples were of European descent. Related individuals had been excluded
607 by using a kinship value threshold of 0.0442 computed by KING⁶³ from UK Biobank,
608 and by using a minimal expected amount of sharing between first cousins from a
609 segmental identity-by descent algorithm from 23andMe. For the FinnGen GWAS

610 analysis, REGENIE accounted for the genetic relatedness by default, and therefore
611 no relatedness exclusions were applied.

612 We excluded multi-allelic variants, and variants with minor allele frequency (MAF) <
613 0.01, IMPUTE2 info or MACH $r^2 < 0.6$, and when available, missingness > 0.05 and
614 Hardy-Weinberg equilibrium (HWE) $P < 1 \times 10^{-6}$ from each study. Consequently, we
615 are only considering biallelic common variants in this work. We recoded indels as
616 insertions (I) and deletions (D). We mapped the FinnGen GWAS summary statistics
617 positions from hg38 to hg37 by UCSC LiftOver⁶⁴. We excluded the SNPs with an
618 effect allele frequency (EAF) discrepancy of >0.30 and indels with an EAF
619 discrepancy of >0.20 compared to UK Biobank from each study following
620 Hautakangas et al. 2022.

621 We conducted an inverse-variance weighted fixed-effects meta-analysis to combine
622 the three studies by GWAMA⁶⁵ with 11,316,120 variants, of which 7,062,924 variants
623 were available in all three studies.

624

625 Genetic correlation and SNP-heritability using LD Score regression

626

627 We estimated genetic correlations between the three GWAS and SNP-heritability
628 from the migraine meta-analysis by LD Score regression v1.0.0^{66,25} with
629 precomputed 1000 Genomes European LD Scores
630 (<https://data.broadinstitute.org/alkesgroup/LDSCORE/>) limiting the analysis to the
631 HapMap3 SNPs. We used munge-tool to reformat and perform additional quality
632 control for all GWAS summary statistics prior to the genetic correlation estimation.

633 We obtained a liability scale SNP-heritability estimate⁶⁷ by using a population
634 prevalence of 16% for migraine.

635

636 Locus definition

637

638 We followed the locus definition of Hautakangas et al. (2022) and defined an LD-
639 independent genome-wide significant (GWS, $P < 5 \times 10^{-8}$) risk locus from the meta-
640 analysis by using the UKB LD. Iteratively, we chose the variant with the smallest P -
641 value as the index variant and excluded all variants that had $r^2 \geq 0.1$ with the index
642 variant, until no variant had $P < 5 \times 10^{-8}$. Next, we formed high LD regions around
643 each index variant based on the combined UKB-FG LD and r^2 threshold of 0.6. The
644 start of the high LD region was the smallest position, and the end of the region was
645 the largest position where any variant had $r^2 > 0.6$ with the index variant. Next, we
646 formed the loci by adding ± 250 kb around the high LD region and merged the
647 overlapping regions. Further, we iteratively added all other GWS variants to their
648 closest loci, and updated the loci boundaries if any of the variants added were
649 outside the existing locus boundaries. Again, the overlapping loci were merged. We
650 named each locus by the lead variant, i.e., the variant with the smallest P -value of
651 the locus.

652

653 Replication in HUNT All-in Headache and IHGC16

654

655 To replicate our new loci, we used two independent data sets with no overlaps with
656 our GWAS data: HUNT All-in Headache²⁸ (N=40,224, 7,801 cases, 32,423 controls)
657 and IHGC16 migraine meta-analysis⁹ excluding 23andMe and the Finnish cohorts (N
658 = 189,000, 27,006 migraine cases and 161,994 controls). The meta-analysis of the
659 replication data thus contained N=229,224 samples (34,807 cases and 194,417

660 controls). We used a one-sided P -value threshold of 0.05 to denote a replication and
661 assessed consistency of the effect directions by a sign test. We also reported the
662 two-sided P -value of a combined analysis of our discovery and replication results to
663 determine which of the new loci remained GWS after observing the replication data.

664

665 Fine-mapping

666

667 For fine-mapping, we first merged loci that were closer than 1.5 Mb leading to 102
668 fine-map regions. We performed fine-mapping for each fine-map region with
669 FINEMAP v1.4^{19,22}. FINEMAP is a Bayesian method that uses summary statistics
670 from a GWAS together with LD information to infer which variants are most likely
671 causal within the genomic region. We used the default prior parameters and set the
672 maximum number of causal variants to 10.

673

674 We estimated the in-sample LD correlations for the individual GWAS cohorts by
675 using LDStore²². We combined the in-sample LD correlations for the meta-analysis
676 data set by combining the study-specific LD matrices by weighting each matrix in
677 proportion to its effective sample size as follows:

$$678 \quad \mathbf{R} = (M_1 \mathbf{R}_1 + \dots + M_C \mathbf{R}_C) / M, \quad (F1)$$

679

680 where \mathbf{R}_i is the LD correlation matrix of study i , $M_i = 4N_i p_i (1-p_i)$ is the effective
681 sample size of study i , with N_i being the total sample size (i.e., the sum of cases and
682 controls) and p_i being the proportion of cases in study i , and $M = M_1 + \dots + M_C$ is the
683 sum of the effective sample sizes.

684

685 For the UK Biobank reference LD (UKB-LD), we used the in-sample LD estimated
686 from the individuals included in the UKB GWAS.

687 For the combined UKB-FG LD reference panel, we combined the UKB and FG in-
688 sample LD matrices by weighting FG in proportion to its effective sample size, and
689 UKB in proportion to the combined UKB+23andMe effective sample size using the
690 above formula (F1).

691

692 LD reference panel sensitivity analyses

693

694 We compared the performance of different LD reference panels (UKB LD, UKB-FG LD
695 and in-sample LD) on the FINEMAP results for the 26 fine-map regions for which the
696 in-sample LD was available. We used the maximum difference between the posterior
697 inclusion probabilities (PIPs) from different panels ($\max\Delta$) to compare the
698 performance of the three LD panels.

699 In addition, we examined the following candidate statistics which could be used for
700 separating the fine-map regions for which the fine-mapping with the reference LD
701 performs poorly when compared to the use of the in-sample LD: 1) the posterior
702 expectation of the number of causal variants (PENC), and, from the top variant(s) of
703 the credible set(s) determined by FINEMAP, 2) the maximum pairwise r^2 , 3) the
704 maximum marginal P -value from the meta-analysis, or 4) the minimum INFO value.

705

706 Variant annotation by VEP and eQTL mapping

707

708 FINEMAP reports 95%-credible sets (CS). We searched for coding variants among
709 the CS from the Ensembl VEP

710 (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP) database by using a default of
711 5 kb window around the index variant.

712

713 For the follow-up analyses, we formed a functional variant group among the CS
714 variants by including the variants that were predicted by VEP to have a moderate or
715 high impact on the transcript

716 (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html). This
717 includes transcript ablation, splice acceptor or donor variants, stop gained, frameshift
718 variant, stop lost, start lost, transcript amplification, inframe insertion or deletion, and
719 missense variant.

720 We mapped the functional variant set, and also another set including all variants with
721 PIP > 0.1 (highPIP), to significant eQTLs of the 49 tissues from GTEx v.8
722 (<https://gtexportal.org/home/>).

723

724 Phenome-wide association scans

725

726 We performed three phenome-wide association scans (PheWAS). First, we scanned
727 all 181 candidate variants of the risk loci (top variants of the credible sets) among the
728 2,399 FinnGen Data Freeze 10 (R10) GWAS endpoints (excluding 9 migraine
729 endpoints) at significance level 1×10^{-5} . Second, we scanned all variants annotated
730 as functional variants with a moderate to high impact on protein function by VEP
731 among neurological and cardiovascular endpoints from FinnGen R10, including the
732 FinnGen endpoint categories Neurological endpoints, VI Diseases of the nervous
733 system (G6_), and IX Diseases of the circulatory system (I9_) with at significance
734 level 1×10^{-4} .

735 Third, we scanned all variants with PIP > 0.1 among the same FinnGen neurological
736 and cardiovascular endpoints at significance level 1×10^{-4} .

737 Results can be browsed from PheWAS app

738 https://hhautakangas.github.io/phewas_migraine_tables.html.

739

740 Data availability

741 The access to the UK biobank data can be applied through

742 <https://www.ukbiobank.ac.uk/>

743 The GWAS summary statistics for FinnGen R8 are publicly available through

744 https://www.finnngen.fi/en/access_results. The Finnish biobank data can be accessed

745 through the Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB.

746 Finnish Health register data can be applied from Findata (<https://findata.fi/en/data/>).

747 The GWAS summary statistics for the 23andMe data set will be made available

748 through 23andMe to qualified researchers under an agreement with 23andMe that

749 protects the privacy of the 23andMe participants. Please visit

750 <https://research.23andme.com/collaborate/#publication> for more information and to

751 apply to access the data.

752

753 Acknowledgements

754

755 We would like to thank the research participants and employees of 23andMe, Inc. for

756 making this work possible. We thank all the study participants, employees, and

757 investigators of FinnGen and the UK Biobank for their contribution to this research.

758 This research has been conducted using the UK Biobank Resource under

759 Application Number 22627. This work was supported by grants no. 336825, 338507,

760 352795 from the Research Council of Finland to M.P., by Sigrid Jusélius foundation
761 (M.P. and A.P.) and by the Doctoral School of University of Helsinki (H.H.).
762 The FinnGen project is funded by two grants from Business Finland (HUS
763 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie
764 Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene
765 Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme
766 LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US
767 Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and
768 Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for
769 delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL
770 Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank
771 Borealis of Northern Finland ([https://www.ppsHP.fi/Tutkimus-ja-](https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx)
772 [opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx](https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx)), Finnish Clinical
773 Biobank Tampere ([www.tays.fi/en-](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)
774 [US/Research_and_development/Finnish_Clinical_Biobank_Tampere](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of
775 Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank
776 (www.ksshP.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank
777 (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank
778 (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic
779 Biobank ([https://www.oulu.fi/en/university/faculties-and-units/faculty-](https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank)
780 [medicine/northern-finland-birth-cohorts-and-arctic-biobank](https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank)). All Finnish Biobanks are
781 members of BBMRI.fi infrastructure (www.bbMRI.fi). Finnish Biobank Cooperative -
782 FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The
783 Finnish biobank data can be accessed through the Fingenious® services
784 (<https://site.fingenious.fi/en/>) managed by FINBB.

785 The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research
786 Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science
787 and Technology NTNU), Trøndelag County Council, Central Norway Regional Health
788 Authority, and the Norwegian Institute of Public Health. The genotyping was financed
789 by the National Institute of health (NIH), University of Michigan, The Norwegian
790 Research council, and Central Norway Regional Health Authority and the Faculty of
791 Medicine and Health Sciences, Norwegian University of Science and Technology
792 (NTNU). The genotype quality control and imputation has been conducted by the
793 K.G. Jebsen center for genetic epidemiology, Department of public health and
794 nursing, Faculty of medicine and health sciences, Norwegian University of Science
795 and Technology (NTNU).

796

797 Consortia

798 A full list of FinnGen members and their affiliations appears in the Supplementary
799 Data 1.

800

801 International Headache Genetics Consortium

802 Verner Anttila^{1,2,3}, Ville Artto⁴, Andrea C Belin⁵, Anna Bjornsdottir⁶, Gyda

803 Bjornsdottir⁷, Dorret I Boomsma⁸, Sigrid Børte^{9,10,11}, Mona A Chalmer¹², Daniel I

804 Chasman^{13,14}, Bru Cormand¹⁵, Ester Cuenca-Leon¹⁶, George Davey-Smith¹⁷, Irene

805 de Boer¹⁸, Martin Dichgans^{19,20}, Tonu Esko²¹, Tobias Freilinger^{22,23}, Padhraig

806 Gormley²⁴, Lyn R Griffiths²⁵, Eija Hämäläinen²⁶, Thomas F Hansen^{12,27}, Aster VE

807 Harder^{18,28}, Heidi Hautakangas²⁶, Marjo Hiekkala²⁹, Maria G Hrafnsdottir³⁰, M. Arfan

808 Ikram³¹, Marjo-Riitta Järvelin^{32,33,34,35}, Risto Kajanne²⁶, Mikko Kallela⁴, Jaakko

809 Kaprio²⁶, Mari Kaunisto²⁹, Lisette JA Kogelman¹², Espen S Kristoffersen^{36,37,38},

810 Christian Kubisch³⁹, Mitja Kurki⁴⁰, Tobias Kurth⁴¹, Lenore Launer⁴², Terho
811 Lehtimäki⁴³, Davor Lessel³⁹, Lannie Ligthart⁸, Sigurdur H Magnusson⁷, Rainer
812 Malik¹⁹, Bertram Müller-Myhsok⁴⁴, Carrie Northover⁴⁵, Dale R Nyholt⁴⁶, Jes Olesen¹²,
813 Aarno Palotie^{26,47}, Priit Palta²⁶, Linda M Pedersen⁴⁸, Nancy Pedersen⁴⁹, Matti
814 Pirinen^{26,50,51}, Danielle Posthuma⁵², Patricia Pozo-Rosich⁵³, Alice Pressman⁵⁴, Olli
815 Raitakari^{55,56,57}, Caroline Ran⁵, Gudrun R Sigurdardottir⁶, Hreinn Stefansson⁷, Kari
816 Stefansson⁷, Olafur A Sveinsson³⁰, Gisela M Terwindt¹⁸, Thorgeir E Thorgeirsson⁷,
817 Arn MJM van den Maagdenberg^{18,28}, Cornelia van Duijn⁵⁸, Maija Wessman^{26,29},
818 Bendik S Winsvold^{9,48,59}, John-Anker Zwart^{9,10,48}

819

820 ¹Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital
821 and Harvard Medical School, Boston, Massachusetts, USA; ²Program in Medical and Population
822 Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; ³Stanley Center for
823 Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA;
824 ⁴Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland; ⁵Department of
825 Neuroscience, Karolinska Institutet, Stockholm, Sweden; ⁶Neurology private practice, Laeknasetrid,
826 Reykjavik, Iceland; ⁷deCODE genetics/Amgen Inc., Reykjavik, Iceland; ⁸Netherlands Twin Register,
827 Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands; ⁹K.G. Jebsen
828 Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and
829 Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway; ¹⁰Institute of
830 Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; ¹¹Research and
831 Communication Unit for Musculoskeletal Health, Department of Research, Innovation and Education,
832 Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway; ¹²Danish Headache Center,
833 Department of Neurology, Copenhagen University Hospital, Copenhagen, Denmark; ¹³Department of
834 Medicine, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts,
835 USA; ¹⁴Harvard Medical School, Boston, Massachusetts, USA; ¹⁵Department of Genetics, Spain
836 Centre for Biomedical Network Research on Rare Diseases, University of Barcelona, Barcelona,
837 Spain; ¹⁶Pediatric Neurology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain;
838 ¹⁷University of Bristol/Medical Research Council Integrative Epidemiology Unit, University of Bristol,
839 Bristol, UK; ¹⁸Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands;
840 ¹⁹Institute for Stroke and Dementia Research, University Hospital, LMU Munich, Munich, Germany;
841 ²⁰Munich Cluster for Systems Neurology, Munich, Germany; ²¹Estonian Biobank Registry, the
842 Estonian Genome Center, University of Tartu, Tartu, Estonia; ²²Department of Neurology, Klinikum
843 Passau, Passau, Germany; ²³Department of Neurology and Epileptology, Hertie Institute for Clinical

844 Brain Research, University of Tuebingen, Tuebingen, Germany; ²⁴GSK Inc., Cambridge,
845 Massachusetts, USA; ²⁵Centre for Genomics and Personalised Health, Queensland University of
846 Technology, Brisbane, Queensland, Australia; ²⁶Institute for Molecular Medicine Finland, Helsinki
847 Institute of Life Science, University of Helsinki, Helsinki, Finland; ²⁷Novo Nordic Foundation Center for
848 Protein Research, Copenhagen University, Copenhagen, Denmark; ²⁸Department of Human
849 Genetics, Leiden University Medical Centre, Leiden, the Netherlands; ²⁹Folkhälsan Research Center,
850 Helsinki, Finland; ³⁰Landspítali University Hospital, Reykjavik, Iceland; ³¹Department of Epidemiology,
851 Erasmus University Medical Center, Rotterdam, the Netherlands; ³²Department of Epidemiology and
852 Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College
853 London, London, UK; ³³Center for Life Course Health Research, Faculty of Medicine, University of
854 Oulu, Oulu, Finland; ³⁴Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland;
855 ³⁵Department of Life Sciences, College of Health and Life Sciences, Brunel University London,
856 London, UK; ³⁶Research and Communication Unit for Musculoskeletal Health, Department of
857 Research, Innovation and Education, Division of Clinical Neuroscience, Akershus University Hospital
858 and University of Oslo, Oslo, Norway; ³⁷Department of General Practice, Institute of Health and
859 Society, University of Oslo, Oslo, Norway; ³⁸Department of Neurology, Akershus University Hospital,
860 Lørenskog, Norway; ³⁹Institute of Human Genetics, University Medical Center Hamburg-Eppendorf,
861 Hamburg, Germany; ⁴⁰Psychiatric and Neurodevelopmental Genetics Unit, Department of Medicine,
862 Massachusetts General Hospital, Boston, Massachusetts, USA; ⁴¹Institute of Public Health, Charité –
863 Universitätsmedizin, Berlin; ⁴²Laboratory of Epidemiology and Population Sciences, Intramural
864 Research Program, National Institute on Aging, Bethesda, Maryland, USA; ⁴³Department of Clinical
865 Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of
866 Medicine and Health Technology, Tampere University, Tampere, Finland; ⁴⁴Max Planck Institute of
867 Psychiatry, Munich, Germany; ⁴⁵23&Me Inc., Mountain View, California, USA; ⁴⁶School of Biomedical
868 Sciences, Faculty of Health, Centre for Genomics and Personalised Health, Centre for Data Science,
869 Queensland University of Technology, Brisbane, Queensland, Australia; ⁴⁷University of Helsinki,
870 Helsinki, Finland; ⁴⁸Department of Research, Innovation and Education, Division of Clinical
871 Neuroscience, Oslo University Hospital, Oslo, Norway; ⁴⁹Department of Medical Epidemiology and
872 Biostatistics, Karolinska Institutet, Stockholm, Sweden; ⁵⁰Department of Mathematics and Statistics,
873 University of Helsinki, Helsinki, Finland; ⁵¹Department of Public Health, University of Helsinki, Helsinki,
874 Finland; ⁵²Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive
875 Research, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands;
876 ⁵³Headache Unit, Neurology Department, Vall d'Hebron University Hospital, Barcelona, Spain;
877 ⁵⁴Sutter Health, Sacramento, California, USA; ⁵⁵Centre for Population Health Research, University of
878 Turku, Turku University Hospital, Turku, Finland; ⁵⁶Research Centre of Applied and Preventive
879 Cardiovascular Medicine, University of Turku, Turku, Finland; ⁵⁷Department of Clinical Physiology and
880 Nuclear Medicine, Turku University Hospital, Turku, Finland; ⁵⁸Department of Epidemiology, Erasmus
881 University Medical Centre, Rotterdam, the Netherlands; ⁵⁹Department of Neurology, Oslo University
882 Hospital, Oslo, Norway
883

885 HUNT All-in Headache

886 Amy E Martinsen^{1,2,3}, Anne Heidi Skogholt³, Ben M Brumpton³, Bendik S
887 Winsvold^{4,5,6}, Cristen J Willer⁷, Erling Tronvik^{8,9}, Espen Saxhaug Kristoffersen^{1,10,11},
888 John-Anker Zwart^{1,2,3}, Jonas B Nielsen^{3,7,12}, Knut Hagen⁸, Kristian Hveem^{3,13,14},
889 Kristian Bernhard Nilsen^{8,15}, Lars G Fritsche¹⁶, Lars Jacob Stovner^{8,17}, Laurent F
890 Thomas^{3,18,19,20}, Linda M Pedersen¹, Maiken E Gabrielsen³, Marianne B Johnsen^{3,21},
891 Marie U Lie^{2,21}, Oddgeir L Holmen¹³, Sigrid Børte^{2,3,21}, Wei Zhou^{22,23}

892

893 ¹Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital,
894 Oslo, Norway, ²Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway,
895 ³K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of
896 Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU),
897 Trondheim, Norway, ⁴Department of Research and Innovation, Division of Clinical Neuroscience, Oslo
898 University Hospital, Oslo, Norway, ⁵K. G. Jebsen Center for Genetic Epidemiology, Department of
899 Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of
900 Science and Technology (NTNU), Trondheim, Norway, ⁶Department of Neurology, Oslo University
901 Hospital, Oslo, Norway, ⁷Department of Internal Medicine, Division of Cardiovascular Medicine,
902 University of Michigan, Ann Arbor, MI, 48109, USA, ⁸Department of Neuromedicine and Movement
903 Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology
904 (NTNU), Trondheim, Norway, ⁹Department of Neurology and Clinical Neurophysiology, St. Olavs
905 Hospital, Trondheim University Hospital, Trondheim, Norway, ¹⁰Department of General Practice,
906 University of Oslo, Oslo, Norway, ¹¹Department of Neurology, Akershus University Hospital,
907 Lørenskog, Norway, ¹²Department of Epidemiology Research, Statens Serum Institut, Copenhagen,
908 Denmark, ¹³HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine
909 and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
910 ¹⁴Department of Research, Innovation and Education, St. Olavs Hospital, Trondheim University
911 Hospital, Trondheim, Norway, ¹⁵Department of Neurology, Oslo University Hospital, Oslo, Norway,
912 ¹⁶Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI,
913 48109, USA, ¹⁷Norwegian Advisory Unit on Headaches, St. Olavs Hospital, Trondheim University
914 Hospital, Trondheim, Norway, ¹⁸Department of Clinical and Molecular Medicine, Norwegian University
915 of Science and Technology (NTNU), Trondheim, Norway, ¹⁹BioCore - Bioinformatics Core Facility,
916 Norwegian University of Science and Technology (NTNU), Trondheim, Norway, ²⁰Clinic of Laboratory
917 Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, ²¹Research and
918 Communication Unit for Musculoskeletal Health (FORMI), Department of Research and Innovation,
919 Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway, ²²Department of

920 Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA,

921 ²³Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

922

923 Competing interests

924

925 A.P. is the Scientific Director of the public-private partnership project FinnGen that

926 has 12 industry partners that provide funding for the FinnGen project.

927 Other authors report no conflicts of interests.

928

929

930 References:

931

- 932 1. (IHS), H.C.C.o.t.I.H.S. Headache Classification Committee of the International
933 Headache Society (IHS) The International Classification of Headache Disorders, 3rd
934 edition. *Cephalalgia* **38**, 1-211 (2018).
- 935 2. Vos, T. *et al.* Global burden of 369 diseases and injuries in 204 countries and
936 territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study
937 2019. *The Lancet* **396**, 1204-1222 (2020).
- 938 3. Gervil, M., Ulrich, V., Kaprio, J., Olesen, J. & Russell, M.B. The relative role of genetic
939 and environmental factors in migraine without aura. *Neurology* **53**, 995-999 (1999).
- 940 4. Anttila, V. *et al.* Genome-wide association study of migraine implicates a common
941 susceptibility variant on 8q22.1. *Nature Genetics* **42**, 869-873 (2010).
- 942 5. Chasman, D.I. *et al.* Genome-wide association study reveals three susceptibility loci
943 for common migraine in the general population. *Nature Genetics* **43**, 695-U116
944 (2011).
- 945 6. Freilinger, T. *et al.* Genome-wide association analysis identifies susceptibility loci for
946 migraine without aura. *Nature Genetics* **44**, 777-782 (2012).
- 947 7. Anttila, V. *et al.* Genome-wide meta-analysis identifies new susceptibility loci for
948 migraine. *Nature Genetics* **45**, 912-U255 (2013).
- 949 8. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42
950 human traits. *Nature Genetics* **48**, 709-717 (2016).
- 951 9. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci
952 for migraine. *Nature Genetics* **48**, 856-866 (2016).
- 953 10. Chen, S.-P. *et al.* Genome-wide association study identifies novel susceptibility loci
954 for migraine in Han Chinese resided in Taiwan. *Cephalalgia* **38**, 466-475 (2018).
- 955 11. Chang, X. *et al.* Common variants at 5q33.1 predispose to migraine in African-
956 American children. *Journal of Medical Genetics* **55**, 831 (2018).
- 957 12. Choquet, H. *et al.* New and sex-specific migraine susceptibility loci identified from a
958 multiethnic genome-wide meta-analysis. *Communications Biology* **4**, 864 (2021).
- 959 13. Hautakangas, H. *et al.* Genome-wide analysis of 102,084 migraine cases identifies
960 123 risk loci and subtype-specific risk alleles. *Nature Genetics* **54**, 152-160 (2022).
- 961 14. Bjornsdottir, G. *et al.* Rare variants with large effects provide functional insights into
962 the pathology of migraine subtypes, with and without aura. *Nature Genetics* **55**,
963 1843-1853 (2023).
- 964 15. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies
965 disease-relevant tissues and cell types. *Nature Genetics* **50**, 621-629 (2018).
- 966 16. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate
967 causal variants by statistical fine-mapping. *Nature Reviews Genetics* **19**, 491-504
968 (2018).
- 969 17. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical
970 fine-mapping studies. *PLoS Genet* **10**, e1004722 (2014).
- 971 18. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B. & Eskin, E. Identifying causal
972 variants at loci with multiple signals of association. *Genetics* **198**, 497-508 (2014).
- 973 19. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from
974 genome-wide association studies. *Bioinformatics* **32**, 1493-1501 (2016).

- 975 20. Newcombe, P.J., Conti, D.V. & Richardson, S. JAM: A Scalable Bayesian Framework
976 for Joint Analysis of Marginal SNP Effects. *Genetic Epidemiology* **40**, 188-201 (2016).
- 977 21. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A Simple New Approach to
978 Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal*
979 *of the Royal Statistical Society Series B: Statistical Methodology* **82**, 1273-1300
980 (2020).
- 981 22. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by
982 Using Summary Statistics from Genome-wide Association Studies. *American Journal*
983 *of Human Genetics* **101**, 539-551 (2017).
- 984 23. Kanai, M. *et al.* Meta-analysis fine-mapping is often miscalibrated at single-variant
985 resolution. *Cell Genomics* **2**(2022).
- 986 24. Trubetsky, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
987 schizophrenia. *Nature* **604**, 502-508 (2022).
- 988 25. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and
989 traits. *Nature Genetics* **47**, 1236-1241 (2015).
- 990 26. Isgut, M., Song, K., Ehm, M.G., Wang, M.D. & Davitte, J. Effect of case and control
991 definitions on genome-wide association study (GWAS) findings. *Genetic*
992 *Epidemiology* **47**, 394-406 (2023).
- 993 27. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for
994 quantitative and binary traits. *Nature Genetics* **53**, 1097-1103 (2021).
- 995 28. Krokstad, S. *et al.* Cohort Profile: The HUNT Study, Norway. *International Journal of*
996 *Epidemiology* **42**, 968-977 (2012).
- 997 29. Yang, A.W., Sachs, A.J. & Nystuen, A.M. Deletion of Inpp5a causes ataxia and
998 cerebellar degeneration in mice. *neurogenetics* **16**, 277-285 (2015).
- 999 30. Liu, Q. *et al.* Cerebellum-enriched protein INPP5A contributes to selective
1000 neuropathology in mouse model of spinocerebellar ataxias type 17. *Nature*
1001 *Communications* **11**, 1101 (2020).
- 1002 31. Zorzetto, M. *et al.* SERPINA1 Gene Variants in Individuals from the General
1003 Population with Reduced α 1-Antitrypsin Concentrations. *Clinical Chemistry* **54**, 1331-
1004 1338 (2008).
- 1005 32. Gupta, R.M. *et al.* A Genetic Variant Associated with Five Vascular Diseases Is a Distal
1006 Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522-533.e15 (2017).
- 1007 33. Wang, X. & Musunuru, K. Confirmation of Causal rs9349379-<i>PHACTR1</i>
1008 Expression Quantitative Trait Locus in Human-Induced Pluripotent Stem Cell
1009 Endothelial Cells. *Circulation: Genomic and Precision Medicine* **11**, e002327 (2018).
- 1010 34. Guo, D.-c. *et al.* Genetic Variants in LRP1 and ULK4 Are
1011 Associated with Acute Aortic Dissections. *The American Journal of Human Genetics*
1012 **99**, 762-769 (2016).
- 1013 35. Georges, A. *et al.* Genetic investigation of fibromuscular dysplasia identifies risk loci
1014 and shared genetics with common cardiovascular diseases. *Nature Communications*
1015 **12**, 6031 (2021).
- 1016 36. Turley, T.N. *et al.* Identification of Susceptibility Loci for Spontaneous Coronary
1017 Artery Dissection. *JAMA Cardiology* **5**, 929-938 (2020).
- 1018 37. Storck, S.E., Kurtyka, M. & Pietrzik, C.U. Brain endothelial LRP1 maintains blood-
1019 brain barrier integrity. *Fluids and Barriers of the CNS* **18**, 27 (2021).
- 1020 38. Liu, Z., Andraska, E., Akinbode, D., Mars, W. & Alvidrez, R.I.M. LRP1 in the Vascular
1021 Wall. *Current Pathobiology Reports* **10**, 23-34 (2022).

- 1022 39. Lee, J. *et al.* ANKS1A regulates LDL receptor-related protein 1 (LRP1)-mediated
1023 cerebrovascular clearance in brain endothelial cells. *Nature Communications* **14**,
1024 8463 (2023).
- 1025 40. Liu, L. *et al.* Regulatory mechanisms in multiple vascular diseases locus
1026 *LRP1* involve repression by SNAIL and extracellular matrix remodeling.
1027 *bioRxiv*, 2023.05.09.539992 (2023).
- 1028 41. Korsching, S., Auburger, G., Heumann, R., Scott, J. & Thoenen, H. Levels of nerve
1029 growth factor and its mRNA in the central nervous system of the rat correlate with
1030 cholinergic innervation. *The EMBO Journal* **4**, 1389-1393 (1985).
- 1031 42. Connor, B. & Dragunow, M. The role of neuronal growth factors in
1032 neurodegenerative disorders of the human brain. *Brain Research Reviews* **27**, 1-39
1033 (1998).
- 1034 43. van Dongen, R.M. *et al.* Migraine biomarkers in cerebrospinal fluid: A systematic
1035 review and meta-analysis. *Cephalalgia* **37**, 49-63 (2017).
- 1036 44. Mozafarihashjin, M. *et al.* Assessment of peripheral biomarkers potentially involved
1037 in episodic and chronic migraine: a case-control study with a focus on NGF, BDNF,
1038 VEGF, and PGE2. *The Journal of Headache and Pain* **23**, 3 (2022).
- 1039 45. Aloe, L., Tuveri, M.A., Carcassi, U. & Levi-Montalcini, R. Nerve growth factor in the
1040 synovial fluid of patients with chronic arthritis. *Arthritis & Rheumatism* **35**, 351-355
1041 (1992).
- 1042 46. Freemont, A.J. *et al.* Nerve growth factor expression and innervation of the painful
1043 intervertebral disc. *The Journal of Pathology* **197**, 286-292 (2002).
- 1044 47. Walsh, D.A. *et al.* Angiogenesis and nerve growth factor at the osteochondral
1045 junction in rheumatoid arthritis and osteoarthritis. *Rheumatology (Oxford)* **49**, 1852-
1046 61 (2010).
- 1047 48. Sanga, P. *et al.* Efficacy, safety, and tolerability of fulranumab, an anti-nerve growth
1048 factor antibody, in the treatment of patients with moderate to severe osteoarthritis
1049 pain. *Pain* **154**, 1910-1919 (2013).
- 1050 49. Tiseo, P.J., Ren, H. & Mellis, S. Fasinumab (REGN475), an antinerve growth factor
1051 monoclonal antibody, for the treatment of acute sciatic pain: results of a proof-of-
1052 concept study. *Journal of Pain Research* **7**, 523-30 (2014).
- 1053 50. Watt, F.E. *et al.* Tropomyosin-related kinase A (TrkA) inhibition for the treatment of
1054 painful knee osteoarthritis: results from a randomized controlled phase 2a trial.
1055 *Osteoarthritis Cartilage* **27**, 1590-1598 (2019).
- 1056 51. Berenbaum, F. *et al.* Subcutaneous tanezumab for osteoarthritis of the hip or knee:
1057 efficacy and safety results from a 24-week randomised phase III study with a 24-
1058 week follow-up period. *Ann Rheum Dis* **79**, 800-810 (2020).
- 1059 52. Wise, B.L., Seidel, M.F. & Lane, N.E. The evolution of nerve growth factor inhibition
1060 in clinical medicine. *Nature Reviews Rheumatology* **17**, 34-46 (2021).
- 1061 53. Koya, Y. *et al.* A novel anti-NGF PEGylated Fab' provides analgesia with lower risk of
1062 adverse effects. *mAbs* **15**, 2149055 (2023).
- 1063 54. Detke, H.C. *et al.* Galcanezumab in chronic migraine: The randomized, double-blind,
1064 placebo-controlled REGAIN study. *Neurology* **91**, e2211-e2221 (2018).
- 1065 55. Dodick, D.W. *et al.* ARISE: A Phase 3 randomized trial of erenumab for episodic
1066 migraine. *Cephalalgia* **38**, 1026-1037 (2018).
- 1067 56. Ferrari, M.D. *et al.* Fremanezumab versus placebo for migraine prevention in
1068 patients with documented failure to up to four migraine preventive medication

- 1069 classes (FOCUS): a randomised, double-blind, placebo-controlled, phase 3b trial. *The*
1070 *Lancet* **394**, 1030-1040 (2019).
- 1071 57. De Smedt, F., Verjans, B., Mailleux, P. & Erneux, C. Cloning and expression of human
1072 brain type I inositol 1,4,5-trisphosphate 5-phosphatase High levels of mRNA in
1073 cerebellar Purkinje cells. *FEBS Letters* **347**, 69-72 (1994).
- 1074 58. Rubin, S. *et al.* PHACTR-1 (Phosphatase and Actin Regulator 1) Deficiency in Either
1075 Endothelial or Smooth Muscle Cells Does Not Predispose Mice to Nonatherosclerotic
1076 Arteriopathies in 3 Transgenic Mice. *Arteriosclerosis, Thrombosis, and Vascular*
1077 *Biology* **42**, 597-609 (2022).
- 1078 59. Kurth, T. *et al.* Migraine and risk of cardiovascular disease in women: prospective
1079 cohort study. *BMJ* **353**, i2610 (2016).
- 1080 60. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
1081 *Nature* **562**, 203-209 (2018).
- 1082 61. Kurki, M.I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated
1083 population. *Nature* **613**, 508-518 (2023).
- 1084 62. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1085 richer datasets. *GigaScience* **4**, s13742-8 (2015).
- 1086 63. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association
1087 studies. *Bioinformatics* **26**, 2867-2873 (2010).
- 1088 64. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids*
1089 *Research* **34**, D590-8 (2006).
- 1090 65. Mägi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-
1091 analysis. *BMC bioinformatics* **11**, 288 (2010).
- 1092 66. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from
1093 polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295
1094 (2015).
- 1095 67. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability
1096 for disease from genome-wide association studies. *American Journal of Human*
1097 *Genetics* **88**, 294-305 (2011).
- 1098