

# **A genome-wide association study finds novel genetic associations with broadly-defined headache in UK Biobank (N = 223,773)**

Weihua Meng<sup>1\*</sup>, Mark J Adams<sup>2\*</sup>, Harry L Hebert<sup>1</sup>, Ian J Deary<sup>3</sup>, Andrew M McIntosh<sup>2,3</sup>, Blair H Smith<sup>1</sup>

1 Division of Population Health Sciences, School of Medicine, University of Dundee, Dundee, UK, DD2 4BF

2 Division of Psychiatry, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK, EH10 5HF

3 Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh, UK, EH8 9JZ

\*Weihua Meng and Mark J Adams contributed equally to this paper.

\*Corresponding author: Weihua Meng [w.meng@dundee.ac.uk](mailto:w.meng@dundee.ac.uk)

Address: Division of Population Health Sciences, School of Medicine, University of Dundee, Dundee, UK, DD2 4BF. Tel.: +44 1382383419; fax: +44 1382383802

## **Abstract**

Headache is the most common neurological symptom and a leading cause of years lived with disability. We sought to identify the genetic variants associated with a broadly-defined headache phenotype in 223,773 subjects from the UK Biobank cohort. We defined headache based on a specific question answered by the UK Biobank participants. We performed a genome-wide association study of headache as a single entity, using 74,461

cases and 149,312 controls. We identified 3,343 SNPs which reached the genome-wide significance level of  $P < 5 \times 10^{-8}$ . The SNPs were located in 28 loci, with the top SNP of rs11172113 in the *LRP1* gene having a  $P$  value of  $4.92 \times 10^{-47}$ . Of the 28 loci, 14 have previously been associated with migraine. Among 14 new loci, rs77804065 with a  $P$  value of  $5.87 \times 10^{-15}$  in the *LINC02210-CRHR1* gene was the top SNP.

Positive relationships ( $P < 0.001$ ) between multiple brain tissues and genetic associations were identified through tissue expression analysis, whereas no vascular related tissues showed significant relationships. We identified several significant positive genetic correlations between headache and other psychological traits including neuroticism, depressive symptoms, insomnia, and major depressive disorder.

Our results suggest that brain function is closely related to broadly-defined headache. In addition, we also found that many psychological traits have genetic correlations with headache.

Key words: headache, genome-wide association study, *LRP1*, UK Biobank, tissue expression

## Introduction

Headache is the most common neurological symptom, with a life time prevalence of over 90% in the general population in the UK (1). It represents 4.4% of consultations in primary care and 30% of outpatient consultations in neurology (2,3).

According to the International Headache Society, headache can be generally divided into two categories: primary headache, if not associated with another disorder; and secondary headache if associated with an underlying medical illness (4). Primary headaches mainly include migraine, tension-type, and cluster headaches. Secondary headaches include any head pain caused by infection, neoplasm, head injury, some metabolic disorders, or drugs (4).

In a comprehensive review of population-based epidemiological studies of headache, the global prevalence of recurrent headache in all ages was found to be 46% for all headaches, including 11% for migraine and 42% for tension-type headache (5). Tension-type headache is the most prevalent type of headache, whereas migraine is the most disabling (6).

Migraine affects around 6 million people in England in the age range 16-65 and it is the sixth cause in terms of years of life lost to disability according to the Global Burden of Diseases 2013 (7). Migraine costs the National Health Service (NHS) almost £2 billion per year (8). It presents with recurrent headache attacks and/or hypersensitivity to light and sound. Around one third of migraineurs experience an aura, which are transient neurological symptoms mostly involving the visual system (9).

Family studies and twin studies have suggested that both migraine and tension-type headache are heritable traits with a heritability over 40% (10,11). Recently, genome-wide association studies (GWAS) have identified many genetic loci associated with migraine (12-16). A GWAS meta-analysis of 375,000 patients involving 22 centres has identified 38 genetic susceptibility loci for migraine with the *LRP1* region in chromosome 12 being the

most strongly associated (17). Along with other GWAS on migraine, the total number of loci identified to be associated with migraine is currently 47 (18). No GWAS have been performed for tension-type headache so far.

There are several phenotypic associations between headache and metabolic, psychological, and other factors such as obesity (19,20). Genome-wide association studies provide a potential route to discover genetic correlations with other complex traits and diseases that in turn may provide clues to shared genetic architectures and aetiologies (21).

To identify the genetic variants associated with headache, we conducted this GWAS using the UK Biobank cohort which has never been contributed to genetic studies of headache including migraine. We used a broadly-defined headache phenotype, the one available in the UK Biobank dataset. Secondly, we sought to test for shared genetic associations with other complex traits and diseases using linkage-disequilibrium score regression (22).

## **Results:**

### **GWAS results**

During the initial assessment visit (2006-2010), at which participants were recruited and consent was given by 501,708 UK Biobank participants, the specific pain question (see the materials and methods section for details) received 775,252 responses to all options. Among these responses, the number of participants who selected the 'Headache' option was 102,994 (cases), and the number of participants who selected the 'None of the above' option was 197,149 (controls). We further removed those whose ancestry was not white British ( $n = 22,694$ ) based on principal component analysis, those who were related to one or more others in the cohort (a cut-off value of 0.025 in the generation of the genetic relationship matrix) ( $n = 52,166$ ), those who were also in a Psychiatric Genomics Consortium MDD cohort ( $n = 597$ ), and those who failed QC ( $n = 913$ ). Thus we finally identified 74,461 cases

(27,350 males and 47,111 females) and 149,312 controls (71,480 males and 77,832 females) for the GWAS association analysis. After quality control, there were 9,304,965 single nucleotide polymorphisms (SNPs) for the GWAS analysis.

The clinical characteristics of these cases and controls are summarised in Table 1. There were statistical differences ( $P < 0.001$ ) in age, sex and body mass index (BMI) between cases and controls.

We identified 3,343 SNPs which reached GWAS significance of  $P < 5 \times 10^{-8}$  (Fig 1, Supplementary Table S1). These SNPs represented 28 independent loci including 14 newly-identified loci (Table 2).

The first cluster was in the LDL receptor related protein 1 (*LRP1*) gene in the chromosome 12 area with a lowest  $P$  value of  $4.92 \times 10^{-47}$  for rs11172113 (C allele, odds ratio (OR): 0.98). The second cluster was in the four and a half LIM domains 5 (*FHL5*) gene in chromosome 6 with a lowest  $P$  value of  $2.58 \times 10^{-31}$  for rs9486715 (C allele, OR: 1.02). Among the 28 loci, 14 of them were newly identified. Rs77804065 in the LINC02210-CRHR1 readthrough (*LINC02210-CRHR1*) gene with a  $P$  value of  $5.87 \times 10^{-15}$  (T allele, OR: 1.01) was the most strongly associated among the newly-identified loci. The Q-Q plot of the GWAS is shown in Supplementary Fig S1.

### **Gene analysis, gene-set analysis and tissue expression analysis by FUMA**

In the gene analysis by MAGMA integrated in FUMA, all the SNPs are mapped to 19,436 protein coding genes if the SNPs are located within genes. The default SNP-wise (mean) model for the gene analysis was applied. The signal transducer and activator of transcription 6 (*STAT6*) gene reached the lowest  $P$  value of  $1.1 \times 10^{-46}$ , followed by UFM1 specific ligase 1 (*UFL1*) ( $P = 2.53 \times 10^{-26}$ ), *FHL5* ( $P = 8.64 \times 10^{-25}$ ) and *LRP1* ( $P = 3.85 \times 10^{-23}$ ). All genes ( $N = 160$ ) with a  $P$  less than  $3 \times 10^{-6}$  ( $0.05/19436$ ) are included in the Supplementary Table S2.

In the gene set analysis by MAGMA integrated in FUMA, a total of 10,894 gene sets were tested and a default competitive test model was applied. Gene sets of positive regulation of gene expression, positive regulation of transcription from RNA polymerase ii promoter, neurogenesis, and excitatory synapse reached a  $P$  value less than 0.0001, but not statistical significant of  $P < 5 \times 10^{-6}$  (0.05/10,894). The top 10 gene sets from the analysis were included in Supplementary Table S3.

In the tissue expression analysis by GTEx integrated in the FUMA, average gene-expression per tissue type was used as gene covariate to test positive relationship between gene expression in a specific tissue type and genetic associations. Two types of tissue analysis were performed. One used 30 general tissue types and the other used 53 specific tissue types. In the expression analysis of 30 general tissue types, the tissue in the brain showed the most significant  $P$  value ( $P = 4.12 \times 10^{-6}$ ), followed by the tissue from blood vessel ( $P = 0.014$ ) (Supplementary Table S4). In the 53 specific tissue types, tissues from the brain cortex reached a lowest  $P$  value of  $1.44 \times 10^{-5}$  and most of the brain specific tissues also reached a significant  $P$  value of 0.001 (0.05/53). None of the vascular-related tissues reached significance. See Fig 2 and Supplementary Table S5.

### **Genetic correlation analysis by LD hub**

Through the genetic correlation analysis (see the materials and methods section for details), we identified multiple significant positive correlations for headache (Supplementary Table S6). The significant genetic correlations ( $r_g$ ) surviving multiple testing correction ( $P < 0.05/234$ ) were: neuroticism ( $r_g = 0.50$ ,  $P = 2.24 \times 10^{-72}$ ), depressive symptoms ( $r_g = 0.52$ ,  $P = 1.60 \times 10^{-46}$ ), years of education ( $r_g = -0.28$ ,  $P = 5.25 \times 10^{-37}$ ), maternal age at first delivery ( $r_g = -0.32$ ,  $P = 3.97 \times 10^{-29}$ ), subjective wellbeing ( $r_g = -0.37$ ,  $P = 9.51 \times 10^{-19}$ ), insomnia ( $r_g = 0.42$ ,  $P = 2.54 \times 10^{-18}$ ), and major depressive disorder ( $r_g = 0.39$ ,  $P = 1.57 \times 10^{-11}$ ).

## Discussion

We have performed a GWAS on broadly-defined headache as a single entity using the UK Biobank resource and found that variants in 28 loci were associated with having experienced headache within the last month to the extent that it interfered with usual activities. Evidence from tissue expression analysis showed that brain function is closely related to this broadly-defined headache. In addition, we found that headache was genetically correlated with a number of psychological factors, including those linked to a higher tendency toward experiencing negative emotional states, and shorter duration of education.

In this study, we defined headache cases and controls based on the responses by UK Biobank participants to a specific pain question. This question focused on headache occurrence, sufficient to cause interference with activities, during the previous month. We can therefore only treat headache as a global condition to perform GWAS. We can hypothesise, based on the reported population prevalence of each subtype (5,6), that tension-type headache will be the most common diagnosis among cases, followed by migraine, and that many will have experienced more than one type of headache. Whether the SNPs identified are associated with one or more of these diagnoses specifically, or with headache globally remains unclear until further research is conducted. It is important to note that half of the 28 loci were previously reported loci for migraine (Table 2).

The UK Biobank genetic resource is especially useful as a screening tool to test whether heterogeneous phenotypes such as headache have genetic components at all, as the UK Biobank has collected many heterogeneous phenotypes which need further genetic investigation. An example of such a GWAS approach was adopted by Deary et al. on self-reported tiredness (a heterogeneous phenotype) using the UK Biobank cohort (23). GWAS like this will help with genetic stratification analysis for heterogeneous phenotypes. This genetic discovery phase could be part of the partitioning or classification of broadly-reported

or broadly-defined headache, improving our understanding of its aetiology, diagnosis, prognosis and the development of treatments.

The findings of this research will facilitate the next essential phases of research into the genetics and causes of headache. They will enable investigators to conduct adequate power calculations for future studies and will provide candidate loci to examine in replication studies. Even with our very heterogeneous group of cases, including many different headache subtypes, we were still able to identify numerous GWAS associations. Therefore, it can be expected that in a GWAS of a more intensively-phenotyped headache cohort or a subtype of headache such as tension-type headache for which no GWAS has yet been performed, genetic contributions could be more distinctive and easier to identify than in this study (for example, GWAS hits can be identified using fewer samples). Since our results are of GWAS significance, the genes (such as *LRP1* and *FHL5*) identified for migraine by other GWAS are also promising candidate genes for tension-type headache. The majority of the patients with migraine (over 90%) also suffer from tension-type headache (24), and it is possible that migraine and tension-type headache might share common genetic components.

In this GWAS, we have identified 28 loci for headache. One was in the *LRP1* gene with a lowest *P* value of  $4.92 \times 10^{-47}$  for rs11172113. The *LRP1* gene encodes a precursor protein that is processed by furin in the trans-Golgi complex, generating a 515 kDa alpha-chain and an 85 kDa beta-chain associated non-covalently (25). This protein is involved in multiple cellular processes, including intracellular signalling, lipid homeostasis, and clearance of apoptotic cells (26). The *LRP1* gene has also been reported to be associated with Alzheimer's disease, cardiovascular disease, and tumours (27-29). Chasman et al first identified the association between the *LRP1* gene and migraine in a Caucasian population.<sup>14</sup> The association was then replicated by Freilinger et al in migraine without aura in German and Dutch patients (15). The locus was further replicated in Indian but not in Chinese populations (30,31). It is worth noting that many nearby supporting SNPs for the *LRP1*



region were located in the nearby gene of *STAT6* (which was also strongly associated with headache).

The second SNP cluster was in the *FHL5* gene area with a lowest *P* value of  $2.58 \times 10^{-31}$  for rs9486715. The protein encoded by this gene is expressed with activator of cAMP-responsive element modulator (CREM) (32). It is associated with CREM and confers a powerful transcriptional activation function (32). The locus was first reported by Anttila et al to be associated with migraine (13). Although replication failed in a Chinese population, the locus was replicated by Gormley et al in a large meta-analysis on migraine (17,33). Both the *LRP1* and *FHL5* genes are also candidate genes for cervical artery dissection suggesting vascular involvement in headache (18). Just as the *LRP1* region extends to the *STAT6* gene, the *FHL5* cluster extends to the *UFL1* gene.

The strongest association among the 14 newly proposed loci is located in the *LINC02210-CRHR1* gene. The corticotropin releasing hormone receptor 1 (*CRHR1*) gene has been reported to be associated with stress, panic disorder, neuroticism (34-36). This is matched with our genetic correlation results by LD hub suggesting headache and psychological traits share genetic architectures (see discussion below). The loci extends to the microtubule associated protein tau (*MAPT*) gene region. Other new identified loci included: calcium/calmodulin dependent protein kinase ID (*CAMK1D*), *NUFIP2*, *FMR1 interacting protein 2* (*NUFIP2*), *butyrophilin subfamily 2 member A2* (*BTN2A2*), *myosin IH* (*MYO1H*), *intraflagellar transport 81* (*IFT81*), *nucleolar protein 4 like* (*NOL4L*), *polypyrimidine tract binding protein 2* (*PTBP2*), *frataxin* (*FXN*), *autophagy related 13* (*ATG13*), *microtubule-actin crosslinking factor 1* (*MACF1*) and some intergenic regions. It is interesting to notice that GWAS have identified that *CAMK1D* and *MACF1* are involved in vascular disorders (hypertension and peripheral artery disease) (37,38), again supporting a vascular contribution to headache.

It is also interesting to note that the *STAT6* gene was the most significant in the gene analysis with headache and not the *LRP1* gene where the top SNP resides. These two genes are next to each other in the genome and have previously been associated with disorders related to the immune system such as food allergen sensitization and Sjogren's Disease (39,40). It is reported that IL4/STAT6 signalling activates neural stem cell proliferation and neurogenesis in zebrafish brain, which indicates the importance of the gene in neuron function (41). The gene sets analysis revealed that genes involved in the neurogenesis are associated with headache which is consistent with the tissue expression analysis.

Both tissue expression analysis on 30 general tissue types and 53 specific tissue types showed significant associations between brain tissues and headache, but not vascular tissues. This conclusion is different from the predominant theory of vascular aetiology for migraine since our results suggest that brain function is closely related to broadly-defined headache. Combining all results from GWAS and FUMA together, it is clear that both neuronal and vascular factors are involved in the headache mechanism.

Consistent with previous studies from our group, we found strong evidence to support a shared genetic aetiology of pain with psychological traits, indicating a vulnerability to depression and other negative mood states (42,43). In addition, we found a negative correlation with factors associated with a longer duration of education. Previous results have suggested that headache sufferers were more likely to have psychiatric disorders than healthy people (44). In addition, a shared genetic link between migraine and depression has been identified (45). Further studies are needed to demonstrate the nature of these correlations and whether any directional 'causal' inferences can be drawn.

The Q-Q plot suggested that there are residual confounding factors between headache cases and controls which have not been adjusted for. Those could be factors associated with psychiatric and vascular disorders. We also noted there is a 90 degree upswing (around

the  $P = 10^{-14}$  level). This is due to the fact that we have a cluster of 1,800 significant SNPs at this level (mainly located at the *LINC02210-CRHR1-MAPT* loci), which is over half of the total number of significant SNPs.

Using the CaTS power calculator, we had 80% power to identify SNP associations with a significance level of  $5 \times 10^{-8}$ , based on 74,461 cases and 149,312 controls, assuming an additive model, a minor disease allele frequency of 0.15, a genotypic relative risk of 1.05, and a prevalence of headache in the general population of 0.2 (a conservative assumption) (46).

In conclusion, we have identified 28 loci for broadly-defined headache as a single entity in a GWAS using the UK Biobank resource, including 14 loci that have previously been associated with migraine, and 14 loci that have not previously been associated with headache. This is the largest GWAS on headache in a single population so far. We also identified evidence that brain function is closely related to broadly-defined headache. In addition, we found several significant correlations with a number of psychological factors, suggesting that the genetic aetiology of headache may also be related to these traits.

## **Materials and Methods**

### **Participants and genetic information of participants**

The UK Biobank is a health research resource that aims to improve the prevention, diagnosis and treatment of a wide range of illnesses. The UK Biobank cohort recruited over 500,000 people aged between 40-69 years in 2006-2010 across the UK. Participants completed a detailed clinical, demographic, and lifestyle questionnaire, underwent clinical measures, provided biological samples (blood, urine and saliva) for future analysis, and agreed to have their health records accessed. The informed consent of all participants has been obtained. Details of the UK Biobank resource can be found at [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk).

UK Biobank received ethical approval from the National Health Service National Research Ethics Service (reference 11/NW/0382). The current analyses were conducted under approved UK Biobank data application number 4844.

The detailed methods of DNA extraction and quality control can be found at <http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/DNA-Extraction-at-UK-Biobank-October-2014.pdf>. Participants' DNA was genotyped by bespoke Affymetrix UK Biobank chips. Standard quality-control (QC) steps were performed by the Wellcome Trust Centre for Human Genetics at Oxford University. In this study, we used BGENIE (<https://jmarchini.org/bgenie/>) to be the main GWAS software and removed SNPs with INFO scores less than 0.1, with minor allele frequency less than 0.5%, or those that failed Hardy-Weinberg tests  $P < 10^{-6}$  (47). SNPs on the X and Y chromosomes and mitochondrial SNPs were excluded from analyses. The detailed QC steps can be found at <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580>.

In September 2017, the genetic information from 501,708 samples was released to UK Biobank project research collaborators. This included imputation with a reference set combining the UK10K haplotype and 1000 Genomes Phase 3 reference panels. The imputation details can be found at <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=157020>.

### **Phenotypic information on pain**

The UK Biobank participants were offered a pain-related questionnaire, which included the question: 'in the last month have you experienced any of the following that interfered with your usual activities?'. The options were: 1. Headache; 2. Facial pain; 3. Neck or shoulder pain; 4. Back pain; 5. Stomach or abdominal pain; 6. Hip pain; 7. Knee pain; 8. Pain all over the body; 9. None of the above; 10. Prefer not to say. Participants could select more than one option. (UK Biobank Questionnaire field ID: 6159)

The headache cases in this study were those who selected the 'Headache' option for the above question, regardless of whether they had selected other options.

The controls in this study were those who selected the 'None of the above' option.

## **Statistical analysis**

Standard Frequentist association tests using BGENIE was used to perform association studies adjusting for age, sex, BMI, 9 population principal components, genotyping arrays, and assessment centres. Gender difference between cases and controls was compared using chi-square testing. Age and BMI were compared using independent t testing in IBM SPSS 22 (IBM Corporation, New York). SNP associations were considered significant if they had a *P* value less than  $5 \times 10^{-8}$ .

SNP functional annotations were applied by the FUMA web application and a Manhattan plot was generated by R (48). R was also used to generate the corresponding Q-Q plot, a tool to evaluate differences between cases and controls caused by potential confounders.

The gene analysis and gene-set analysis were performed with MAGMA v1.6, which was integrated in FUMA (49). Both analyses were based on GWAS summary statistics. In gene analysis, summary statistics of SNPs are aggregated to the level of whole genes, testing the joint association of all SNPs in the gene with the phenotype. In gene-set analysis, individual genes are aggregated to groups of genes sharing certain biological, functional or other characteristics. This will provide insight into the involvement of specific biological pathways or cellular functions in the genetic aetiology of a phenotype. Tissue expression analysis was obtained from GTEx (<https://www.gtexportal.org/home/>) which was also integrated in FUMA.

In order to identify genetic correlations with other complex traits and diseases, we used linkage disequilibrium score regression through LD Hub v1.4.1 (available at <http://ldsc.broadinstitute.org/ldhub/>) (50). This web-tool uses individual SNP allele effect sizes and the average linkage disequilibrium in a region to estimate the bivariate genetic correlations of headache with 234 traits. Those with *P* values of  $2.1 \times 10^{-4}$  ( $0.05/234$ ) or less should be regarded as surviving Bonferroni adjustment for multiple testing.

## Figure legends:

Figure 1. the Manhattan plot of the GWAS on headache using the UK Biobank cohort

Figure 2. the tissue expression results on 53 specific tissue types by GTEx in FUMA

The line indicates statistical significance of  $P = 0.001$  (0.05/53)

## Competing Interests:

Ian Deary is a participant in UK Biobank. Other authors declare no conflict of interests.

## Acknowledgements/Funding

This work was supported by the DOLORisk project [EU Horizon2020, grant number: 633491], the STRADL project [Wellcome Trust, grant number: 104036/Z/14/Z], and the Centre for Cognitive Ageing and Cognitive Epidemiology [Medical Research Council and Biotechnology and Biological Sciences Research Council, grant number: MR/K026992/1]. We are grateful for support from the the Dr Mortimer And Theresa Sackler Foundation.

We would like to thank all participants of the UK Biobank cohort who have provided necessary genetic and phenotypic information.

## References

1. Boardman, H.F., Thomas, E., Croft, P.R. and Millson, D.S. (2003) Epidemiology of headache in an English district. *Cephalalgia*, **23**, 129-137.
2. Larner, A.J. (2006) Guidelines for primary headache disorders in primary care: an “intervention” study. *Headache. Care.*, **3**, 1-2.

3. Latinovic, R., Gulliford, M. and Ridsdale, L. (2006) Headache and migraine in primary care: consultation, prescription, and referral rates in a large population. *J. Neurol. Neurosurg. Psychiatry.*, **77**, 385-387.
4. Headache Classification Committee of the International Headache Society (IHS). (2013) The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia*, **33**, 629-808.
5. Stovner, L.j., Hagen, K., Jensen, R., Katsarava, Z., Lipton, R., Scher, A., Steiner, T. and Zwart, J.A. (2007) The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia*, **27**, 193-210.
6. Riesco, N., Cernuda-Morollón, E. and Pascual, J. (2007) Neuropeptides as a Marker for Chronic Headache. *Curr. Pain. Headache. Rep.*, **21**, 18.
7. Global Burden of Disease Study 2013 Collaborators. (2015) Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **386**, 743-800.
8. Steiner, T.J., Scher, A.I., Stewart, W.F., Kolodner, K., Liberman, J. and Lipton, R.B. (2003) The prevalence and disability burden of adult migraine in England and their relationships to age, gender and ethnicity. *Cephalalgia*, **23**, 519-527.
9. Silberstein, S.D. (2004) Migraine. *Lancet*, **363**, 381-391.
10. Russell, M., Levi, N. and Kaprio, J. (2007) Genetics of tension-type headache: a population based twin study. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.*, **144B**, 982-986.
11. Schürks, M. (2012) Genetics of migraine in the age of genome-wide association studies. *J. Headache. Pain.*, **13**, 1-9.
12. Anttila, V., Stefansson, H., Kallela, M., Todt, U., Terwindt, G.M., Calafato, M.S., Nyholt, D.R., Dimas, A.S., Freilinger, T., Müller-Myhsok, B. *et al.* (2010) Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. *Nat. Genet.*, **42**, 869-873.

13. Anttila, V., Winsvold, B.S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., Kallela, M., Malik, R., de Vries, B., Terwindt, G. *et al.* (2013) Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.*, **45**, 912-917.
14. Chasman, D.I., Schürks, M., Anttila, V., de Vries, B., Schminke, U., Launer, L.J., Terwindt, G.M., van den Maagdenberg, A.M., Fendrich, K., Völzke, H. *et al.* (2011) Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nat. Genet.*, **43**, 695-698.
15. Freilinger, T., Anttila, V., de Vries, B., Malik, R., Kallela, M., Terwindt, G.M., Pozo-Rosich, P., Winsvold, B., Nyholt, D.R., van Oosterhout, W.P. *et al.* (2012) Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat. Genet.*, **44**, 777-782.
16. Ligthart, L., de Vries, B., Smith, A.V., Ikram, M.A., Amin, N., Hottenga, J.J., Koelewijn, S.C., Kattenberg, V.M., de Moor, M.H., Janssens, A.C. *et al.* (2011) Meta-analysis of genome-wide association for migraine in six population-based European cohorts. *Eur. J. Hum. Genet.*, **19**, 901-907.
17. Gormley, P., Anttila, V., Winsvold, B.S., Palta, P., Esko, T., Pers, T.H., Farh, K.H., Cuenca-Leon, E., Muona, M., Furlotte, N.A. *et al.* (2016) Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.*, **48**, 856-866.
18. Gormley, P., Winsvold, B.S., Nyholt, D.R., Kallela, M., Chasman, D.I. and Palotie, A. (2016) Migraine genetics: from genome-wide association studies to translational insights. *Genome. Med.*, **8**, 86.
19. Scher, A.I., Stewart, W.F., Ricci, J.A. and Lipton, R.B. (2003) Factors associated with the onset and remission of chronic daily headache in a population-based study. *Pain*, **106**, 81-89.
20. Waldie, K.E. and Poulton, R. (2002) Physical and psychological correlates of primary headache in young adulthood: a 26 year longitudinal study. *J. Neurol. Neurosurg. Psychiatry.*, **72**, 86-92.



21. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium, Duncan, L. *et al.* (2015) An atlas of genetic correlations across human diseases and traits. *Nat. Genet.*, **47**, 1236-1241.
22. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291-295.
23. Deary, V., Hagenaars, S.P., Harris, S.E., Hill, W.D., Davies, G., Liewald, D.C., International Consortium for Blood Pressure GWAS, CHARGE Consortium Aging and Longevity Group, CHARGE Consortium Inflammation Group, McIntosh, A.M. *et al.* (2016) Genetic contributions to self-reported tiredness. *Mol. Psychiatry.*, doi: 10.1038/mp.2017.5.
24. Lyngberg, A.C., Rasmussen, B.K., Jørgensen, T. and Jensen, R. (2005) Has the prevalence of migraine and tension-type headache changed over a 12-year period? A Danish population survey. *Eur. J. Epidemiol.*, **20**, 243-249.
25. Lillis, A.P., Van Duyn, L.B., Murphy-Ullrich, J.E. and Strickland, D.K. (2008) LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. *Physiol. Rev.*, **88**, 887-918.
26. Etique, N., Verzeaux, L., Dedieu, S. and Emonard, H. (2013) LRP-1: a checkpoint for the extracellular matrix proteolysis. *Biomed. Res. Int.*, **2013**, 152163.
27. Kanekiyo, T. and Bu, G. (2014) The low-density lipoprotein receptor-related protein 1 and amyloid- $\beta$  clearance in Alzheimer's disease. *Front. Aging. Neurosci.*, **6**, 93.
28. Llorente-Cortés, V. and Badimon, L. (2005) LDL receptor-related protein and the vascular wall: implications for atherothrombosis. *Arterioscler. Thromb. Vasc. Biol.*, **25**, 497-504.

29. Song, H., Li, Y., Lee, J., Schwartz, A.L. and Bu, G. (2009) Low-density lipoprotein receptor-related protein 1 promotes cancer cell migration and invasion by inducing the expression of matrix metalloproteinases 2 and 9. *Cancer. Res.*, **69**, 879-886.
30. Fan, X., Wang, J., Fan, W., Chen, L., Gui, B., Tan, G. and Zhou, J. (2014) Replication of migraine GWAS susceptibility loci in Chinese Han population. *Headache*, **54**, 709-715.
31. Ghosh, J., Pradhan, S. and Mittal, B. (2013) Genome-wide-associated variants in migraine susceptibility: a replication study from North India. *Headache*, **53**, 1583-1594.
32. Fimia, G.M., De Cesare, D. and Sassone-Corsi, P. (1999) CBP-independent activation of CREM and CREB by the LIM-only protein ACT. *Nature*, **398**, 165-169.
33. Lin, Q.F., Fu, X.G., Yao, L.T., Yang, J., Cao, L.Y., Xin, Y.T., Hou, J.X., Ye, L.F. and Huang, G.B. (2015) Association of genetic loci for migraine susceptibility in the she people of China. *J. Headache. Pain.*, **16**, 553.
34. Henckens, M.J., Deussing, J.M. and Chen, A. (2016) Region-specific roles of the corticotropin-releasing factor-urocortin system in stress. *Nat. Rev. Neurosci.*, **17**, 636-651.
35. Weber, H., Richter, J., Straube, B., Lueken, U., Domschke, K., Schartner, C., Klauke, B., Baumann, C., Pané-Farré, C., Jacob, C.P. *et al.* (2016) Allelic variation in CRHR1 predisposes to panic disorder: evidence for biased fear processing. *Mol. Psychiatry.*, **21**, 813-822.
36. DeYoung, C.G., Cicchetti, D. and Rogosch, F.A. (2011) Moderation of the association between childhood maltreatment and neuroticism by the corticotropin-releasing hormone receptor 1 gene. *J. Child. Psychol. Psychiatry.*, **52**, 898-906.
37. Frau, F., Zaninello, R., Salvi, E., Ortu, M.F., Braga, D., Velayutham, D., Argiolas, G., Fresu, G., Troffa, C., Bulla, E. *et al.* (2014) Genome-wide association study identifies CAMKID variants involved in blood pressure response to losartan: the SOPHIA study. *Pharmacogenomics*, **15**, 1643-1652.

38. Ward-Caviness, C.K., Neas, L.M., Blach, C., Haynes, C.S., LaRocque-Abramson, K., Grass, E., Dowdy, E., Devlin, R.B., Diaz-Sanchez, D., Cascio, W. *et al.* (2016) Genetic Variants in the Bone Morphogenic Protein Gene Family Modify the Association between Residential Exposure to Traffic and Peripheral Arterial Disease. *PLoS. One.*, **11**, e0152670.
39. Hancock, D.B., Romieu, I., Chiu, G.Y., Sienna-Monge, J.J., Li, H., Estela Del Rio-Navarro, B. and London, S.J. (2012) STAT6 and LRP1 polymorphisms are associated with food allergen sensitization in Mexican children. *J. Allergy. Clin. Immunol.*, **129**, 1673-1676.
40. Johar, A.S., Mastronardi, C., Rojas-Villarraga, A., Patel, H.R., Chuah, A., Peng, K., Higgins, A., Milburn, P., Palmer, S., Silva-Lara, M.F. *et al.* (2015) Novel and rare functional genomic variants in multiple autoimmune syndrome and Sjögren's syndrome. *J. Transl. Med.*, **13**, 173.
41. Bhattarai, P., Thomas, A.K., Cosacak, M.I., Papadimitriou, C., Mashkaryan, V., Froc, C., Reinhardt, S., Kurth, T., Dahl, A., Zhang, Y. *et al.* (2016) IL4/STAT6 Signaling Activates Neural Stem Cell Proliferation and Neurogenesis upon Amyloid- $\beta$ 42 Aggregation in Adult Zebrafish Brain. *Cell. Rep.*, **17**, 941-948.
42. McIntosh, A.M., Hall, L.S., Zeng, Y., Adams, M.J., Gibson, J., Wigmore, E., Hagenaaars, S.P., Davies, G., Fernandez-Pujals, A.M., Campbell, A.I. *et al.* (2016) Genetic and Environmental Risk for Chronic Pain and the Contribution of Risk Variants for Major Depressive Disorder: A Family-Based Mixed-Model Analysis. *PLoS. Med.*, **13**, e1002090.
43. van Hecke, O., Hocking, L.J., Torrance, N., Campbell, A., Padmanabhan, S., Porteous, D.J., McIntosh, A.M., Burri, AV., Tanaka, H., Williams, F.M. *et al.* (2017) Chronic pain, depression and cardiovascular disease linked through a shared genetic predisposition: Analysis of a family-based cohort and twin study. *PLoS. One.*, **12**, e0170653.

44. Mitsikostas, D.D. and Thomas, A.M. (1999) Comorbidity of headache and depressive disorders. *Cephalalgia*, **19**, 211-217.
45. Stam, A.H., de Vries, B., Janssens, A.C., Vanmolkot, K.R., Aulchenko, Y.S., Henneman, P., Oostra, B.A., Frants, R.R., van den Maagdenberg, A.M., Ferrari, M.D. *et al.* (2010) Shared genetic factors in migraine and depression: evidence from a genetic isolate. *Neurology*, **74**, 288-294.
46. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209-213.
47. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2017) Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*, doi: <https://doi.org/10.1101/166298>.
48. Watanabe, K., Taskesen, E., van Bochoven, A. and Posthuma, D. (2017) FUMA: Functional mapping and annotation of genetic associations. *bioRxiv*, <https://doi.org/10.1101/110023>.
49. de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS. Comput. Biol.*, **11**, e1004219.
50. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, *et al.* (2017) LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, **33**, 272-279.

**Table 1.** Clinical characteristics of headache cases (74,461) and controls (149,312)

| Covariates               | Cases          | Controls       | <i>P</i> |
|--------------------------|----------------|----------------|----------|
| Sex (male:female)        | 27,350: 47,111 | 71,480: 77,832 | <0.001   |
| Age (years)              | 54.38 (7.95)   | 56.93 (7.97)   | <0.001   |
| BMI (kg/m <sup>2</sup> ) | 27.50 (5.05)   | 26.66 (4.30)   | <0.001   |

BMI: body mass index

A chi-square test was used to test the difference of gender frequency between cases and controls and an independent t test was used for other covariates.

Continuous covariates were presented as mean (standard deviation).

**Table 2.** Summary of the 28 loci associated with broadly defined headache

| Locus rank | Gene                        | Chromosome | Lead SNP   | P                      | Effective allele | MAF  | Beta    | SE     | Reported in references? |
|------------|-----------------------------|------------|------------|------------------------|------------------|------|---------|--------|-------------------------|
| 1          | <i>LRP1-STAT6-SDR9C7</i>    | 12         | rs11172113 | $4.92 \times 10^{-47}$ | C                | 0.38 | -0.0202 | 0.0014 | 13-15, 17               |
| 2          | <i>FHL5-UFL1</i>            | 6          | rs9486715  | $2.58 \times 10^{-31}$ | C                | 0.35 | 0.0171  | 0.0015 | 13, 17                  |
| 3          | <i>TRPM8-HJURP</i>          | 2          | rs2362290  | $6.25 \times 10^{-16}$ | A                | 0.17 | -0.0141 | 0.0017 | 13-15, 17               |
| 4          | <i>PHACTR1</i>              | 6          | rs9349379  | $3.59 \times 10^{-15}$ | G                | 0.40 | -0.0110 | 0.0014 | 13, 15, 17              |
| 5          | <i>LINC02210-CRHR1-MAPT</i> | 17         | rs77804065 | $5.87 \times 10^{-15}$ | T                | 0.25 | 0.0130  | 0.0017 | New loci                |
| 6          | <i>Intergenic</i>           | 1          | rs12740679 | $3.66 \times 10^{-12}$ | G                | 0.26 | 0.0110  | 0.0016 | New loci                |
| 7          | <i>Intergenic</i>           | 10         | rs78438709 | $8.57 \times 10^{-12}$ | G                | 0.07 | -0.0186 | 0.0027 | 17                      |
| 8          | <i>MEF2D</i>                | 1          | rs1050316  | $1.54 \times 10^{-11}$ | T                | 0.33 | -0.0098 | 0.0015 | 13, 15, 17              |
| 9          | <i>ASTN2</i>                | 9          | rs17220352 | $9.42 \times 10^{-11}$ | G                | 0.25 | 0.0104  | 0.0016 | 13, 15, 17              |
| 10         | <i>Intergenic</i>           | 3          | rs34097149 | $1.33 \times 10^{-10}$ | C                | 0.03 | -0.0297 | 0.0046 | 17                      |
| 11         | <i>Intergenic</i>           | 6          | rs9490318  | $1.72 \times 10^{-10}$ | T                | 0.14 | 0.0124  | 0.0019 | 17                      |
| 12         | <i>Intergenic</i>           | 1          | rs12134493 | $1.93 \times 10^{-10}$ | A                | 0.11 | 0.0136  | 0.0021 | 13, 17                  |
| 13         | <i>Intergenic</i>           | 18         | rs4941139  | $2.22 \times 10^{-10}$ | C                | 0.32 | 0.0093  | 0.0015 | New loci                |
| 14         | <i>CAMK1D</i>               | 10         | rs2895526  | $4.61 \times 10^{-10}$ | A                | 0.48 | -0.0086 | 0.0014 | New loci                |
| 15         | <i>PRDM16</i>               | 1          | rs56304645 | $5.74 \times 10^{-10}$ | T                | 0.26 | 0.0102  | 0.0016 | 13, 14 17               |
| 16         | <i>NUFIP2</i>               | 17         | rs8614     | $4.25 \times 10^{-9}$  | A                | 0.19 | 0.0105  | 0.0018 | New loci                |
| 17         | <i>Intergenic</i>           | 12         | rs10774231 | $4.55 \times 10^{-9}$  | C                | 0.42 | -0.0082 | 0.0014 | 17                      |
| 18         | <i>MRVI1</i>                | 11         | rs4909945  | $5.13 \times 10^{-9}$  | C                | 0.30 | 0.0087  | 0.0015 | 17                      |
| 19         | <i>BTN2A2</i>               | 6          | rs2072806  | $5.30 \times 10^{-9}$  | G                | 0.12 | -0.0121 | 0.0021 | New loci                |
| 20         | <i>Intergenic</i>           | 1          | rs7555006  | $5.87 \times 10^{-9}$  | G                | 0.44 | 0.0081  | 0.0014 | New loci                |
| 21         | <i>MYO1H</i>                | 12         | rs6606710  | $8.44 \times 10^{-9}$  | C                | 0.42 | 0.0084  | 0.0015 | New loci                |
| 22         | <i>IFT81</i>                | 12         | rs7300001  | $8.86 \times 10^{-9}$  | G                | 0.10 | -0.0144 | 0.0025 | New loci                |
| 23         | <i>NOL4L</i>                | 20         | rs1555132  | $1.16 \times 10^{-8}$  | A                | 0.34 | 0.0083  | 0.0015 | New loci                |
| 24         | <i>CFDP1</i>                | 16         | rs1011121  | $1.46 \times 10^{-8}$  | G                | 0.43 | -0.0080 | 0.0014 | 17                      |
| 25         | <i>PTBP2</i>                | 1          | rs3748784  | $1.75 \times 10^{-8}$  | G                | 0.47 | -0.0078 | 0.0014 | New loci                |
| 26         | <i>FXN</i>                  | 9          | rs4596713  | $2.30 \times 10^{-8}$  | T                | 0.41 | -0.0078 | 0.0014 | New loci                |
| 27         | <i>ATG13</i>                | 11         | rs56349329 | $3.69 \times 10^{-8}$  | A                | 0.16 | -0.0104 | 0.0019 | New loci                |
| 28         | <i>MACF1</i>                | 1          | rs2036465  | $4.01 \times 10^{-8}$  | C                | 0.20 | -0.0093 | 0.0017 | New loci                |

MAF: minor allele frequency

SE: standard error

## Supporting information

Supplementary Fig. S1: The Q-Q plot of the GWAS on headache using UK Biobank

Supplementary Table S1: A summary of 3343 GWAS significant SNPs

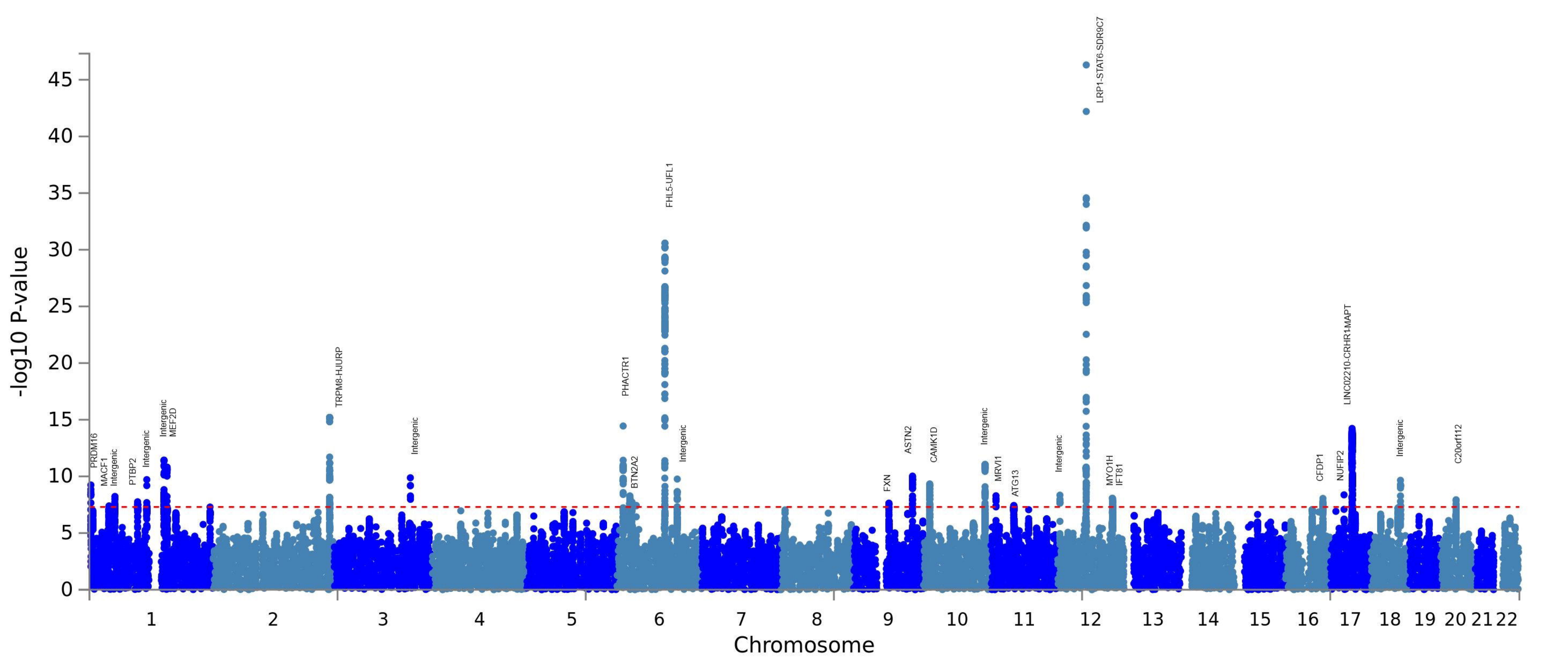
Supplementary Table S2: Significant genes based on the gene analysis by MAGMA integrated in FUMA

Supplementary Table S3: Top 10 gene sets results by MAGMA integrated in FUMA

Supplementary Table S4: Tissue expression analysis on 30 general tissue types

Supplementary Table S5: Tissue expression analysis on 53 specific tissue types

Supplementary Table S6: Genetic correlations between headache and other phenotypes which reached significance of 0.05/234





-log<sub>10</sub> P-value

