Genome-Wide Association Study Identifies Risk Loci for Cluster Headache

Emer O'Connor, MB, BCH, BAO,[1,2,3†] Carmen Fourier, MSc,[4†] Caroline Ran, PhD,[4†] Prasanth Sivakumar, PhD,[1] Franziska Liesecke, PhD,[4] Laura Southgate, PhD,[5,6] Aster V. E. Harder, MD,[7,8] Lisanne S. Vijfhuizen, BSc,[8] Janice Yip, MSc,[1] Nicola Giffin, MD,[9] Nicholas Silver, PhD,[10] Fayyaz Ahmed, PhD,[11] Isabel C. Hostettler, MD,[1] Brendan Davies, MD,[12] M. Zameel Cader, DPhil,[13] Benjamin S. Simpson, MSc,[14] Roisin Sullivan, MSc,[1] Stephanie Efthymiou, MSc,[1] Joycee Adebimpe,[1] Olivia Quinn,[1] Ciaran Campbell, BA,[15] Gianpiero L. Cavalleri, PhD,[15] Michail Vikelis, PhD,[16] Tim Kelderman, MD,[17] Koen Paemeleire, PhD,[17] Emer Kilbride, MB, BCH, BAO,[18] Lou Grangeon, MD,[3,19] Susie Lagrata, MSc,[3] Daisuke Danno, MD,[3] Richard Trembath, PhD,[6] Nicholas W. Wood, PhD,[1,2] Ingrid Kockum, PhD,[20] Bendik S. Winsvold, PhD,[21,22,23] Anna Steinberg, PhD,[20] Christina Sjöstrand, PhD,[20] Elisabet Waldenlind, PhD,[20] Jana Vandrovcova, PhD,[1†] Henry Houlden, PhD,[1†] Manjit Matharu, PhD,[3†] and Andrea Carmine Belin, PhD[4†]

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.26150

Received Sep 25, 2020, and in revised form Jun 26, 2021. Accepted for publication Jun 28, 2021.

Address correspondence to Dr Belin, Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden, E-mail: andrea.carmine.belin@ki.se; and Dr Houlden, Department of Neuromuscular Disease, UCL Institute of Neurology, London WC1N 3BG, UK, E-mail: h.houlden@ucl.ac.uk

[†]E.O., C.F., C.R., J.V., H.H., M.M., and A.C.B. contributed equally.

From the [1]Department of Neuromuscular Diseases, Institute of Neurology, University College London, London, UK; [2]Neurogenetics Laboratory, Institute of Neurology, University College London, London, UK; [3]Headache and Facial Pain Group, University College London Queen Square Institute of Neurology and National Hospital for Neurology and Neurosurgery, London, UK; [4]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden; [5]Molecular and Clinical Sciences Research Institute, St George's, University of London, London, UK; [6]Department of Medical & Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London, London, UK; [7]Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands; [8]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; [9]Neurology Department, Royal United Hospital, Bath, UK; [10]Walton Centre, Liverpool, UK; [11]Department of Neurology, Hull Royal Infirmary, Hull, UK; [12]Department of Neurology, University Hospital North Midlands National Health Service Trust, Stoke-on-Trent, UK; [13]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; [14]Division of Surgery and Interventional Science, University College London, London, UK; [15]Science Foundation Ireland FutureNeuro Research Centre, Royal College of Surgeons, Ireland School of Pharmacy and Biomolecular Sciences, Royal College of Surgeons in Ireland Dublin, Dublin, Ireland; [16]Glyfada Headache Clinic, Glyfada, Greece; [17]Department of Neurology, Ghent University Hospital, Ghent, Belgium; [18]University College Hospital London, London, UK; [19]Department of Neurology, Rouen University Hospital, Rouen, France; [20]Division of Neurology, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden; [21]Department of Research, Innovation, and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway; [22]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway; and [23]Department of Neurology, Oslo University Hospital, Oslo, Norway

Additional supporting information can be found in the online version of this article.

Objective: This study was undertaken to identify susceptibility loci for cluster headache and obtain insights into relevant disease pathways.
Methods: We carried out a genome-wide association study, where 852 UK and 591 Swedish cluster headache cases were compared with 5,614 and 1,134 controls, respectively. Following quality control and imputation, single variant association testing was conducted using a logistic mixed model for each cohort. The 2 cohorts were subsequently combined in a merged analysis. Downstream analyses, such as gene-set enrichment, functional variant annotation, prediction and pathway analyses, were performed.
Results: Initial independent analysis identified 2 replicable cluster headache susceptibility loci on chromosome 2. A merged analysis identified an additional locus on chromosome 1 and confirmed a locus significant in the UK analysis on chromosome 6, which overlaps with a previously known migraine locus. The lead single nucleotide polymorphisms were rs113658130 ($p = 1.92 \times 10^{-17}$, odds ratio [OR] = 1.51, 95% confidence interval [CI] = 1.37–1.66) and rs4519530 ($p = 6.98 \times 10^{-17}$, OR = 1.47, 95% CI = 1.34–1.61) on chromosome 2, rs12121134 on chromosome 1 ($p = 1.66 \times 10^{-8}$, OR = 1.36, 95% CI = 1.22–1.52), and rs11153082 ($p = 1.85 \times 10^{-8}$, OR = 1.30, 95% CI = 1.19–1.42) on chromosome 6. Downstream analyses implicated immunological processes in the pathogenesis of cluster headache.
Interpretation: We identified and replicated several genome-wide significant associations supporting a genetic predisposition in cluster headache in a genome-wide association study involving 1,443 cases. Replication in larger independent cohorts combined with comprehensive phenotyping, in relation to, for example, treatment response and cluster headache subtypes, could provide unprecedented insights into genotype–phenotype correlations and the pathophysiological pathways underlying cluster headache.

ANN NEUROL 2021;00:1–10

Cluster headache (CH) is a rare, debilitating disorder with an estimated prevalence of 1 in 1,000 world-wide, with higher rates in northern countries, further from the equator.[1] It presents with unilateral pain distributed along the trigeminal nerve's first branch.[2] It is clearly distinct from other headache disorders based on attack duration, prominent ipsilateral cranial autonomic features, restlessness/agitation, and response to specific treatments.[3] Concomitant migraine can occur, but the reported frequency varies.[4,5] Some 85% of CH patients experience attack periods interspersed with attack-free periods of at least 3 months per year (episodic CH). The remainder have chronic CH with limited remissions. CH exhibits heritability evidenced by familial aggregation and cases of concordance among monozygotic twins.[6] In familial cases, segregation analysis predominantly shows an autosomal dominant inheritance model with reduced penetrance.[6]

Attempts to determine the underlying genetic architecture include candidate association studies of genes with a putative pathogenic role in CH. The pathophysiology remains unclear; a neurovascular process involving the trigeminovascular system, trigeminal autonomic reflex, and posterior hypothalamus is hypothesized.[7] Functional imaging studies observed activation of the ipsilateral inferior hypothalamic gray matter in CH attacks.[8] These findings, together with the circadian periodicity of CH, influenced selection of candidate genes. Unfortunately, reported associations with, for example, *HCRTR2*, *CLOCK*, and *ADH4*, presently lack replicability.[6]

Genome-wide association studies (GWASs) have provided insight into many neurological disorders, including migraine.[9] A large migraine meta-analysis yielded associations suggestive of vascular and neuronal mechanisms.[9] To date, there is 1 GWAS on CH, involving 99 cases and 360 controls.[10] It lacked power to detect loci of genome-wide significance (GWS), and suggested associations were not reproducible.[10,11] This present study aims to identify novel genetic risk variants for CH by performing a GWAS.

Subjects and Methods

Participant Recruitment and Phenotyping

This is a multicenter study comprising a GWAS of 2 independent cohorts sourced from specialized headache clinics; one from the UK and one from Sweden (Supplementary Table S1).

UK Cohort. Recruitment occurred between 2006 and 2018, starting at the National Hospital for Neurology and Neurosurgery, London, UK. Ethical approval was obtained for 4 additional UK sites (RAC#2060008 and University College London Hospitals 04/N034). All were specialist-led headache clinics. A diagnosis was made using the International Classification of Headache Disorders-3beta edition (ICHD-3b).[2] Review by 2 independent neurologists was required if 1 ICHD-3b criterion was not met. Control genotype data consisted of a cohort of UK individuals without headache (n = 463), the 1958 birth cohort (n = 2,699) from the Wellcome Trust Case Control Consortium (WTCCC), and the National Blood Survey cohort (NBS; n = 2,501).

Swedish Cohort. Recruitment occurred between 2014 and 2017 at the Karolinska University Hospital neurology clinic, Stockholm, Sweden. Ethical approval was obtained from the Stockholm Regional Ethical Review Board (registration number 2014/656–31). Patients (n = 643) fulfilling the ICHD-3b diagnostic criteria were included.[2] Diagnosis was confirmed by headache specialists and through medical records. Additional information was derived using a diagnostic questionnaire.

Control genotype data were obtained from neurologically healthy controls from the Immunomodulation and Multiple Sclerosis Epidemiology study (IMSE, https://ki.se/cns/imse) (n = 1,299).

DNA Extraction and Genotyping

Participants provided blood or saliva for DNA extraction. UK cases were genotyped at the Human Genotyping Facility, the Netherlands, using the Illumina (San Diego, CA) Infinium 24v1.0 Global Screening Array (GSA). Genotyping for the WTCCC and NBS controls used the custom 1.2 M Illumina chip. Swedish cases were genotyped using the SNP&SEQ Technology Platform, Sweden and controls at deCODE, Iceland using the GSA.

Data Processing and Quality Control

Raw IDAT files were processed using GenomeStudio (Illumina). PLINK and Peddy software was used for data quality control (QC).[12,13] The QC procedure was performed according to standard guidelines, and details for each study are summarized in Supplementary Table S2.

Imputation

The HRC/1KG imputation preparation and checking tool was used to identify errors related to strand, reference, and alternate allele assignments, and allele frequency differences (>0.2) relative to the Haplotype Reference Consortium panel v1.1 (HRCv1.1).[14] Estimated haplotypes were phased using Eagle v2.3, and imputation was performed on the Michigan server using HRCv1.1.[15,16] Monomorphic single nucleotide polymorphisms (SNPs) or those with an imputation quality score R^2 of < 0.3 or minor allele frequency (MAF) < 0.01 were excluded. Because of the smaller sample size, results for the Swedish-only results were filtered for MAF < 0.05.

Single Variant Association Testing

Single variant association testing was performed using Scalable and Accurate Implementation of Generalized Mixed Model (SAIGE),[17] fitting a null logistic mixed model. Imputed genotype data were used for all markers, and only autosomal SNPs were analyzed. In the UK cohort, principal component analysis (PCA) axes, generated with PLINK to identify population stratification, were used as covariates.[12] PCA axes 1–6 and sex were used as covariates in the Swedish cohort. In the merged analysis, we used separately imputed data from UK and Sweden, which were again subjected to a PCA after merging. The first 10 axes were used as covariates for combined SAIGE analysis. In each linkage disequilibrium (LD) cluster of SNPs containing GWS hits, the SNP with the lowest p value in case versus control comparisons was defined as the lead SNP. The Manhattan and Q-Q plots from association tests were created using R v3.6.2, and regional association plots were created with LocusZoom.[18,19] Downstream analysis was conducted using R v3.6.2 unless otherwise stated.[18]

Functional Variant Annotation and Prediction

Annotation and functional consequence prediction of disease-associated SNPs were conducted using FUMA v1.3.6a and the Variant Annotation Integrator tool.[20,21] The selected lead SNP from each associated locus was annotated. Utilized metrics included alternate allele frequency by population; Variant Effect Predictor annotation, including the Combined Annotation Dependent Depletion (CADD) score; and Genomic Evolutionary Rate Profiling (GERP) score annotation.[22,23] CADD scores of >10 are predicted to be the 10% most deleterious possible substitutions in the human genome, of >20 are predicted to be the 1%, and of >30 are predicted to be the 0.1%. GERP score ranges from −12.3 to 6.17, where higher scores indicate higher evolutionary constraint and a score greater than 2 can be considered constrained.

Gene-Based Association Testing

Gene-based association analysis was conducted using MAGMA through FUMA v1.3.6a.[20,24] To identify candidate genes associated with CH, the mean association of all SNPs within a gene was calculated, accounting for LD. Gene windows were extended 35kb upstream and 10kb downstream of the annotated gene start and end sites to include regulatory regions.

Gene Expression and Expression Quantitative Trait Locus Analysis

Gene expression was determined using GTEx v8; spatiotemporal analysis of gene expression was accessed using the Human Brain Transcriptome dataset and a cell type–specific Brain RNA-seq atlas.[25–27] Gene mapping and expression quantitative trait loci (eQTLs) of interest were determined through FUMA using default settings and eQTL data from the brain, spleen, and vascular and immunological tissue based on GTEx v8 and the eQTL catalogue.[20] Only cis-eQTLs with multiple testing correction false discovery rate (FDR) < 0.05 were included in downstream analysis.

Pathway Analysis

Pathway analysis was conducted using regions surrounding identified lead SNPs. In each case, protein-coding genes within a 1Mb window on either side of the lead SNP were used as input for the gprofiler2 R package with the default background of annotated genes.[28] Pathways with a multiple testing-adjusted p value < 0.05 (FDR) underwent further analysis. Sources for predefined pathways and complexes included Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome (REAC).[29–31]

Genetic Colocalization Analysis

Bayesian colocalization analysis was used to determine shared causal regions or lead variants between CH and the migraine GWAS from the UK Biobank GWAS database, publicly available at http://www.nealelab.is/uk-biobank, using the R package coloc.[32] Causal regions were defined as the range of positions of SNPs between 2 recombination hotspots. Coloc reports posterior probability for each of 5 hypotheses tested: H0, neither CH nor migraine had a genetic association within the tested region; H1, only CH had a genetic association within the tested region; H2, only migraine had a genetic association within the tested region; H3, both CH and migraine had a genetic association within the tested region, but did not share causal variants; and H4, both CH and migraine shared a

single causal variant with the tested region. High posterior probability for H4 supports colocalization of the signals.

Results

The clinical phenotypes of CH cases within each cohort are summarized in Supplementary Table S1.

UK Cohort. A total of 852 cases and 5,614 controls were included after QC. Eighty-six SNPs passed the GWS threshold ($p < 5 \times 10^{-8}$) and clustered in 3 independent loci. Two loci were located on chromosome 2 (chr2q13, chr2q33). The lead SNPs in these regions were rs4519530 ($p = 2.49 \times 10^{-8}$, odds ratio [OR] = 1.39,

TABLE. Summary of Lead SNPs at Each Locus Associated with Cluster Headache

Locus	*LINC01705/DUSP10*	*MERTK*	*LINC01877/SATB2*	*FHL5*
Chr	1q41	2q13	2q33	6q16
UK analysis				
rsID	rs6687758	rs4519530	rs113658130	rs9386670
Variant details	1:222164948: A/G	2:112759182: C/T	2:200504209: C/T	6:97060688: C/A
EA	G	C	C	A
OR (95% CI)	1.27 (1.12–1.45)	1.39 (1.24–1.55)	1.63 (1.39–1.90)	1.40 (1.25–1.57)
p	3.29×10^{-4}	2.49×10^{-8}	7.39×10^{-10}	1.41×10^{-8}
EAF cases	0.23	0.67	0.76	0.38
EAF controls	0.19	0.6	0.7	0.31
Swedish analysis				
rsID	rs6671564 ($r^2 = 0.09$ with rs6687758)	rs72825689 ($r^2 = 0.03$ with rs4519530)	rs4675692 ($r^2 = 0.6$ with rs113658130)	rs4098006 ($r^2 = 0.003$ with rs9386670)
Variant details	1:222014257: G/A	2:112790104: T/C	2:200449911: G/A	6:97084953: G/A
EA	A	C	G	A
OR (95% CI)	1.39 (1.19–1.63)	2.82 (1.98–4.03)	1.61 (1.37–1.90)	1.28 (1.09–1.51)
p	4.90×10^{-5}	1.07×10^{-8}	1.22×10^{-8}	2.72×10^{-3}
EAF cases	0.50	0.08	0.67	0.50
EAF controls	0.43	0.04	0.56	0.46
Combined analysis				
rsID	rs12121134 ($r^2 = 1$ with rs6687758)	rs4519530	rs113658130	rs11153082 ($r^2 = 0.98$ with rs9386670)
Variant details	1:222194880: C/T	2:112759182: C/T	2:200504209: C/T	6:97059666: A/G
EA	T	C	C	G
OR (95% CI)	1.36 (1.22–1.52)	1.47 (1.34–1. 61)	1.51 (1.37–1.66)	1.30 (1.19–1.42)
p	1.66×10^{-8}	6.98×10^{-17}	1.92×10^{-17}	1.85×10^{-8}
EAF cases	0.25	0.71	0.75	0.39
EAF controls	0.19	0.62	0.69	0.33

Chr = chromosome; CI = confidence interval; *DUSP10* = dual specificity phosphatase 10; EA = effect allele; EAF = effect allele frequency; *FHL5* = four and a half LIM domains 5; *LINC01705* = long intergenic non-protein coding RNA 1705; *LINC01877* = long intergenic non-protein coding RNA 1877; *MERTK* = MER proto-oncogene tyrosine kinase; OR = odds ratio; rsID = reference SNP ID number; *SATB2* = SATB homeobox 2; SNP = single nucleotide polymorphism.

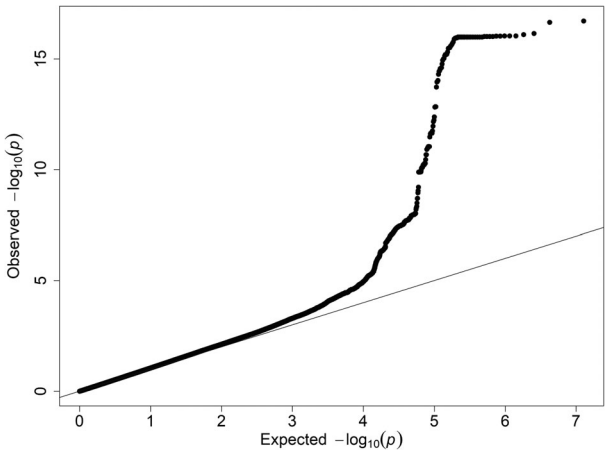

FIGURE 1: Quantile-quantile (Q-Q) plot of combined genome-wide association study (GWAS) showing SNP *p* values in GWAS analysis versus expected *p* values. The straight line in the Q-Q plot indicates the distribution of single nucleotide polymorphisms under the null hypothesis.

95% confidence interval [CI] = 1.24–1.55; 2q13) and rs113658130 ($p = 7.39 \times 10^{-10}$, OR = 1.63, 95% CI = 1.39–1.90; 2q33). An additional locus on chromosome 6 also reached GWS, with lead SNP rs9386670 ($p = 1.41 \times 10^{-8}$, OR = 1.4, 95% CI = 1.25–1.57; Table).

Swedish Cohort. A total of 591 cases and 1,134 controls were included after QC. Fifty-three SNPs reached GWS threshold ($p < 5 \times 10^{-8}$). The 2 independent loci located on chromosome 2 identified in the UK association analysis (chr2q13, chr2q33) were replicated. The lead SNPs in these regions included rs72825689 ($p = 1.07 \times 10^{-8}$, OR = 2.82, 95% CI = 1.98–4.03) and rs4675692 ($p = 1.22 \times 10^{-8,}$ OR = 1.61, 95% CI = 1.37–1.90), respectively (see Table). The third locus on chromosome 6 did not reach genome-wide significance.

UK–Swedish Combined Results. Following QC, there were 1,443 cases and 6,748 controls of European ancestry. A genomic inflation factor of 1.06 suggested some occurrence of population stratification between Sweden and the UK (Fig 1).

The combined association analysis identified 4 loci with GWS ($p < 5 \times 10^{-8}$; Fig 2, Table). Two loci on chromosome 2 had significant independent association with CH in both cohorts. The locus with the strongest association, 2q33 with lead SNP rs113658130 ($p = 1.92 \times 10^{-17}$, OR = 1.51, 95% CI = 1.37–1.66), is located in a long intergenic noncoding RNA *LINC01877*. The closest protein coding gene is *SATB2* (SATB homeobox 2; Fig 3, Table). The second locus is represented by lead SNP rs4519530; it is an intronic variant at 2q13, and it is in the *MERTK* (MER proto-oncogene, tyrosine kinase) gene ($p = 6.98 \times 10^{-17}$, OR = 1.47, 95% CI = 1.34–1.61). The third locus corresponds to the one identified on 6q16 in the UK cohort, lead SNP rs11153082 ($p = 1.85 \times 10^{-8}$, OR = 1.30, 95% CI = 1.19–1.42). It is located in the *FHL5* (four and a half LIM domains 5) gene. A new locus was identified on chromosome 1q41 with lead SNP rs12121134 ($p = 1.66 \times 10^{-8}$, OR = 1.36, 95% CI = 1.22–1.52). This locus has a p value $< 10^{-3}$ in the UK respective to the Swedish cohort ($p = 3.29 \times 10^{-4}$ and $p = 4.90 \times 10^{-5}$). This region does not contain any known genes. The closest gene is *LINC01705*, a long noncoding RNA, located 11kb upstream from rs12121134, and the nearest coding gene is *DUSP10* (dual specificity phosphatase 10). All of the lead SNPs were imputed and well above the imputation quality score R^2 threshold.

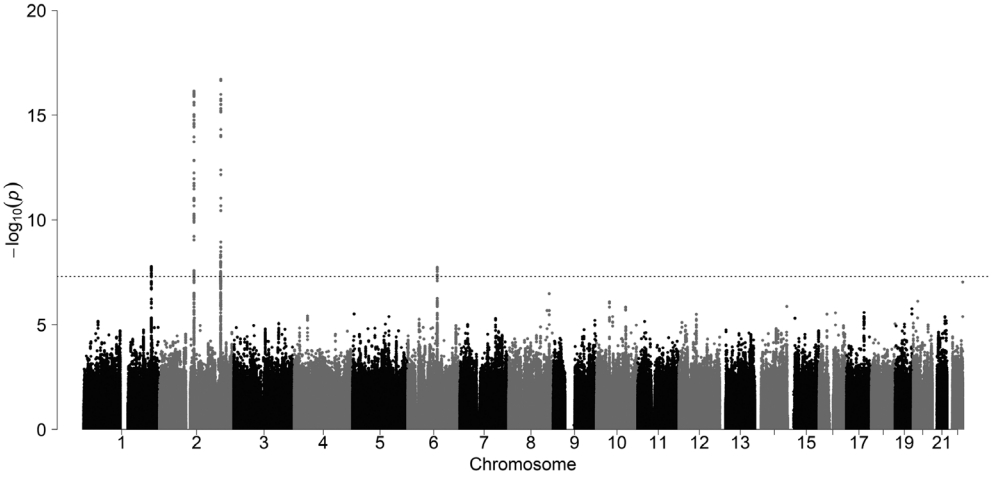

FIGURE 2: Manhattan plot of combined cohort cluster headache (CH) association analysis for 1,443 CH cases and 6,748 controls highlighting genome-wide significant associations at chromosomes 1, 2, and 6. The broken gray line is indicative of the threshold for genome-wide significance.

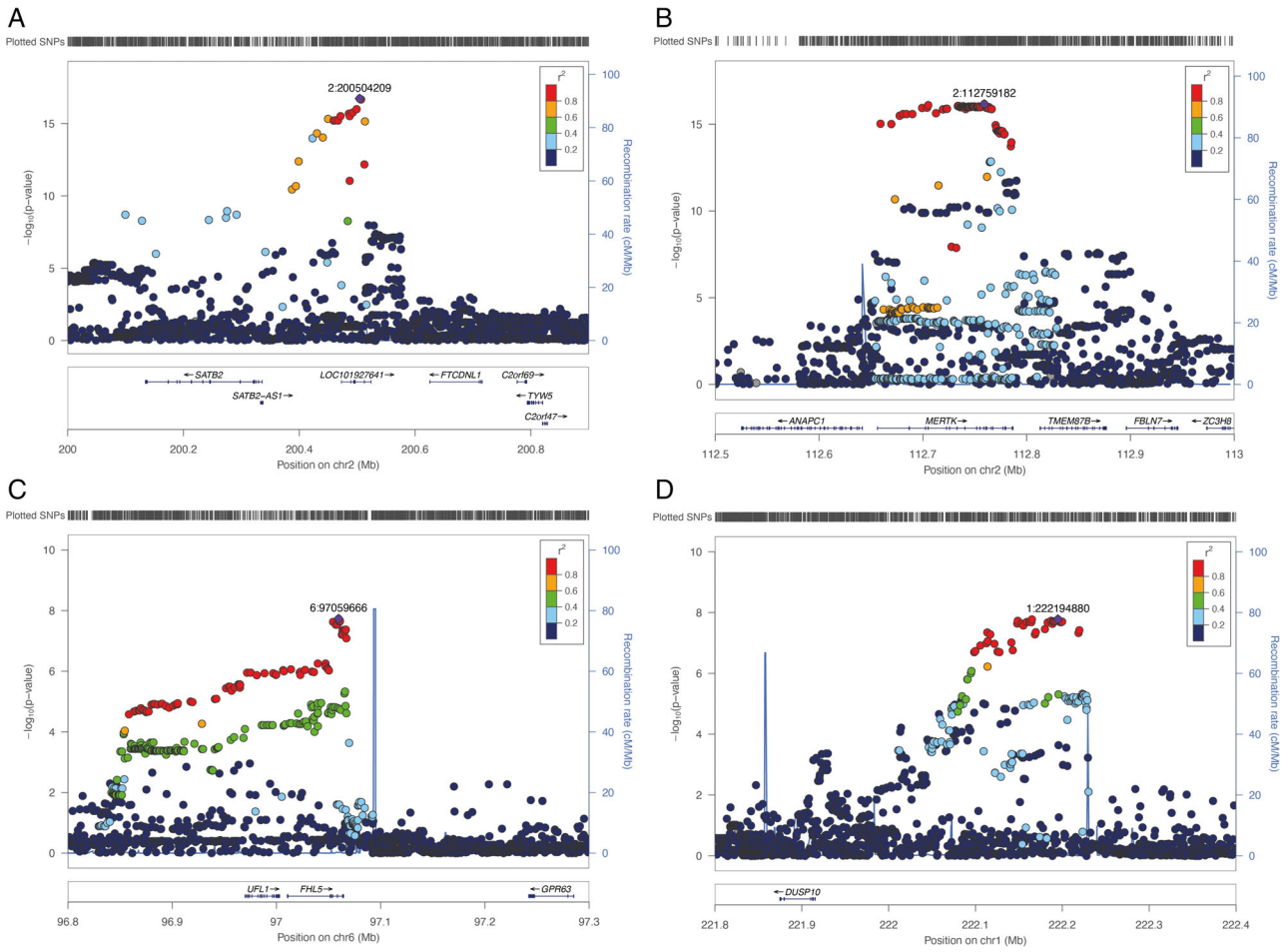

FIGURE 3: Regional association plots using imputed single nucleotide polymorphism (SNP) data. SNP positions, recombination rates, and gene boundaries are based on GRCh37/hg19. (A) The lead SNP rs113658130 ($p = 1.92 \times 10^{-17}$) at chromosome 2, which overlies the long intergenic noncoding RNA *LINC01877*. (B) The lead SNP rs4519530, an intronic variant in *MERTK* ($p = 6.98 \times 10^{-17}$). (C) The lead SNP rs11153082 ($p = 1.85 \times 10^{-8}$) at chromosome 6, which overlaps *FHL5*. (D) Chromosome 1q41 locus with the lead SNP rs12121134 ($p = 1.66 \times 10^{-8}$).

Functional Variant Annotation and Prediction

Closer investigation of our GWS loci revealed that the 2q13 locus in *MERTK* is represented by 2 independent lead SNPs: rs4519530 (reported in Table) and rs72825689 ($r^2 = 0.03$ with rs451930), which lies intergenic to *MERTK* ($p = 1.79 \times 10^{-12}$, OR 2.19, 95% CI = 1.76–2.73) in the merged analysis. All lead SNPs were located in noncoding regions of the genome. Four variants in high LD with the lead SNPs ($r^2 > 0.9$) were exonic variants with moderate impact. Two of these were missense variants in *MERTK* (rs7604639, rs2230515) and 2 in *FHL5* (rs2273621, rs9373985). Both *FHL5* variants had high CADD scores of 23.6 and 15.5, respectively, and all variants showed a high level of mammalian conservation, with GERP scores > 2.

Gene-Based Association Testing

Gene-based testing used the mean association signal from all SNPs within each gene, accounting for LD. A total of 18,559 genes were analyzed. A multiple testing-corrected p value threshold of 2.69×10^{-6} was applied to identify genes significantly associated with CH. Five candidate genes passed this threshold: *TMEM87B* (transmembrane protein 87B; $p = 1.06 \times 10^{-12}$), *MERTK* ($p = 5.55 \times 10^{-11}$), *ANAPC1* (anaphase promoting complex subunit 1; $p = 3.63 \times 10^{-10}$), *FBLN7* (fibulin 7; $p = 2.57 \times 10^{-7}$) on chr2q13, and *FHL5* ($p = 2.03 \times 10^{-6}$) on chr6q16.

Gene Expression and eQTL Analysis

All of the candidate genes (*MERTK*, *ANAPC1*, *TMEM87B*, *FHL5*, and *FBLN7*) are expressed in the human brain. Cell type analysis showed that *ANAPC1* and *FBLN7* have the highest RNA expression in neurons, whereas *MERTK* and *TMEM87B* have the highest RNA expression in brain support cells, namely microglia and astrocytes. *FHL5* had minimal expression in brain but was highly expressed in brain endothelial cells. Spatiotemporal expression pattern of the genes expressed in the human brain showed that *MERTK*, *TMEM87B*, and *ANAPC1*,

were expressed in all regions of the brain. *MERTK* and *FBLN7* were shown to be more highly expressed in the adult brain; the temporal expression in *FBLN7* particularly differed in the neocortex. Using lead SNPs (rs12121134, rs4519530, rs72825689, rs113658130, and rs11153082), eQTLs were identified in 11 genes, specifically looking at brain, spleen, vascular, and immunological tissues through FUMA. We observed overlapping eQTLs significant for all these tissues; eQTLs with the lowest p values, which were mapped by GWS SNPs, are represented for each tissue type in Table S3.

Pathway Analysis

Seventy-four pathways were significantly enriched for 46 genes in candidate regions ($p < 0.05$). The 5 most significant were cytokine–cytokine receptor interaction (KEGG 04060, $p = 3.97 \times 10^{-4}$), interleukin (IL)-1 family signaling (REAC R-HSA-446652, $p = 6.66 \times 10^{-5}$), growth factor receptor binding (GO 0070851 $p = 4.64 \times 10^{-5}$), IL-36 pathway (REAC R-HSA-9014826, $p = 4.21 \times 10^{-5}$), and IL-1 receptor binding (GO 0005149, $p = 1.30 \times 10^{-10}$). Of special interest for the brain were pathways related to neuroglial cells: positive regulation of glial cell proliferation (GO 0060252 $p = 0.0297$) and regulation of gliogenesis (GO 0014013, $p = 0.0407$). Moreover, there were a number of pathways relating to differentiation and activation of immune cells, as well as cell adhesion, many of which included the *MERTK* gene.

Migraine GWAS Overlap

Based on the significant lead SNPs of the combined analysis, 3 of the 4 loci showed no association with migraine. This was confirmed using colocalization analysis, which showed that for the loci on chromosomes 1 and 2, H1 (only CH had a genetic association within the tested region) was most likely. However, the lead SNP (rs11153082) at the chromosome 6 locus showed overlap with migraine. We confirmed this signal resulting from the same causal variant as in the migraine GWAS from the UK Biobank GWAS database (posterior probability for shared causal variant at chr6q16 = 97.4%). To exclude concurrent migraine driving this association, 655 UK patients with known migraine status (n = 195 had coexisting migraine) were reassessed. Fisher exact tests for the alternate allele frequency across the 2 groups, CH only compared to CH with concurrent migraine, in the lead SNP showed no significant differences (rs9386670: $p = 0.93$, OR = 1.02, 95% CI = 0.71–1.47).

Discussion

We conducted a GWAS in CH, identifying 4 susceptibility loci with large effect sizes, of which 1 has previously been associated with migraine. The strongest association was for a region on chromosome 2 containing a long intergenic noncoding RNA *LINC01877*. *LINC01877* is highly expressed in brain, most abundantly in the hippocampus and hypothalamus. The *SATB2* gene is 168kb proximal to the lead SNP. Mutations causing haploinsufficiency in this gene cause *SATB2*-associated syndrome, a disorder characterized by neurodevelopmental delay and craniofacial abnormalities.[33] In the developing brain, *SATB2* is required for cell-type specification of the upper layer pyramidal neurons in the neocortex and formation of the corpus callosum.[33] In adult mice, it is strongly expressed in hypothalamic regions and the A12 cell group of dopaminergic neurons.[34] Conditional knockout mice exhibit abnormalities in structures with a role in nociceptive processing, namely the somatosensory cortex and thalamocortical projection axons.[35] A second independently significant region on chromosome 2 overlies an intronic region of *MERTK* and is in LD with 2 missense mutations reaching GWS. *MERTK* encodes a TAM (TYRO33, AXL, and MERTK) receptor, regulators of microglial function, and the phagocytosis of apoptotic cells.[36] Homozygous mutations in *MERTK* cause retinitis pigmentosa, a condition sometimes associated with headache.[37,38] Cell-type expression analysis showed *MERTK* is highly expressed in microglia, and *Mer*-deficient mice exhibit an aggregation of apoptotic cells in neurogenic regions of the central nervous system.[36] MERTK has a role in neuroinflammation and has been associated with multiple sclerosis.[39] Furthermore, MERTK mediates astrocyte elimination of excess synapses, regulating synapse remodeling through neural circuit refinement.[40]

An association identified at rs11153082 on chromosome 6 correlates with a locus implicated in migraine and headache.[9] This association was stronger in the UK cohort, although the subsequent subgroup analysis found this was not dependent on the presence of coexistent migraine, indicating possible pleiotropy at this locus. The lead SNP driving this association is an intronic variant in the *FHL5* gene and overlaps the *UFL1* (UFM1-specific ligase 1) gene, which was identified as an eQTL in several tissues relevant for CH pathophysiology. In addition, *UFL1* was below GWS ($p = 8.46 \times 10^{-6}$) in the gene-based association testing. *UFL1* encodes a protein constituting part of the *UFM1* (ubiquitin-fold modifier 1) conjugation system involved in the apoptosis and trafficking of vesicles in the endoplasmic reticulum.[41]

Cell-type analysis showed enhanced microglial and neuronal expression. Pathways related to neuroglial cells such as positive regulation of glial cell proliferation and regulation of gliogenesis were also implicated in our results. Neuroinflammation is involved in several pain disorders, and microglia mediate the generation of neuropathic pain through amplification of excitatory neuronal

currents and attenuation of inhibitory currents.[42] Microglia influence central sensitization events in chronic pain and are responsible for synaptic pruning in brain development, modulating functional connectivity.[43,44] Connectivity defects have been demonstrated in CH.[45] Although the mechanism remains unclear, microglial dysfunction may dysregulate synaptic elimination and plasticity, causing connectivity impairment in CH. Microglia are also potential therapeutic targets for medications currently used in CH management. For example, verapamil, the prophylactic agent of choice, exhibits neuroprotective action through inhibition of microglial phagocyte oxidase (PHOX) activity, mediating generation of reactive oxygen species via binding to its catalytic subunit gp91.[46,47] Similarly, valproic acid, a histone deacetylase inhibitor, triggers microglial apoptosis to inhibit overactivation.[48]

Our results implicate immunological processes in the pathogenesis of CH. Immune eQTLs reaching significance include monocytes, T cells, and the spleen. Gene set enrichment analysis was significant for pathways involved in cytokine activity, especially the regulation of IL-1 and IL-36. The role of cytokines in the generation of headache has previously been suggested by enhanced nociceptive neuronal responses of the trigeminal nucleus caudalis, and subsequent hyperalgesia, following the injection of recombinant human IL-1β into the cerebrum of rats.[49] Several candidate genes from our analysis appear to influence the cAMP-responsive element binding protein (CREB) pathway, and *MERTK* activates CREB.[50] *FHL5* is also an activator of CREB, which subsequently is a transcription factor for *UFL1*.[51] The CREB pathway is critical for light entrainment of the circadian clock and also contributes to sensitization of nociceptive cells and meningeal pain hypersensitivity. It has a role in pain transmission evidenced by CREB activation in the trigeminal ganglion after stimulation of nociceptive neurons. Triptans, used to treat CH, reduce CREB activity within the trigeminal system.[52]

There are limitations to this study. This is a relatively small cohort due to the rarity of CH. A larger study or meta-analysis is required to derive additional associations with variants with a lower effect size. Genotyping of the UK controls on a different array platform introduces a potential confounder. The independent replication of loci in the Swedish cohort, which we performed on cases and controls genotyped on the same array (GSA), reduces the likelihood of spurious associations. Ensuring a phenotypically homogenous cohort is challenging, as CH is reliant upon clinical diagnosis. Our cohorts were carefully phenotyped to minimize possible confounding with migraine, which is phenotypically distinct.

We have identified replicable genome-wide significant associations that contribute to a genetic predisposition in CH. Microglial expression appears predominant among candidate genes, and pathway analysis implicated cytokine and immune activity as a pathogenic driver.

Future targeted sequencing of loci will fine-map these regions to help identify pathogenic variants and facilitate functional and mechanistic studies. In combination with deep phenotyping, this has potential in providing genotype–phenotype correlations such as genotype and treatment response, which can lead to future tailor-made treatment options respective of CH subtypes, such as episodic versus chronic CH, which can improve diagnosis. This could offer unprecedented insights into the pathophysiological pathways underlying CH and novel targets for therapeutic intervention.

Postscript

Two parallel manuscripts (Harder et al and O'Connor et al), submitted to this journal, report the first replicated genomic loci associated with CH. Whereas Harder et al investigated Dutch CH cases (n = 840) and controls (n = 1,457) and Norwegian CH cases (n = 144) and controls (n = 1,800), O'Connor et al investigated UK cases (n = 852) and controls (n = 5,614) as well as Swedish cases (n = 591) and controls (n = 1,134). The 4 loci reported by Harder et al correspond to 4 loci reported by O'Connor et al, with the index variants reported in the 2 studies being in LD with each other (D' = 0.86 and r^2 = 0.36 for rs12121134 and rs11579212; D' = 0.98 and r^2 = 0.95 for rs4519530 and rs6541998; D' = 0.95 and r^2 = 0.34 for rs113658130 and rs10184573; D' = 0.93 and r^2 = 0.38 for rs11153082 and rs2499799, in the 1000 Genomes data for European populations). The independent discovery of the 4 loci in the 2 studies provides additional support that they represent genuine risk loci for CH.

Next, we combined the summary statistics from the 4 studies (Dutch, Norwegian, UK, Swedish) using inverse-variance–weighted meta-analysis as implemented in METAL (with the "STDERR" option), after harmonizing the datasets using EasyQC.[53,54] In total, 8,039,373 variants were analyzed. The association to CH remained significant for all the 8 index variants (in the 4 loci) reported in the 2 papers: rs11579212 (effect allele [EA]: C), OR = 1.31, 95% CI = 1.21–1.41, $p = 8.98 \times 10^{-13}$; rs12121134 (EA: T), OR = 1.40, 95% CI = 1.29–1.53, $p = 9.18 \times 10^{-15}$; rs6541998 (EA: C), OR = 1.40, 95% CI = 1.30–1.51, $p = 2.37 \times 10^{-19}$; rs4519530 (EA: C), OR = 1.41, 95% CI = 1.31–1.52, $p = 4.18 \times 10^{-29}$; rs10184573 (EA: T), OR = 1.38, 95% CI = 1.28–1.50, $p = 3.35 \times 10^{-16}$; rs113658130 (EA: C), OR = 1.54, 95% CI = 1.41–1.69, $p = 1.28 \times 10^{-21}$; rs2499799 (EA: C), OR = 0.77, 95%

CI = 0.70–0.84, $p = 2.73 \times 10^{-8}$; rs11153082 (EA: G), OR = 1.33, 95% CI = 1.23–1.43, $p = 2.98 \times 10^{-14}$. The 8 index variants in the overlapping loci showed a consistent effect direction across the 2 studies. Colocalization analysis identified a high posterior probability for 3 loci (those on chromosomes 1 and 2) to represent the same causal variant.[32] rs12121134 and rs11579212 have a posterior probability that the causal variants are the same (H4) of 80.4%; for rs4519530 and rs6541998, H4 is 87.4%, and for rs113658130 and rs10184573, H4 is 96.9%. For the locus on chromosome 6, the colocalization analysis shows a higher probability that the loci in the 2 studies represent distinct causal variants (H3, 78.7%) rather than the same causal variant (H4, 21.2%).

Finally, the meta-analysis resulted in 3 additional loci becoming genome-wide significant: (1) a locus on chromosome 7 with 31 significant ($p < 5 \times 10^{-8}$) variants with index variant rs6966836 (chromosome 7: 117002998, EA: C), OR = 1.25, 95% CI = 1.16–1.35, $p = 2.46 \times 10^{-9}$; (2) a locus on chromosome 10 with 2 significant variants with index variant rs10786156 (chromosome 10: 96014622, EA: C), OR = 1.24, 95% CI = 1.15–1.33, $p = 7.61 \times 10^{-9}$; and (3) a locus on chromosome 19 with 2 significant variants with index variant rs60690598 (chromosome 19: 55052198, EA: T), OR = 1.87, 95% CI = 1.51–2.33, $p = 1.70 \times 10^{-8}$.

Acknowledgments

Sweden: This study was supported by the Swedish Research Council (2017-01096), the Swedish Brain Foundation, and the Mellby Gård Foundation (FO2018-0008), Karolinska Institutet Research Funds (2018-01738). We thank F. Xiang for excellent technical assistance and A.-C. Karlsson for help with patient recruitment. UK: E.O. is supported by Brain Research UK. Patients were collected as part of the SYNaPS Study Group collaboration funded by the Wellcome Trust and strategic award (Synaptopathies) funding (WT093205 MA and WT104033AIA). J.V. is supported by Health Education England and the Medical Research Council. L.S. is supported by the Wellcome Trust Institutional Strategic Support Fund (204809/Z/16/Z) awarded to St George's, University of London. We thank Dr. J. E. Burns for his assistance in preparing the manuscript.

Author Contributions

E.O., C.F., C.R., L.S., G.L.C., M.V., K.P., R.T., N.W.W., J.V., H.H., M.M., and A.C.B. contributed to the conception and design of the study. E.O., P.S., C.F., F.L., L.S., J. Y., N.G., N.S., F.A., I.C.H., B.D., M.Z.C., B.S.S., R.S., S. E., J.A., O.Q., C.C., G.L.C., T.K., E.K., L.G., A.V.E.H., B.S.W., L.S.V., S.L., D.D., R.T., N.W.W., I.K., A.S., C.S., E.W., C.R., J.V., H.H., M.M., and A.C.B. contributed to the acquisition and analysis of data. E.O., P.S., C.F., J.V., C.R., and A.C.B. contributed to drafting a significant portion of the manuscript or figures.

Potential Conflicts of Interest

Nothing to report.

References

1. Fischera M, Marziniak M, Gralow I, Evers S. The incidence and prevalence of cluster headache: a meta-analysis of population-based studies. Cephalalgia 2008;28:614–618.
2. Headache Classification Committee of the International Headache Society (IHS). The international classification of headache disorders, 3rd edition. Cephalalgia 2018;38:629–808.
3. Vollesen AL, Benemei S, Cortese F, et al. Migraine and cluster headache—the common link. J Headache Pain 2018;19:89.
4. Steinberg A, Fourier C, Ran C, et al. Cluster headache—clinical pattern and a new severity scale in a Swedish cohort. Cephalalgia 2018; 38:1286–1295.
5. Bahra A, May A, Goadsby PJ. Cluster headache: a prospective clinical study with diagnostic implications. Neurology 2002;58:354–361.
6. Gibson KF, Dos Santos A, Lund N, et al. Genetics of cluster headache. Cephalalgia 2019;39:1298–1312.
7. Goadsby PJ. Pathophysiology of cluster headache: a trigeminal autonomic cephalgia. Lancet Neurol 2002;1:251–257.
8. May A, Ashburner J, Buchel C, et al. Correlation between structural and functional changes in brain in an idiopathic headache syndrome. Nat Med 1999;5:836–838.
9. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 2016;48:856–866.
10. Bacchelli E, Cainazzo MM, Cameli C, et al. A genome-wide analysis in cluster headache points to neprilysin and PACAP receptor gene variants. J Headache Pain 2016;17:114.
11. Ran C, Fourier C, Michalska JM, et al. Screening of genetic variants in ADCYAP1R1, MME and 14q21 in a Swedish cluster headache cohort. J Headache Pain 2017;18:88.
12. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
13. Pedersen BS, Quinlan AR. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. Am J Hum Genet 2017;100:406–413.
14. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48: 1279–1283.
15. Loh P-R, Palamara PF, Price AL. Fast and accurate long-range phasing in a UKbiobank cohort. Nat Genet 2016;48:811–816.
16. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–1287.
17. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335–1341.
18. RStudio Team. *RStudio: integrated development environment for R.* Boston, MA: RStudio, 2015.

19. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26:2336–2337.

20. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.

21. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006.

22. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. Genome Biol 2016;17:122.

23. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 2005; 15:901–913.

24. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11: e1004219.

25. Consortium GTe. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580–585.

26. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. Nature 2011;478:483–489.

27. Zhang Y, Sloan SA, Clarke LE, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron 2016;89:37–53.

28. Raudvere U, Kolberg L, Kuzmin I, et al. G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47:W191–W198.

29. Consortium GO. The gene ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258–D261.

30. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2015; 44:D457–D462.

31. Croft D, O'Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 2011;39: D691–D697.

32. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.

33. Döcker D, Schubach M, Menzel M, et al. Further delineation of the SATB2 phenotype. Eur J Hum Genet 2014;22:1034–1039.

34. Huang Y, Song N-N, Lan W, et al. Expression of transcription factor Satb2 in adult mouse brain. Anat Rec (Hoboken) 2013;296:452–461.

35. Zhang Q, Huang Y, Zhang L, et al. Loss of Satb2 in the cortex and hippocampus leads to abnormal behaviors in mice. Front Mol Neurosci 2019;12:33.

36. Fourgeaud L, Través PG, Tufail Y, et al. TAM receptors regulate multiple features of microglial physiology. Nature 2016;532:240–244.

37. Walsh AD, Johnson LJ, Harvey AJ, et al. Identification and characterisation of cis-regulatory elements upstream of the human receptor tyrosine kinase gene MERTK. Brain Plast 2020;Preprint:1–14.;

38. Tongue AC. Clinical findings and common symptoms in retinitis pigmentosa. Am J Ophthalmol 1988;106:507–508.

39. Ma GZM, Stankovich J, Australia, et al. Polymorphisms in the receptor tyrosine kinase MERTK gene are associated with multiple sclerosis susceptibility. PLoS One 2011;6:e16964.

40. Chung W-S, Clarke LE, Wang GX, et al. Astrocytes mediate synapse elimination through MEGF10 and MERTK pathways. Nature 2013; 504:394–400.

41. Zhang Y, Zhang M, Wu J, et al. Transcriptional regulation of the Ufm1 conjugation system in response to disturbance of the endoplasmic reticulum homeostasis and inhibition of vesicle trafficking. PLoS One 2012;7:e48587.

42. Zhao H, Alam A, Chen Q, et al. The role of microglia in the pathobiology of neuropathic pain development: what do we know? Br J Anaesth 2017;118:504–516.

43. Kobayashi K, Yamanaka H, Fukuoka T, et al. P2Y12 receptor upregulation in activated microglia is a gateway of p38 signaling and neuropathic pain. J Neurosci 2008;28:2892–2902.

44. Schafer DP, Lehrman EK, Kautzman AG, et al. Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. Neuron 2012;74:691–705.

45. Chou K-H, Yang F-C, Fuh J-L, et al. Altered white matter microstructural connectivity in cluster headaches: a longitudinal diffusion tensor imaging study. Cephalalgia 2014;34:1040–1052.

46. Leone M, D'Amico D, Frediani F, et al. Verapamil in the prophylaxis of episodic cluster headache: a double- blind study versus placebo. Neurology 2000;54:1382–1385.

47. Liu Y, Lo Y-C, Qian L, et al. Verapamil protects dopaminergic neuron damage through a novel anti-inflammatory mechanism by inhibition of microglial activation. Neuropharmacology 2011;60:373–380.

48. Chen PS, Wang CC, Bortner CD, et al. Valproic acid and other histone deacetylase inhibitors induce microglial apoptosis and attenuate lipopolysaccharide-induced dopaminergic neurotoxicity. Neuroscience 2007;149:203–212.

49. Oka T, Aou S, Hori T. Intracerebroventricular injection of interleukin-1 beta enhances nociceptive neuronal responses of the trigeminal nucleus caudalis in rats. Brain Res 1994;656:236–244.

50. Cummings CT, Deryckere D, Earp HS, Graham DK. Molecular pathways: MERTK signaling in cancer. Clin Cancer Res 2013;19:5275–5280.

51. The ENCODE (ENCyclopedia of DNA elements) project. Science 2004;306:636–640.

52. Mitsikostas DD, Knight YE, Lasalandra M, et al. Triptans attenuate capsaicin-induced CREB phosphorylation within the trigeminal nucleus caudalis: a mechanism to prevent central sensitization? J Headache Pain 2011;12:411–417.

53. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191.

54. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 2014;9:1192–1212.