

1 **Title: Rare risk variants associate with epigenetic dysregulation in migraine**

2

3 Tanya Ramdal Techlo,¹ Mona Ameri Chalmer,¹ Peter L. Møller,² Lisette J. A. Kogelman,¹ Isa A.
4 Olofsson,¹ DBDS Genomic Consortium, Karina Banasik,³ Mette Nyegaard,⁴ Jes Olesen,¹ Thomas
5 Folkmann Hansen*^{1,3}

6

7 ¹Danish Headache Center, Department of Neurology, Copenhagen University Hospital,
8 Rigshospitalet-Glostrup, DK-2600 Glostrup, Denmark

9 ²Department of Biomedicine, University of Aarhus, DK-8000 Aarhus C, Denmark

10 ³Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, DK-2200
11 Copenhagen N, Denmark

12 ⁴Department of Health Science and Technology, University of Aalborg, DK-9220 Aalborg E,
13 Denmark

14

15 ***Corresponding author:**

16 Thomas Folkmann Hansen, Danish Headache Center, Department of Neurology, Copenhagen
17 University Hospital, Rigshospitalet-Glostrup, Valdemar Hansens Vej 5, DK-2600 Glostrup,
18 Denmark.

19 Phone: +45 38633051, Email: thomas.folkmann.hansen@regionh.dk

20

21 **ABSTRACT**

22

23 Migraine has a heritability of up to 65%. Genome-wide association studies (GWAS) on migraine
24 have identified 123 risk loci, explaining only 10.6% of migraine heritability. Thus, there is a

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

25 considerable genetic component not identified with GWAS. Further, the causality of the identified
26 risk loci remains inconclusive. Rare variants contribute to the risk of migraine but GWAS are often
27 underpowered to detect these. Whole genome sequencing is reliable for analyzing rare variants but
28 is not frequently used in large-scale. We assessed if rare variants in the migraine risk loci associated
29 with migraine. We used a large cohort of whole genome sequenced migraine patients (1,040
30 individuals from 155 families). The findings were replicated in an independent case-control cohort
31 (2,027 migraine patients, 1,650 controls). We found rare variants (minor allele frequency<0.1%)
32 associated with migraine in a Polycomb Response Element in the *ASTN2* locus. The association was
33 independent of the GWAS lead risk variant in the locus. The findings place rare variants as risk
34 factors for migraine. We propose a biological mechanism by which epigenetic regulation by
35 Polycomb Response Elements plays a crucial role in migraine etiology.

36

37 INTRODUCTION

38

39 Migraine is a complex neurovascular and genetic disorder which affects 15-20% of the population.¹
40 It is characterized by episodes of severe headache, which is typically unilateral and pulsating,
41 accompanied by nausea, vomiting, photophobia and/or phonophobia.² Identifying factors that
42 predispose to migraine is highly relevant. The Global Burden of Disease Study (2019) ranks
43 migraine the highest cause of disability worldwide in young women (aged 15-49) and the second
44 highest cause of disability worldwide among both sexes.³ Additionally, migraine has a large
45 socioeconomic impact through treatment costs and loss of productivity estimated at €27 billion in
46 Europe alone.⁴

47

48 Migraine has a significant genetic component and is polygenic. There is an increased familial risk
49 and twin studies have shown a heritability of 34%–65%.^{5,6} Multiple, rare variants, each with large
50 effect sizes, may contribute to disease risk together with common variants.^{7,8} The most recent meta-
51 analysis of migraine genome-wide association studies (GWAS) identified 123 migraine risk loci, of
52 which 121 were autosomal. Still, only 10.6% of the total heritability was explained by the GWAS.⁹
53 GWAS are often underpowered to detect rare variants, as these are imputed and not measured.¹⁰
54 Whole genome sequencing (WGS) provides a direct assessment of rare variants. However, it is a
55 technology that is still not frequently used in large-scale and is costly compared to genotyping using
56 arrays.

57 How the migraine risk loci affect migraine pathogenesis is unknown. Many reside in non-protein
58 coding regions of the DNA and may alter the regulation of gene expression.¹¹ Increased risk of
59 migraine has been linked to disruption in gene regulation¹² and epigenetic regulation.^{12–14} While
60 candidate genes have been identified for increased genetic risk in migraine, the findings have been
61 reported with a high risk of false positive results.^{15–18} This suggests that migraine pathogenesis is
62 affected by mutations outside of protein-coding regions of the DNA, such as in regulatory regions.

63

64 We hypothesized that migraine is influenced by rare variants in regulatory regions. We tested if rare
65 variants in the migraine risk loci increase the risk of migraine. We defined blocks of linkage
66 disequilibrium (LD) around the 121 autosomal GWAS lead risk variants using a Danish cohort of
67 90,312 individuals.^{9,19} Within the LD-blocks, we analyzed rare variants (minor allele frequency
68 (MAF)<0.1%, MAF<1%, and MAF<5%) in genes and regulatory regions using a Sequence Kernel
69 Association Test. For the analysis, we used WGS data from a large cohort of 155 families
70 ($n_{individuals}=1,040$) with clustering of migraine. The findings were replicated in an independent case-
71 control cohort of migraine patients with no familial history of migraine and controls.

72

73 **MATERIALS AND METHODS**

74

75 **Migraine family cohort**

76

77 The migraine family cohort consisted of 155 families with a Mendelian-like inheritance pattern of
78 migraine. The participants were recruited from the Danish Headache Center, Rigshospitalet-
79 Glostrup, Denmark, as described elsewhere.^{20,21} All participants were examined by a neurology
80 resident or a senior medical student specifically trained in headache diagnostics. The participants
81 were examined using a semi-structured interview^{22,23} based on the International Classification of
82 Headache Disorders.² The families consisted of 1,040 participants, including 746 individuals
83 diagnosed with migraine and 294 without migraine. The mean number of participants per family
84 was 6.7 and the mean number of migraine patients per family was 4.8. Most of the families were
85 extended beyond the nuclear family. The smallest families consisted of at least two individuals with
86 at least one migraine patient.

87

88 **Replication cohort**

89

90 The replication cohort consisted of 2,027 migraine patients with no familial history of migraine, i.e.
91 unrelated migraine patients, and 1,650 controls. The unrelated migraine patients were recruited and
92 assessed at the Danish Headache Center using the same procedures as described for the family
93 cohort. The controls were recruited for the Danish study of Non-Invasive testing in Coronary Artery
94 Disease (Dan-NICAD) according to procedures described by Nissen *et al.*²⁴

95

96 **Whole genome sequencing**

97

98 Blood samples were taken from all participants and genomic DNA was extracted from whole blood.

99 The sequencing was performed on an Illumina NovaSeq 6000 sequencing platform with S4 flow
100 cells. The WGS data was subjected to quality control by deCODE genetics, as described
101 elsewhere.²⁵ All samples from both the migraine family cohort and the replication cohort were
102 subjected to the same sequencing and quality control procedures.

103

104 **Analytical approach**

105

106 **Selecting genomic regions for analysis**

107

108 We calculated blocks of linkage disequilibrium (LD) around the 121 autosomal migraine risk loci,⁹
109 using the genomic positions of the GWAS lead risk variants, i.e. the variants with the smallest
110 GWAS *p*-value. For this we used 90,312 participants from the Danish Blood Donor Study
111 (DBDS).¹⁹ The genomic positions of the GWAS lead risk variants were based on SNP positions
112 from the Database of Single Nucleotide Polymorphisms (dbSNP) build 153 in the GRCh38.p12
113 (hg38) assembly. The LD-blocks were calculated using a LD-block recognition algorithm proposed
114 by Gabriel *et al.*²⁶ with PLINK2.²⁷ Here, we assessed variant-pairs within 1000 kilobases of each
115 other with a MAF>5% using a 90% D' confidence interval threshold >0.70. Subsequently, we
116 annotated genes and regulatory regions within the LD-blocks and analyzed these. If a LD-block
117 partially spanned a gene or regulatory region, we included the entire gene or regulatory region in
118 the analysis. If a LD-block was located in an intergenic region, we extended the genomic region for
119 analysis to include the nearest protein- or RNA-coding gene. No LD-block was generated for 22

120 loci. In such case, the nearest protein- or RNA-coding gene was included in the analysis. The
121 genomic positions of the regions we analyzed are found in Table S1.

122

123 **Annotating genes and regulatory regions for analysis**

124

125 The gene transcripts were derived from GENCODE release 36 (basic gene-set only). They included
126 protein-coding genes, non-coding RNA-genes, and pseudogenes. We analyzed the genes in their
127 full length. In addition, we analyzed the coding exons, introns, 5' untranslated regions (UTR)
128 exons, and 3' UTR exons by themselves. For regulatory regions, we analyzed promoters, enhancers,
129 CpG islands, insulators, Polycomb Response Elements (PREs), and Transcription Factor Binding
130 Sites (TFBSs). We included promoters from The Eukaryotic Promoter Database (EPD) version 6
131 and EPD non-coding promoters version 1.²⁸ The enhancers were included from the GeneHancer
132 project,²⁹ using only enhancers that are supported by multiple studies. The genomic positions of the
133 CpG islands were based on predictions by Gardiner-Garden *et al.*³⁰ The insulators and PREs were
134 included from the Broad Institute Chromatin State Segmentation by HMM³¹ from The
135 Encyclopedia of DNA Elements (ENCODE) Consortium³² in H1-hESC cells in the GRCh37 (hg19)
136 assembly. The TFBS were included from ENCODE³² in the GRCh37 (hg19) assembly. All genomic
137 positions in hg19 were converted to the CRCh38 (hg38) assembly. Table S2 displays the number of
138 genes and regulatory regions that were analyzed in this study.

139

140 **Preparing VCF files for analysis**

141

142 The VCF files of each participant in the study were merged into one file. Multiallelic sites were
143 converted to a biallelic representation. Subsequently, we annotated genes and regulatory regions to

144 the merged VCF file. We isolated variants, including SNVs, insertions, and deletions, with a
145 MAF<0.1%, MAF<1%, or MAF<5%, using BCFtools.³³

146

147 **Rare variant association analysis**

148

149 To test for an association between the rare variants in genes or regulatory regions and migraine, we
150 used the software of Family Sequence Kernel Association Test (F-SKAT).³⁴ Age and gender were
151 used as covariates. The age was defined as the age at the time of the interview (time of diagnosis).
152 To avoid association merely given that longer genes and regulatory regions can contain more rare
153 variants, we used modified beta-values.³⁵ The beta-values were modified by dividing them with the
154 length of the genes or regulatory regions. The *p*-values were controlled for multiple testing using
155 the Bonferroni method based on the number of genes and regulatory regions tested. A Bonferroni-
156 corrected *p*-value<0.05 was considered significant.

157 To assess if the results were dependent of the GWAS lead risk variants, the number of risk alleles of
158 the variants (0, 1, or 2) were included as covariates in a separate rare variant association analysis
159 with F-SKAT.

160

161 **Replication**

162

163 The findings were replicated using an independent cohort of unrelated migraine patients and
164 controls. Aggregated data were available on rare variants (MAF<0.1%, MAF<1%, MAF<5%) in
165 the genes and regulatory regions for the controls. We tested if the frequency of the rare variants was
166 increased in the migraine patients compared with the controls, using a Fisher's Exact Test.³⁶ We
167 used the null hypothesis of no association with migraine under a dominant model of penetrance.

168 The resulting p -values were controlled for multiple testing using the Bonferroni method. A
169 Bonferroni-corrected p -value <0.05 was considered significant.

170

171 **Characterization of findings**

172

173 To explore the effect of the rare variants of the replicated results, we used Ensembl Variant Effect
174 Predictor.³⁷ To discover the target genes of the replicated results, we used *cis*-eQTLs data from the
175 Genotype-Tissue Expression (GTEx) version 8.0 dataset (dbGaP Accession phs000424.v8.p2
176 release).

177

178 **Ethics**

179

180 Oral and written consent was obtained from all participants. All procedures performed in studies
181 involving human participants were in accordance with the ethical standards of the National
182 Committee on Health Research Ethics (file no. H-2-2010-122) and with the 1964 Helsinki
183 declaration and its later amendments. The study was approved by the data-protection agency
184 (01080/GLO-2010-10).

185

186 **RESULTS**

187

188 **Rare variants in regulatory regions associate with migraine**

189

190 We found an association between migraine and rare variants with a $MAF<0.1\%$ in 7 regulatory
191 regions and an intron of the *MACF1* gene (see Table 1 for genomic regions and p -values). No

192 association was found for rare variants with a MAF<1% or a MAF<5% nor in coding regions (full
193 gene transcripts, coding exons, 5' UTR exons, 3' UTR exons). In line with what is expected when
194 analyzing rare variants for polygenic traits,^{9,38,39} we found that the distribution of the test statistics
195 deviated from the null with inflated genomic inflation factors ($\lambda=1.89$, $\lambda=1.63$, and $\lambda=1.68$) (Figure
196 S1).

197

198 Subsequently, we replicated the Polycomb Response Element (PRE) located at chr9:117,411,901-
199 117,412,101 in the *ASTN2* locus ($p=1.1 \cdot 10^{-15}$) using an independent case-control cohort, henceforth
200 referred to as the replicated PRE.

201

202 We found that the findings were independent of the GWAS lead risk variants, with unaffected
203 results ($p=1.3 \cdot 10^{-4}$ in the original analysis (see Table 1) and $p=1.2 \cdot 10^{-4}$ with the allele count of the
204 risk variants as covariates).

205

206 **Table 1. Summary of the results obtained from the rare variant association analysis.** The
207 results from the rare variant association analysis together with the locus ID, as given in the meta-
208 analysis of migraine GWAS by Hautakangas *et al.*⁹, the genomic positions in GRCh38.p12 (hg38),
209 and type of genomic feature.

210

Locus ID	Genomic positions	Type	p -value	Corrected p -value
<i>MACFI</i>	chr1:39,303,078-39,309,569	Intron	$1.8 \cdot 10^{-5}$	0.019
<i>THADA</i>	chr2:43,400,466-43,400,944	Transcription Factor Binding Site	$5.3 \cdot 10^{-6}$	0.013

<i>PHACTR1</i>	chr6:12,748,985-12,750,439	Enhancer	$1.0 \cdot 10^{-4}$	0.021
<i>ASTN2</i>	chr9:117,411,901-117,412,101	Polycomb Response Element	$1.3 \cdot 10^{-4}$	0.012
near <i>SPATA19</i>	chr11:133,939,661-133,940,767	CpG island	$1.9 \cdot 10^{-4}$	0.027
	chr14:26,917,154-26,917,554	Insulator	$1.2 \cdot 10^{-4}$	0.014
near <i>LRFN5</i>	chr14:41,610,647-41,610,847	Polycomb Response Element	$6.7 \cdot 10^{-5}$	$6.2 \cdot 10^{-3}$
	chr14:41,610,658-41,610,787	Enhancer	$6.7 \cdot 10^{-5}$	0.028

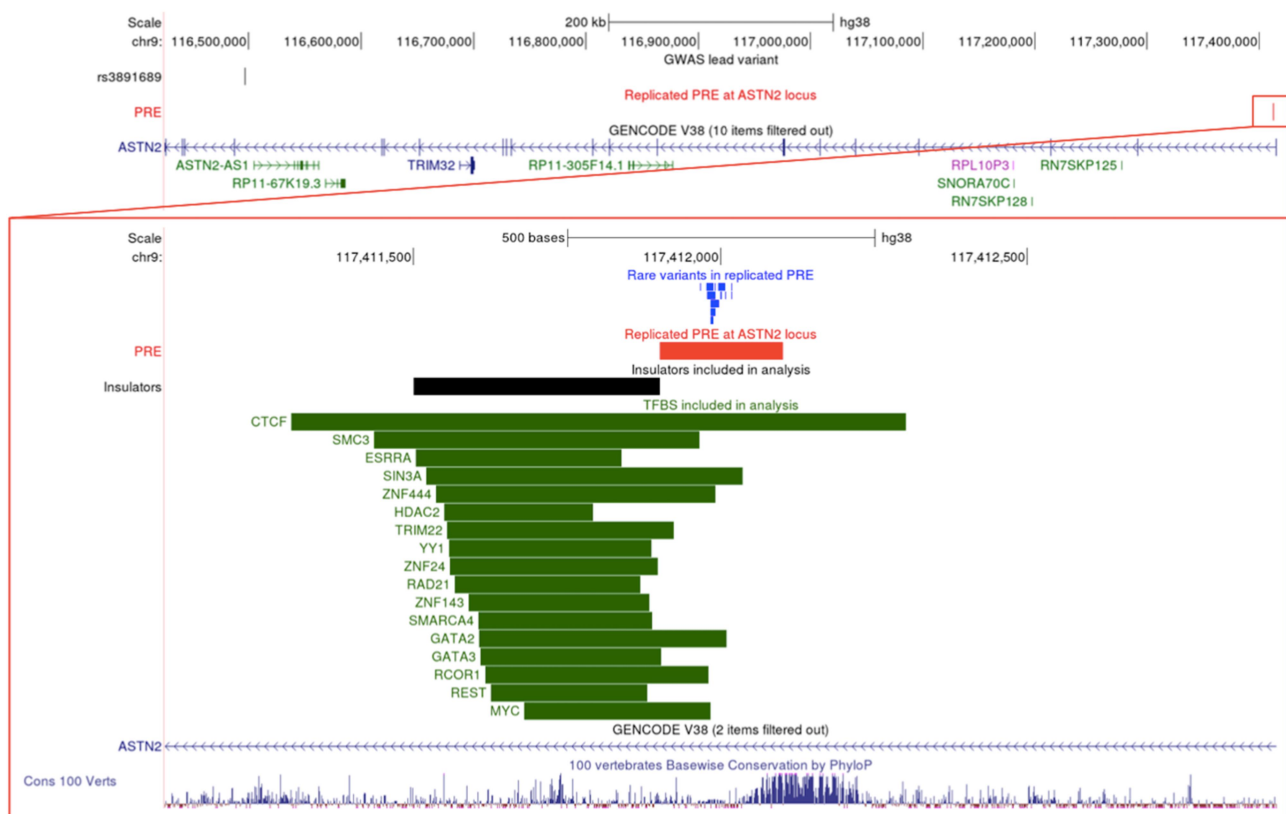
211

212 **Rare variants in PRE map to CTCF binding sites**

213

214 The replicated PRE was located in the first intron of the *ASTN2* gene and overlapped the TFBS of
 215 CTCF, SIN3A, ZNF444, GATA2, RCOR1, and MYC (Figure 1). There were 12 rare variants with
 216 a MAF<0.1% located in the replicated PRE. These rare variants consisted of 1 SNV, 9 deletions,
 217 and 2 insertions (Table 2). We found that all rare variants mapped to CTCF binding sites (Table 2)
 218 using the Ensembl Variant Effect Predictor (VEP).

219



220

221

222 **Figure 1. Overview of the genomic region around the *ASTN2* migraine risk locus. Top:** The
 223 genomic region that was included for analysis (see Table S1 for genomic positions). Displayed is
 224 the GWAS lead risk variant (rs3891689), the replicated Polycomb Response Element, and the genes
 225 included for analysis. **Bottom:** A close-up of the replicated Polycomb Response Element and its
 226 surrounding genomic region. Displayed are the rare variants with a MAF<0.1%, that associated
 227 with migraine, the replicated Polycomb Response Element, near-by regulatory regions (an insulator
 228 and 17 transcription factor binding sites), the first intron of the *ASTN2* gene, and the evolutionary
 229 conservation across 100 vertebrates.

230

231 **Table 2. Information on the rare variants in the replicated Polycomb Response Element.** The
 232 table presents the genomic positions in GRCh38.p12 (hg38), the reference allele, the alternative

233 allele, the variant type, the rs number, and the Ensembl regulatory feature of the rare variants
234 (MAF<0.1%) in the replicated Polycomb Response Element.

Genomic position	Reference allele	Alternative allele	Variant type	rs number	Ensembl regulatory feature
chr9:117411967- 117411967	A	C	SNV	rs866789329	CTCF binding site
chr9:117411977- 117411988	CCACCCCTACCA	C	Deletion		CTCF binding site
chr9:117411978- 117411991	CACCCCTACCACCA	C	Deletion	rs158802387 2	CTCF binding site
chr9:117411983- 117411997	CTACCA	C	Deletion	rs144647713 7	CTCF binding site
chr9:117411983- 117411991	CTACCACCA	C	Deletion		CTCF binding site
chr9:117411983- 117411988	CTACCACCACCCCA A	C	Deletion		CTCF binding site
chr9:117411990- 117411991	CA	C	Deletion		CTCF binding site
chr9:117411996- 117412007	CACCTCACCCCT	C	Deletion		CTCF binding site
chr9:117411999- 117412002	CTCA	C	Deletion	rs138331219 5	CTCF binding site
chr9:117412007- 117412008	TC	T	Deletion	rs145255224	CTCF binding site
chr9:117412018- 117412018	A	ACCCCCCCCCC	Insertion		CTCF binding site
chr9:117412018- 117412018	A	ACCCCCCCCCC C	Insertion		CTCF binding site

235

236 **Possible regulatory targets unidentified**

237

238 Finally, we assessed the possible regulatory targets of the replicated PRE. We found no significant
239 *cis*-eQTLs in the PRE using the GTEx database for the identification of regulatory targets.

240

241 **DISCUSSION**

242

243 Our study is, to our knowledge, the first to assess if rare variants in 121 autosomal migraine risk
244 loci increase the risk of migraine. We assessed if rare variants in genes and regulatory regions
245 associated with migraine, using WGS data from families with clustering of migraine. We replicated
246 a Polycomb Response Element (PRE) in the *ASTN2* locus using an independent case-control cohort.

247

248 **Regulatory regions harbor rare variants associated with migraine**

249

250 We found an association between migraine and rare variants with a MAF<0.1% in 7 regulatory
251 regions and an intron of the *MACF1* gene. Our results indicate that migraine can be associated with
252 rare variants in regulatory regions, since introns are likely to harbor elements with a regulatory
253 function.⁴⁰ We replicated a PRE in the *ASTN2* locus using an independent case-control cohort.
254 Interestingly, we have previously reported a rare variant association with migraine in a PRE in the
255 same locus, using a different study design.¹² Importantly, we show that the effect of the replicated
256 PRE was independent of the GWAS lead risk variant. Thus, our results show that migraine GWAS
257 loci also contain independent rare risk variants.

258

259 **Disruption in epigenetic regulation as a possible biological mechanism for migraine**

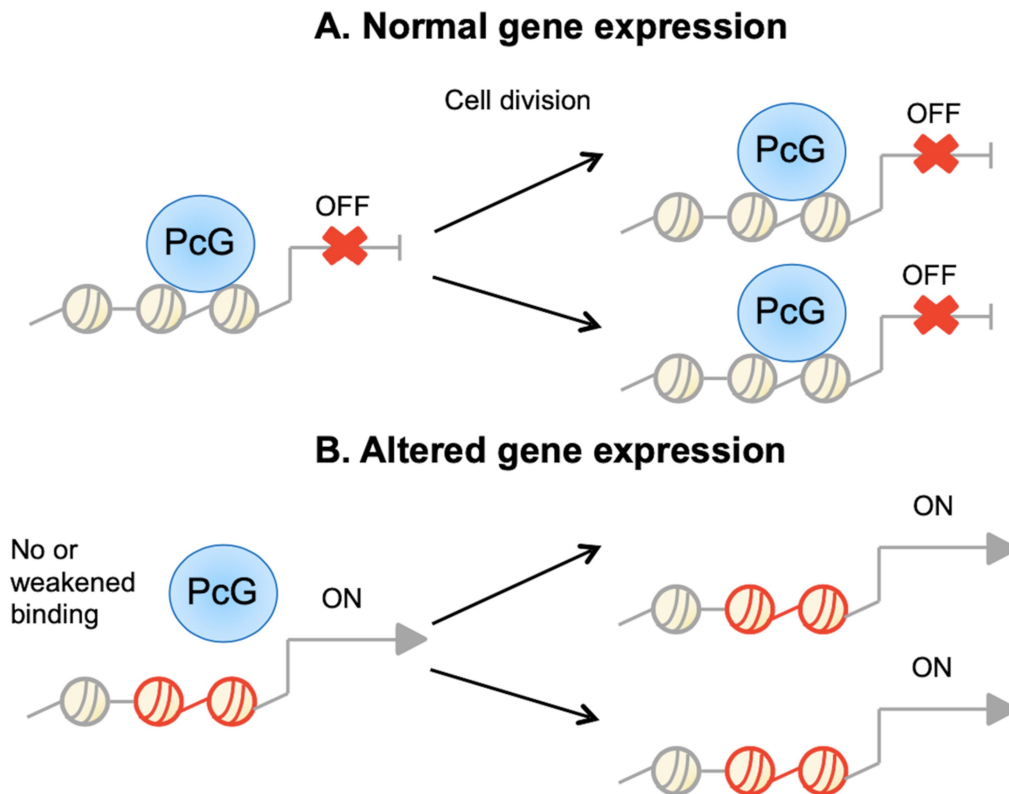
260

261 The function of mammalian PREs is still largely unknown.⁴¹ Mammalian PREs are likely capable
262 of binding Polycomb Group (PcG) proteins as a part of transcriptional silencing during cell
263 division, where the silencing is maintained in subsequent daughter cells.^{41–43} Associated with the
264 silencing is remodeling of the chromatin.^{44,45} Possibly, the rare variants we find associated with
265 migraine in the replicated PRE can cause a disruption in the binding of PcG proteins and chromatin
266 remodeling. The disruption may lead to transcriptional silencing not being maintained after cell
267 division. This ultimately leads to an altered gene expression (Figure 2). Epigenetic mechanisms
268 have been linked to brain plasticity in the transition from episodic to chronic migraine.^{13,14}
269 However, we are, to our knowledge, the first to report an indication of an actual biological
270 mechanism for migraine, where chromatin remodeling is disrupted by rare variants in PREs.

271

272 We found that the rare mutations in the replicated PRE were situated in binding sites of the
273 transcription factor CTCF. CTCF is a transcriptional repressor, with many functions including
274 regulation of gene expression through chromatin modifications.⁴⁶ Again, our findings indicate a link
275 to a disruption in chromatin remodeling as a biological mechanism for migraine.

276



277

278 **Figure 2. Schematic overview of transcriptional silencing by PcG proteins in absence or**
279 **presence of rare variants in the DNA. A.** The recruitment of PcG proteins to the DNA results in
280 transcriptional silencing that is maintained after cell division. **B.** A disruption of PcG binding, e.g.
281 by rare variants, results in no binding or a weakened binding of the PcG proteins. A disrupted
282 binding leads to an altered gene expression where gene silencing is disturbed. The altered gene
283 expression is maintained after cell division.

284

285 ***ASTN2* as a possible regulatory target gene**

286

287 We were not able to identify regulatory targets of the replicated PRE based on significant *cis*-
288 eQTLs. Thus, the target gene could be assigned to the nearest gene, as done by Ringrose *et al.*⁴⁷
289 The nearest gene is *ASTN2*, of which the PRE is overlapping. The *ASTN2* gene encodes the
290 Astrotactin 2 protein. The Astrotactin 2 protein is expressed in the brain and is suggested to play a

291 part in neuronal migration. Thus, our findings support the hypothesis of migraine being a neuronal
292 disorder.⁴⁸

293

294 **Implications for genetic studies on migraine**

295

296 The most recent meta-analysis of migraine GWAS only explains 10.6% of the total heritability of
297 migraine.⁹ GWAS are often underpowered to detect rare variants with low-penetrant alleles.¹⁰ We
298 found that migraine associated with very rare variants, i.e. variants with a MAF<0.1%, in the
299 migraine risk loci. Our results indicate that a part of the unexplained heritability of migraine could
300 arise from rare variants not detected by GWAS. Also, our results suggest that the unexplained
301 heritability could be hidden in epigenetics. For discovering the unexplained heritability of migraine,
302 future studies focusing on the identification of heritable epigenetic marks for migraine can be highly
303 relevant.

304

305 **CONCLUSION**

306

307 We found that rare variants (MAF<0.1%) in a Polycomb Response Element in the *ASTN2* locus
308 associated with migraine. The association was independent of the original, common variant in the
309 locus. Thus, migraine GWAS loci also contain independent rare risk variants. We propose a
310 biological mechanism for migraine by which chromatin remodeling is disrupted by rare variants in
311 Polycomb Response Elements.

312

313 **ABBREVIATIONS**

314

315	GWAS	Genome-Wide Association Study
316	LD	Linkage disequilibrium
317	MAF	Minor allele frequency
318	PcG	Polycomb Group protein
319	PRE	Polycomb Response Element
320	TFBS	Transcription factor binding site
321	UTR	Untranslated region
322	WGS	Whole genome sequencing

323

324 **FUNDING**

325

326 The study was funded by a grant from Candys Foundation “CEHEAD” and a grant from Jascha
327 Fonden (web site: <https://jaschafonden.dk/>). The funders had no role in study design, data collection
328 and analysis, decision to publish, or preparation of the manuscript.

329

330 **COMPETING INTERESTS**

331

332 The authors have declared that no competing interests exist.

333

334 **DATA AVAILABILITY**

335

336 The data supporting the conclusions of this article are unavailable, as they contain information that
337 may be used for identifying study participants. However, summary data supporting the conclusions
338 of this article are available on request to the corresponding author of this article.

339

340 **ACKNOWLEDGEMENTS**

341

342 We thank deCODE genetics for whole genome sequencing and quality-controlling our samples. We
343 thank the Danish blood donors for their valuable participation in the Danish Blood Donor Study as
344 well as the staff at the blood centers for making this study possible. Finally, we thank all of the staff
345 and participating patients at the Danish Headache Center for their willing assistance.

346

347 **CONSORTIA**

348

349 Members list of the DBDS Genomic Consortium:

350 Steffen Andersen¹, Karina Banasik², Søren Brunak², Kristoffer Burgdorf³, Christian Erikstrup⁴,
351 Thomas Folkmann Hansen⁵, Henrik Hjalgrim⁶, Gregor Jemec⁷, Poul Jenum⁸, Per Ingemar
352 Johansson³, Kasper Rene Nielsen⁹, Mette Nyegaard¹⁰, Mie Topholm Bruun¹¹, Ole Birger
353 Pedersen¹², Susan Mikkelsen⁴, Khoa Manh Dinh⁴, Erik Sørensen³, Henrik Ullum³, Sisse
354 Ostrowski³, Thomas Werge¹³, Daniel Gudbjartsson¹⁴, Kari Stefansson¹⁴, Hreinn Stefánsson¹⁴,
355 Unnur Þorsteinsdóttir¹⁴, Margit Anita Hørup Larsen³, Maria Didriksen³, Susanne Sækmose¹².

356

357 ¹ Department of Finance, Copenhagen Business School, Copenhagen, Denmark.

358 ² Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences,
359 University of Copenhagen, Copenhagen, Denmark

360 ³ Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark

361 ⁴ Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

362 ⁵ Danish Headache Center, Department of Neurology, Copenhagen University Hospital, Glostrup,
363 Denmark

364 ⁶ Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

365 ⁷ Department of Clinical Medicine, Sealand University Hospital, Roskilde, Denmark

366 ⁸ Department of Clinical Neurophysiology, University of Copenhagen, Copenhagen, Denmark

367 ⁹ Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark

368 ¹⁰ Department of Biomedicine, Aarhus University, Aarhus, Denmark

369 ¹¹ Department of Clinical Immunology, Odense University Hospital, Odense, Denmark

370 ¹² Department of Clinical Immunology, Næstved Hospital, Næstved, Denmark

371 ¹³ Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen University
372 Hospital, Roskilde, Denmark

373 ¹⁴ deCODE genetics, Reykjavik, Iceland

374

375 REFERENCES

376

377 1. Stovner, L. J. *et al.* Global, regional, and national burden of migraine and tension-type

378 headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.

379 *Lancet Neurol.* **17**, 954–976 (2018).

380 2. Headache Classification Committee of the International Headache Society (IHS) The

381 International Classification of Headache Disorders, 3rd edition. *Cephalalgia* **38**, 1–211

382 (2018).

383 3. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a

384 systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England)*

385 **396**, 1204–1222 (2020).

- 386 4. Stovner, L. J. & Andrée, C. Impact of headache in Europe: a review for the Eurolight project
387 On behalf of the Eurolight Steering Committee. *J Headache Pain* **9**, 139–146 (2008).
- 388 5. Mulder, E. J. *et al.* Genetic and environmental influences on migraine: a twin study across
389 six countries. *Twin Res* **6**, 422–431 (2003).
- 390 6. Ulrich, V., Gervil, M., Kyvik, K. O., Olesen, J. & Russell, M. B. The inheritance of migraine
391 with aura estimated by means of structural equation modelling. *J. Med. Genet.* **36**, 225–227
392 (1999).
- 393 7. Hansen, R. D., Christensen, A. F. & Olesen, J. Family studies to find rare high risk variants
394 in migraine. *J Headache Pain* **18**, 32 (2017).
- 395 8. Levy, D., Strassman, A. M. & Burstein, R. A critical view on the role of migraine triggers in
396 the genesis of migraine pain. *Headache* **49**, 953–957 (2009).
- 397 9. Hautakangas, H. *et al.* Genome-wide analysis of 102,084 migraine cases identifies 123 risk
398 loci and subtype-specific risk alleles. *medRxiv* 2021.01.20.21249647 (2021)
399 doi:10.1101/2021.01.20.21249647.
- 400 10. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–
401 753 (2009).
- 402 11. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to
403 enhance discovery from GWAS. *PLoS Genet* **6**, e1000888 (2010).
- 404 12. Techlo, T. R. *et al.* Familial analysis reveals rare risk variants for migraine in regulatory
405 regions. *Neurogenetics* (2020) doi:10.1007/s10048-020-00606-5.
- 406 13. Eising, E., A Datson, N., van den Maagdenberg, A. M. J. M. & Ferrari, M. D. Epigenetic
407 mechanisms in migraine: a promising avenue? *BMC Med.* **11**, 26 (2013).
- 408 14. Winsvold, B. S. *et al.* Epigenetic DNA methylation changes associated with headache
409 chronification: A retrospective case-control study. *Cephalalgia* **38**, 312–322 (2018).

- 410 15. Corominas, R. *et al.* Association study of the serotonergic system in migraine in the
411 Spanish population. *Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. Off. Publ. Int. Soc.*
412 *Psychiatr. Genet.* **153B**, 177–184 (2010).
- 413 16. Lafreniere, R. G. *et al.* A dominant-negative mutation in the TRESK potassium channel is
414 linked to familial migraine with aura. *Nat Med* **16**, 1157–1160 (2010).
- 415 17. Nyholt, D. R. *et al.* A high-density association screen of 155 ion transport genes for
416 involvement with common migraine. *Hum. Mol. Genet.* **17**, 3318–3331 (2008).
- 417 18. de Vries, B. *et al.* Systematic re-evaluation of genes from candidate gene association studies
418 in migraine using a large genome-wide association data set. *Cephalalgia* **36**, 604–614
419 (2016).
- 420 19. Hansen, T. F. *et al.* DBDS Genomic Cohort, a prospective and comprehensive resource for
421 integrative and temporal analysis of genetic, environmental and lifestyle factors affecting
422 health of blood donors. *BMJ Open* **9**, e028401 (2019).
- 423 20. Chalmer, M. A., Hansen, T. F. & Olesen, J. Nosographic analysis of osmophobia and field
424 testing of diagnostic criteria including osmophobia. *Cephalalgia* **39**, 38–43 (2019).
- 425 21. Eriksen, M. K., Thomsen, L. L., Andersen, I., Nazim, F. & Olesen, J. Clinical characteristics
426 of 362 patients with familial migraine with aura. *Cephalalgia* **24**, 564–575 (2004).
- 427 22. Gervil, M., Ulrich, V., Olesen, J. & Russell, M. B. Screening for migraine in the general
428 population: validation of a simple questionnaire. *Cephalalgia* **18**, 342–348 (1998).
- 429 23. Rasmussen, B. K., Jensen, R. & Olesen, J. Questionnaire versus clinical interview in the
430 diagnosis of headache. *Headache* **31**, 290–295 (1991).
- 431 24. Nissen, L. *et al.* Danish study of Non-Invasive testing in Coronary Artery Disease (Dan-
432 NICAD): study protocol for a randomised controlled trial. *Trials* **17**, 262 (2016).
- 433 25. Jonsson, H. *et al.* Whole genome characterization of sequence diversity of 15,220 Icelanders.

- 434 *Sci. data* **4**, 170115 (2017).
- 435 26. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**,
436 2225–2229 (2002).
- 437 27. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
438 linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
- 439 28. Dreos, R., Ambrosini, G., Cavin Perier, R. & Bucher, P. EPD and EPDnew, high-quality
440 promoter resources in the next-generation sequencing era. *Nucleic Acids Res* **41**, D157-64
441 (2013).
- 442 29. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in
443 GeneCards. *Database (Oxford)* **2017**, (2017).
- 444 30. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**,
445 261–282 (1987).
- 446 31. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic
447 annotation of the human genome. *Nat Biotechnol* **28**, 817–825 (2010).
- 448 32. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE
449 data. *Nature* **489**, 91–100 (2012).
- 450 33. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
451 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–
452 2993 (2011).
- 453 34. Yan, Q. *et al.* A Sequence Kernel Association Test for Dichotomous Traits in Family
454 Samples under a Generalized Linear Mixed Model. *Hum Hered* **79**, 60–68 (2015).
- 455 35. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence
456 kernel association test. *Am J Hum Genet* **89**, 82–93 (2011).
- 457 36. Fisher, R. A. *Statistical methods for research workers*. (Oliver and Boyd) (1934).

- 458 37. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
- 459 38. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in
460 spatially structured populations. *Nat Genet* **44**, 243–246 (2012).
- 461 39. Pirie, A., Wood, A., Lush, M., Tyrer, J. & Pharoah, P. D. The effect of rare variants on
462 inflation of the test statistics in case-control analyses. *BMC Bioinformatics* **16**, 53 (2015).
- 463 40. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the
464 human genome. *Annu Rev Genomics Hum Genet* **7**, 29–59 (2006).
- 465 41. Klymenko, T. & Müller, J. The histone methyltransferases Trithorax and Ash1 prevent
466 transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).
- 467 42. Reinig, J., Ruge, F., Howard, M. & Ringrose, L. A theoretical model of Polycomb/Trithorax
468 action unites stable epigenetic memory and dynamic regulation. *Nat. Commun.* **11**, 4782
469 (2020).
- 470 43. Chan, C. S., Rastelli, L. & Pirrotta, V. A Polycomb response element in the Ubx gene that
471 determines an epigenetically inherited state of repression. *EMBO J.* **13**, 2553–2564 (1994).
- 472 44. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in
473 Polycomb repression. *Nat. Struct. Mol. Biol.* **21**, 569–571 (2014).
- 474 45. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*
475 **298**, 1039–1043 (2002).
- 476 46. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211
477 (2009).
- 478 47. Ringrose, L., Rehmsmeier, M., Dura, J.-M. & Paro, R. Genome-wide prediction of
479 Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev. Cell* **5**, 759–771
480 (2003).
- 481 48. Goadsby, P. J. Pathophysiology of migraine. *Neurol Clin* **27**, 335–360 (2009).

