

ARTICLE OPEN



Genome-wide by trait interaction analyses with neuroticism reveal chronic pain-associated depression as a distinct genetic subtype

Sandor Krause^{1,2,3}, Dora Torok^{1,3,4}, Gyorgy Bagdy^{1,3,4}, Gabriella Juhasz^{1,3,4} and Xenia Gonda^{1,2,3,4,5}✉

© The Author(s) 2025

The frequent co-occurrence of chronic pain (CP) and depression is a well-known phenomenon, supported by both the high prevalence of major depression among CP patients and studies describing a substantial genetic correlation between the two phenotypes. Neuroticism, a trait characterised by maladaptive stress responses and a tendency to experience negative emotions, has been linked to both depression and the experience of pain. This study aimed to determine whether depression associated with CP represents a genetically distinct subtype and to explore the role of neuroticism in modulating genetic susceptibility to depression. To address these questions, we performed genome-wide association analyses for current depression utilising the UK Biobank dataset, followed by genome-wide by trait interaction analyses to assess the interaction effect of neuroticism, and polygenic risk score analyses to compare predictions. Our findings suggest that CP-related depression is a valid subtype of depression. In association with current depression, we identified a total of 49 novel genetic risk polymorphisms meeting the genome-wide significance threshold, including variants involved in synaptic plasticity and transcriptional regulation. Additionally, our results support that neuroticism has a prominent role in modulating the genetic risk of current depression independently of CP, which highlights the importance of considering personality traits and stress factors in understanding the genetic background of complex and heterogeneous phenotypes like depression.

Translational Psychiatry (2025)15:108; <https://doi.org/10.1038/s41398-025-03331-5>

INTRODUCTION

The frequent co-occurrence of chronic pain (CP) and depression is a well-known phenomenon. The mean prevalence of major depressive disorder (MDD) among patients with CP is reported to be as high as 50 percent [1, 2]. Prior research has indicated a substantial genetic correlation between CP and depression [3–6]. Furthermore, Johnston et al. [7] have demonstrated a causal effect of multisite CP on MDD with widespread pleiotropy. Other authors emphasised the differentiation between chronic and recent pain [8, 9], which was substantiated by the observation that the polygenic risk score (PRS) for MDD only predicted the latter. Based on these findings, Campos et al. [8] conceptualised a depression subtype characterised by comorbidity of CP and poor treatment response.

Our research group has previously described a transitive relationship between pain-related variables and lifetime depression mediated by current depressive symptoms and, specifically in the case of headache, by neuroticism [10]. The association between neuroticism and the experience of pain is widely recognized [4, 6, 11]: trait neuroticism may predispose to maladaptive coping [12], lower pain thresholds [13], and result in a hypervigilance towards pain [14]. Furthermore, a twin study conducted by

researchers at VU Amsterdam suggested that a neurotic personality may increase the risk of developing migraine [15].

Trait neuroticism is a relatively stable and conveniently assessable [16, 17] aspect of an individual's personality, which affects a wide range of psychological processes beyond pain perception [11]. Most importantly, it contributes to poor adaptation and reaction to stressful conditions and life events in general, and to an increased likelihood of experiencing negative emotions. There is extensive research confirming the association between high neuroticism and the development and recurrence of depression in part related to stress sensitivity and reactivity and difficulties in regulating emotions [18]. Furthermore, longitudinal studies have reported that trait neuroticism predicts both the onset and recurrence of major depressive episodes independently of environmental stress [19]. Subsequent meta-analyses have further confirmed this relationship, also revealing that neuroticism contributes to depression through increased risk of maladaptive coping strategies and rumination [20, 21]. Very recently, the central role of neuroticism in the development of affective disorders has been reconfirmed [22]. Importantly, genetic studies have also pointed toward an association between neuroticism and depression [23].

¹Department of Pharmacodynamics, Faculty of Pharmaceutical Sciences, Semmelweis University, Budapest, Hungary. ²Department of Psychiatry and Psychotherapy, Semmelweis University, Budapest, Hungary. ³Center of Pharmacology and Drug Research & Development, Semmelweis University, Budapest, Hungary. ⁴NAP3.0-SE Neuropsychopharmacology Research Group, Hungarian Brain Research Program, Semmelweis University, Budapest, Hungary. ⁵Department of Clinical Psychology, Semmelweis University, Budapest, Hungary. ✉email: gonda.xenia@semmelweis.hu

Received: 31 August 2024 Revised: 23 February 2025 Accepted: 18 March 2025

Published online: 29 March 2025

In our present study, we wanted to investigate if depression occurring in the context of CP is a distinct genetic subtype of depression. In order to test this concept, we compared the genetic architecture of depression in two groups distinguished by the occurrence or absence of CP using genome-wide association analysis (GWAS). Furthermore, we hypothesised that the interaction effect of the genetic variation and neuroticism on depression would be more prominent in the CP group, which we investigated by performing a genome-wide by trait interaction study. Finally, we validated whether PRS scores calculated from the results of the above GWAS studies would be able to predict the presence of current depression and the presence of a depression diagnosis as a function of CP in an independent target sample.

MATERIALS AND METHODS

Participants

We used the UK Biobank dataset (Application Number 1602) as a foundation for our association analyses. As our analyses, specifically calculating and testing PRS, required a discovery and a target dataset that were independent of one other at the individual level, we divided the UK Biobank cohort into two samples according to the availability of phenotypic data: the discovery sample comprised those with no missing data on the Patient Health Questionnaire (PHQ-9) assessing current depression severity, while all others with missing data were included in the target sample. Both samples underwent the same quality control steps.

We performed a phenotype-based filtering to exclude possible confounders: we excluded every individual with a medical history of bipolar disorder, any psychotic condition, mental retardation, or substance use disorder using 'First occurrences' (Category 2405) and UK Biobank data-fields 20002 and 20126 [24]. The data-fields of Category 2405 contain data showing the 'first occurrence' of any code mapped to 3-character ICD-10, aggregating primary care data, hospital inpatient data, death register records, and self-reported medical condition codes derived from data-field 20002, reported at the UK Biobank assessment centre visit. Whereas data-field 20126 contains data on bipolar and major depression status, including probable cases, based on prior research at the University of Glasgow [25].

Furthermore, we only included participants with white British ancestry (data-field 22006) who had no missing values regarding sex (data-field 31), age (data-field 21003), genotyping array (data-field 22000), and the phenotypes of interest resulting in a total of 121,574 subjects in the discovery sample, and 196,167 subjects in the target sample.

All participants provided written informed consent, and all procedures were carried out in accordance with the Declaration of Helsinki. Details of the data collection were previously outlined by Bycroft and colleagues [26]. Ethical approval was given by the National Research Ethics Service Committee North West-Haydock. Further information on sample characteristics and the number of subjects excluded can be found in Supplementary Table 1.

Phenotypes

For grouping, we derived a dichotomous variable (CP status) based on UK Biobank data-fields 3799, 4067, 3404, 3571, 3741, 3414, 3773, and 2956, which measured the different types of pain experienced by participants for more than 3 months [5]. If any of the above items indicated pain, we assigned the subject to the CP group ($N = 47,313$). If all questions were answered in the negative, the subject was sorted into the no-pain (NP) group ($N = 74,261$). As our data on CP do not reflect its temporality relative to the diagnosis of MDD, we have opted to measure genetic susceptibility to current depression rather than lifetime depression. To assess current depression, we used the total score of PHQ-9 (UK Biobank data-fields 20514, 20510, 20517, 20519, 20511, 20507, 20508, 20518, and 20513) as primary phenotype which has been proven to be a reliable and valid measure of depression severity [27] with good internal consistency (Cronbach's $\alpha = 0.845$). Neuroticism trait was measured by 12 items (data-fields 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020, and 2030) from the Eysenck Personality Inventory Neuroticism scale (EPIN-R) [16, 25], which also has good internal consistency and a Cronbach's α of 0.834. We classified neuroticism further into 'low', 'medium', and 'high' categories for visualisation purposes, as described in the supplementary materials by Hullam and colleagues [10].

In the target sample, where PHQ-9 data were not available, we used the paper of Hullam et al. [10] as a reference and measured current depression using the 'current depressive symptoms score' based on UK Biobank data-fields 2050, 2060, 2070, and 2080. In addition, in the target sample, we used lifetime depression derived from 'First occurrences', based on ICD-10 codes F32 (data-field 130894) and F33 (data-field 130896).

To test for multicollinearity, we computed a matrix of Pearson's correlation and we found a weak but significant correlation between PHQ-9 and EPIN-R scores in both discovery samples with a coefficient of 0.4002 ($p < 2.2 \times 10^{-16}$, CI: 0.3918–0.4085) in for CP and 0.3801 ($p < 2.2 \times 10^{-16}$, CI: 0.3733 – 0.3868) for the NP group. In our target sample, we found a similar correlation between EPIN-R and 'Current depression score' with a Pearson's coefficient of 0.5505 ($p < 2.2 \times 10^{-16}$, CI: 0.5474–0.5535). Since all of these correlation coefficients are small, they do not interfere with the interpretation of the results.

Genetic data

We followed the same genomic quality control steps for both genetic datasets as described in the supplementary materials of Eszlar and colleagues [28]. We performed filtering steps (minor allele frequency, Hardy-Weinberg equilibrium, missingness, and linkage disequilibrium pruning) and carried out kinship check, sex check, heterozygosity outlier detection, and principal component analysis [28]. Regarding UK Biobank data, 38,939 samples remained in the CP group, counting 6,214,289 genetic variants, and 61,641 samples in the NP group, counting 6,223,429 genetic variants, which were consequently used in our genome-wide association analyses (GWAS).

In the case of the target dataset, we excluded genetic variants on the sex chromosomes and performed filtering based on HapMap3+ variants recommended by Privé and colleagues [29], which provide good coverage of the whole genome. After matching the variants in the summary statistics and the variants in the target sample, 972,321 single-nucleotide polymorphisms (SNP) remained in the CP group and 974,515 SNPs in the NP group, which served as a basis for the polygenic risk calculation.

Analyses

SNP-level analyses. We performed separate GWASs in the CP and NP groups of the discovery sample to examine the direct effect of genetic variants on the PHQ-9 score, followed by a genome-wide by trait interaction analysis to assess the SNP \times Neuroticism effect on the same phenotype. In order to eliminate the effect of neuroticism on the SNP main effect, we included it as a covariate in addition to the top ten genomic principal components (PC), sex, age, and genotyping array. All SNP-based association tests were computed using linear regression models in PLINK 2.0 (www.cog-genomics.org/plink/2.0/) [30, 31]. In this way, a total of 4 models were tested and taken into account when adjusting for multiple comparisons.

Genetic correlation. We used the SumHer software (<https://dougsspeed.com/sumher/>) [32], assuming LDAK-Thin heritability model [33], to estimate genetic correlation between the two GWAS summary statistics, containing information from the discovery sample about the genetic predisposition to current depression in the NP and CP groups.

Functional analyses. We conducted functional annotation for all genetic variants within genomic regions identified by lead SNPs by employing FUMA v1.5.2 (<https://fuma.ctglab.nl/>) [34, 35]. Lead SNPs were defined as independent SNPs ($R^2 \leq 0.5$) within a locus that had the lowest p-value within a 10,000 base pair range. A significance threshold of $p \leq 5.0 \times 10^{-8}$ was applied for lead SNPs, while $p \leq 1.0 \times 10^{-5}$ was considered for suggestive significance. To adjust for multiple comparisons, we used a Bonferroni-corrected threshold of $p \leq 2.0115 \times 10^{-9}$ for the CP group and $p \leq 2.0085 \times 10^{-9}$ for the NP group.

Gene-based tests. By performing gene-based tests with MAGMA v1.08 (<https://ctg.cncr.nl/software/magma>) [36], SNP-based p-values were aggregated into gene-based p-values using a SNP-wise mean model. SNPs were assigned to protein-coding genes based on their positions, with gene boundaries extended by 10,000 base pairs. The number of mapped genes was 19,884 in the CP group and 19,886 in the NP group, resulting in Bonferroni-adjusted thresholds of $p \leq 6.2865 \times 10^{-7}$ and $p \leq 6.2858 \times 10^{-7}$, respectively.

Gene-set analyses. The MAGMA gene-set analysis [36] tests whether the genes in a gene-set are jointly associated with the phenotypes by regressing transformed gene p-values on the gene-set indicator variable. Gene-set definitions were derived from MsigDB v2023.1.Hs [37], including curated gene-sets (C2) and Gene Ontology gene-sets [38, 39] (C5 GO:BP - biological process; C5 GO:CC - cellular component; C5 GO:MF - molecular function). In our analyses, this entailed 17,012 gene-sets and thus a Bonferroni-corrected p-value threshold of 7.3478×10^{-7} for both groups.

To overcome the limitations of using only position-based assignments and protein-coding genes within MAGMA, top SNPs of SNP-level tests were mapped to genes based not only on position but also on functional annotations. The mapped genes were further analysed using hypergeometric tests to check for overrepresentation within predefined gene-sets. In these enrichment analyses conducted through the 'GENE2FUNC' function of FUMA [34, 35], a Benjamini-Hochberg FDR adjusted threshold of $p \leq 0.05$ was set for each subcategory of gene-sets. For this study, the following MsigDB subcategories were considered: C3 MIR (microRNA targets), C3 TFT (transcription factor targets), and the above-mentioned three subcategories from Gene Ontology gene-sets.

Calculation of polygenic risk scores. We used the automatic model of the LDpred2 software [29, 40] (LDpred2-auto in short), which calculates polygenic risk by inferring and aggregating the posterior mean effect size of each SNP, based on prior knowledge from the above GWAS summary statistics, taking into account information on linkage disequilibrium (LD) and genetic architecture. This method requires two hyper-parameters: heritability and the proportion of causal variants, which LDpred2-auto can infer directly using multiple iterations and an initial heritability estimate generated by LD Score regression [41].

We calculated PRS scores on the target sample providing effect sizes from our own GWAS summary statistics on PHQ-9 total score. As both the target sample and the GWAS summary statistics were sourced from the UK Biobank database, we only performed a basic SNP-based harmonisation restricted to HapMap3+ variants [29]. We carried out a genome-wide computation running 50 Gibbs sampler chains in parallel. As a result, 2 separate PRS scores were obtained, namely PRS_{CP} and PRS_{NP}, which encompassed the genetic predisposition to current depression in the CP and NP groups, respectively.

Evaluation of polygenic risk scores. We tested each PRS score with generalised linear models in the target sample, using the PRS score and the same set of covariates as predictors as in the GWAS analyses regarding SNP main effects. We then computed an estimate of the predictive ability of the PRS score by subtracting R^2 between this model and the corresponding covariate-only model and denoted it ΔR^2 . In a subsequent step, we performed another set of regression analyses to assess the explained phenotypic variance by the interaction of neuroticism and the genetic risk for depression, embodied in our PRS scores. Eqs 1, 2 provide the R formulas of the main and interaction models, respectively.

Main model:

$$\text{Outcome} \sim \text{Age} + \text{Sex} + \text{Array} + \text{PCs} + \text{Neuroticism} + \text{PRS} \quad (1)$$

Interaction model:

$$\text{Outcome} \sim \text{Age} + \text{Sex} + \text{Array} + \text{PCs} + \text{Neuroticism} + \text{PRS} + \text{PRS} : \text{Neuroticism} \quad (2)$$

For our PRS analyses in the target sample, two outcome variables were considered: current depression as measured by the current depressive symptoms score, and lifetime depression reflecting the presence of a depression diagnosis. We used linear regression for the first and logistic regression for the second variable. In order to allow for a comparison of the goodness of fit of linear and logistic regression models, a pseudo R^2 was calculated for the latter using the method proposed by Nagelkerke [42]. Finally, we fitted additional models to compare the predictive performance of PRS scores across subsamples stratified by CP status (CP target subsample: $N = 85,420$; NP target subsample: $N = 110,747$). Altogether, a total of 24 generalised linear models were applied, and p-values were adjusted accordingly using Bonferroni correction.

Data cleaning, statistics and visualisation. Data cleaning, descriptive statistics, and regression analyses testing the predictive effect of the calculated PRS scores were conducted using R version 4.1.2 (<https://www.r-project.org/>) [43] and the tidyverse meta-package [44]. The R packages

bigstatsr (v1.5.12) and bigsnpr (v1.12.2) were used as part of the LDpred2 method [45]. Manhattan plots were created via topR (v2.0.2) [46, 47], all other plots were made using ggplot2 [48]. Nagelkerke's R^2 was computed using the R package fmsb (v0.7.6) [49, 50].

RESULTS

Main effect analyses in the NP and CP groups

SNP-level findings. Regarding the main effect of SNPs, 18 variants in the NP group and 11 in the CP group reached suggestive significance with no overlap between them. None of these variants met the genome-wide significance threshold or survived multiple testing correction. Fig. 1 presents these findings in a Manhattan plot. A complete list of the genetic variants organised by groups can be found in Supplementary Table 2 and Supplementary Table 3.

Results of MAGMA analyses. Gene-based tests and gene-set analyses performed using MAGMA yielded no significant results in either group.

Gene-set enrichment analyses. In the NP group, a single gene-set from GO:BP was identified (Supplementary Table 6), whereas in the CP group 5 gene-sets from GO:CC and 10 gene-sets from GO:MF remained significant (Supplementary Table 7).

Genetic correlation for current depression between the CP and NP groups

We found a genetic correlation of 0.9102 ($SE = 0.2048$) between the two GWAS summary statistics, with a heritability estimate of 0.0352 ($SE = 0.0096$) for current depression in the NP group and 0.0767 ($SE = 0.0148$) in the CP group.

Interaction analyses between genetic variants and neuroticism in the NP and CP groups

SNP-level findings. Regarding SNP \times Neuroticism interactions, 165 variants reached suggestive significance in the CP group, of which 2 variants (rs61096498: $\beta = -0.0940$, $p = 6.6414 \times 10^{-9}$; rs10941193: $\beta = 0.1602$, $p = 3.4766 \times 10^{-8}$) met the genome-wide significance threshold, but did not survive the Bonferroni-correction (Supplementary Table 2). In the NP group, of the 777 suggestive significant variants, 47 reached the genome-wide significance threshold and 5 (rs6897383: $\beta = -0.0346$, $p = 4.0404 \times 10^{-10}$; rs2769569: $\beta = 0.0321$, $p = 8.5806 \times 10^{-10}$; rs12345850: $\beta = 0.0799$, $p = 1.0054 \times 10^{-9}$; rs138137561: $\beta = 0.1051$, $p = 1.0401 \times 10^{-9}$; rs75133540: $\beta = -0.0908$, $p = 1.9418 \times 10^{-9}$) remained significant after adjusting for multiple testing (Supplementary Table 3). Furthermore, we identified only 1 overlapping variant between the interaction analysis and the main effect analysis results in each group, namely rs35519714 in the CP group and rs72762916 in the NP group. A Manhattan plot summarising the SNP-level findings of these interaction analyses is presented in Fig. 2.

Results of MAGMA analyses. Gene-based tests showed significant associations with 16 genes in the NP group, and with 1 gene ($SDC2$: $p = 3.078 \times 10^{-7}$) in the CP group (Supplementary Table 4). Among the former, the top four significant findings were $ANXA13$ ($p = 5.0656 \times 10^{-9}$), $SRSF10$ ($p = 1.4593 \times 10^{-8}$), $TSHZ2$ ($p = 2.2068 \times 10^{-8}$), and $HDAC4$ ($p = 2.4605 \times 10^{-8}$). MAGMA gene-set analyses identified a single significant gene-set (GOBP_HISTONE_H3_K9_DIMETHYLATION: $p = 2.18 \times 10^{-6}$) in the CP group, compared to none in the NP group (Supplementary Table 5).

Gene-set enrichment analyses. In the NP group, we identified 271 microRNA targets and 25 transcription factor targets, as well as 5 GO gene-sets: 2 from GO:BP, 2 from GO:CC, and 1 from GO:MF,

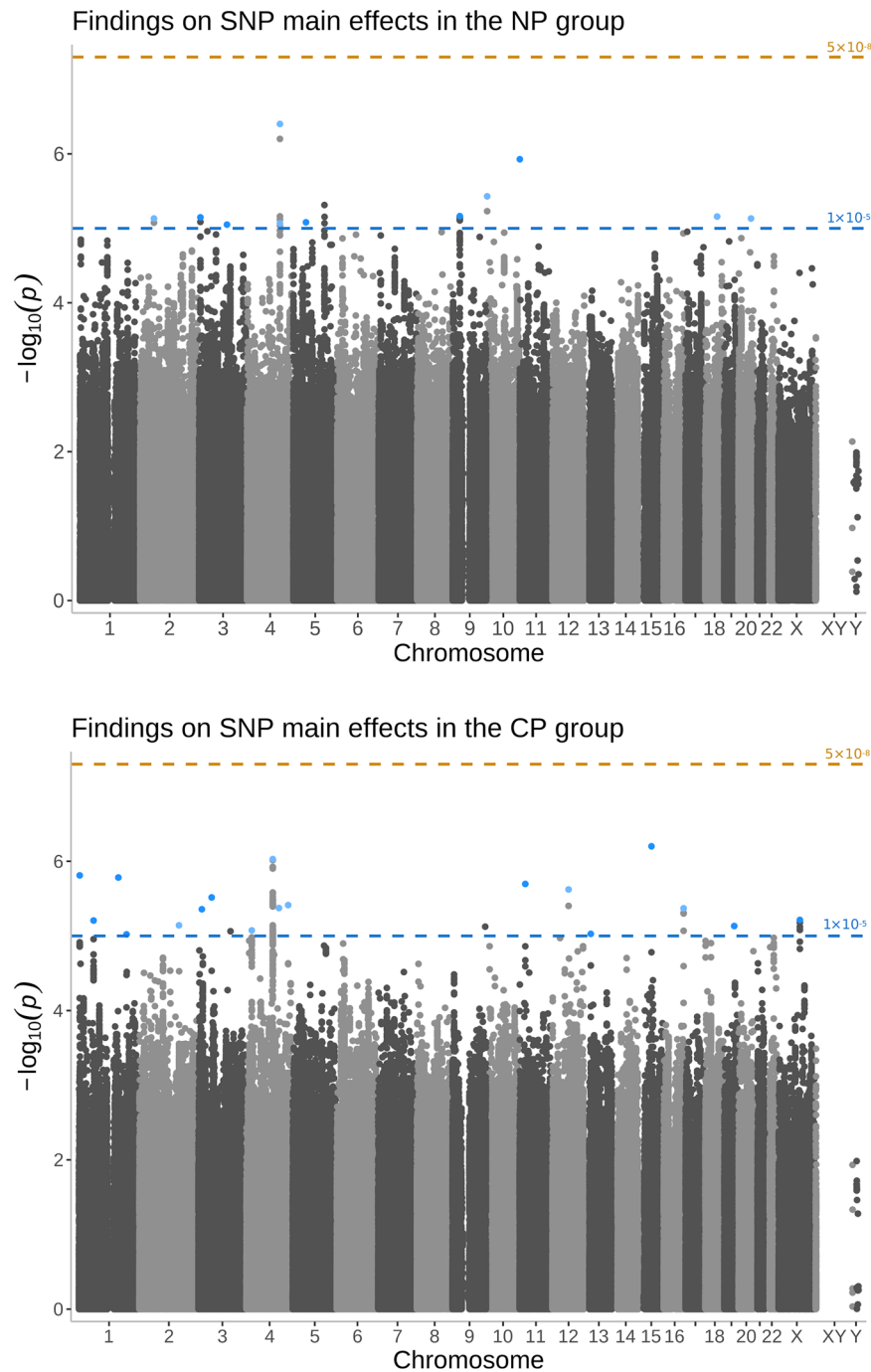


Fig. 1 Manhattan plots presenting genetic variants with suggestive significance identified by our main effect analyses in the no-pain (NP) and chronic pain (CP) groups. The plot above displays results from the no-pain (NP) group, while the plot below displays data from the chronic pain (CP) group. The x-axis shows the physical position of genetic variants, arranged by chromosomes. The y-axis shows the level of significance on a negative logarithmic scale, with a blue line denoting the suggestive significance threshold ($p \leq 1.0 \times 10^{-5}$) and an orange line denoting the genome-wide significance threshold ($p \leq 5.0 \times 10^{-8}$). The highlighted single-nucleotide polymorphisms (SNP) are lead SNPs, each within a locus of 10,000 base pair range, colour-coded according to significance.

respectively. In comparison, the results for the CP group showed 18 significant gene-sets from GO:BP, 7 from GO:CC, and 2 from GO:MF, with 4 additional transcription factor targets. A complete list of identified gene-sets is presented in Supplementary Table 6 and Supplementary Table 7, organised by groups.

Polygenic risk score analyses in the target sample

Predictive effect of PRS scores on current depression. Both PRS_{NP} and PRS_{CP} , representing an aggregated genetic risk for

depression based on data from the NP and CP groups respectively, made a significant contribution to the CDS score of the total target sample (PRS_{NP} : $\beta = 1.4197$, $p_{adj} < 2 \times 10^{-16}$, $\Delta R^2 = 4 \times 10^{-4}$; PRS_{CP} : $\beta = 1.5671$, $p_{adj} < 2 \times 10^{-16}$, $\Delta R^2 = 6.538 \times 10^{-4}$). When looking at the results of the interaction models, the main effects of these PRS scores seemingly diminished and were taken over by the interaction effect of $PRS \times$ Neuroticism ($PRS_{NP} \times$ Neuroticism: $\beta = 0.2569$, $p_{adj} = 5.7953 \times 10^{-9}$, $\Delta R^2 = 1 \times 10^{-4}$; $PRS_{CP} \times$

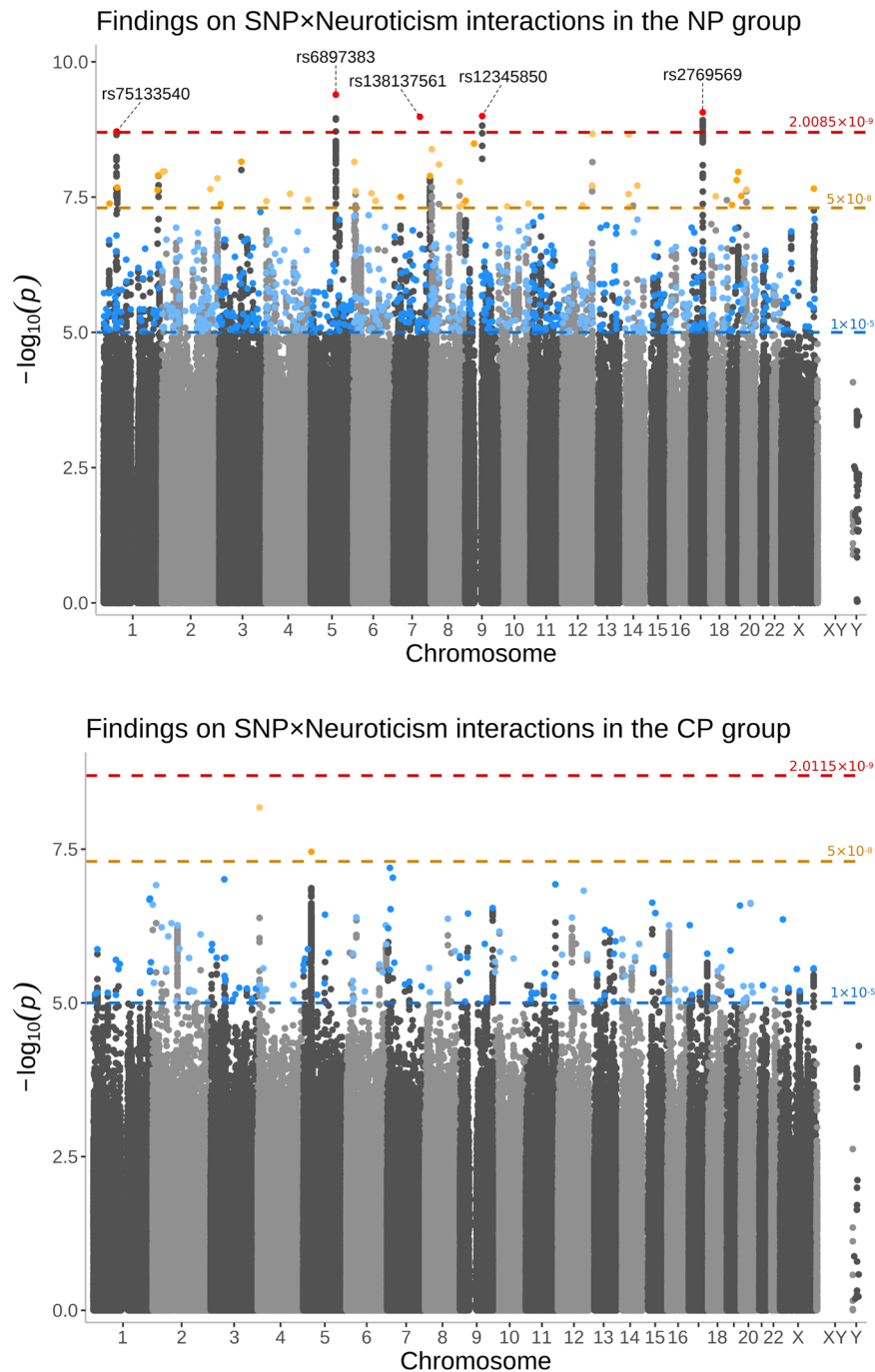


Fig. 2 Manhattan plots presenting significant genetic variants identified by our interaction analyses between genetic variants and neuroticism in the no-pain (NP) and chronic pain (CP) groups. The plot above displays results from the no-pain (NP) group, while the plot below displays data from the chronic pain (CP) group. The x-axis shows the physical position of genetic variants, arranged by chromosomes. The y-axis shows the level of significance on a negative logarithmic scale, with a blue line denoting the suggestive significance threshold ($p \leq 1.0 \times 10^{-5}$), an orange line denoting the genome-wide significance threshold ($p \leq 5.0 \times 10^{-8}$), and a red line denoting the Bonferroni-corrected threshold ($p \leq 2.0115 \times 10^{-9}$ for the CP and $p \leq 2.0085 \times 10^{-9}$ for the NP group, respectively). The highlighted single-nucleotide polymorphisms (SNP) are lead SNPs, each within a locus of 10,000 base pair range, colour-coded according to significance. Lead SNPs surviving multiple testing correction were annotated with the corresponding reference SNP identifier (rsID) assigned by dbSNP.

Neuroticism: $\beta = 0.2794$, $p_{\text{adj}} = 1.3019 \times 10^{-15}$, $\Delta R^2 = 3 \times 10^{-4}$). These findings were independent of CP status and remained consistent across subsamples. The predictive effect of PRS scores on current depression is visualised in Fig. 3, stratified by neuroticism levels to better illustrate their interaction. A comprehensive presentation of these results is available in Supplementary Table 8.

Predictive effect of PRS scores on lifetime depression. Regarding our logistic regression analyses, we observed that PRS_{CP} had a significant predictive effect in the main model ($\beta = 1.2317$, $p_{\text{adj}} = 1.9192 \times 10^{-6}$, $\Delta R^2 = 3 \times 10^{-4}$), while PRS_{NP} did not ($\beta = 0.8228$, $p_{\text{adj}} = 7.5235 \times 10^{-2}$, $\Delta R^2 = 9.09 \times 10^{-5}$). When considering target subsamples stratified by CP status, the main effect of PRS_{CP} appeared to be significant only in the CP target subsample

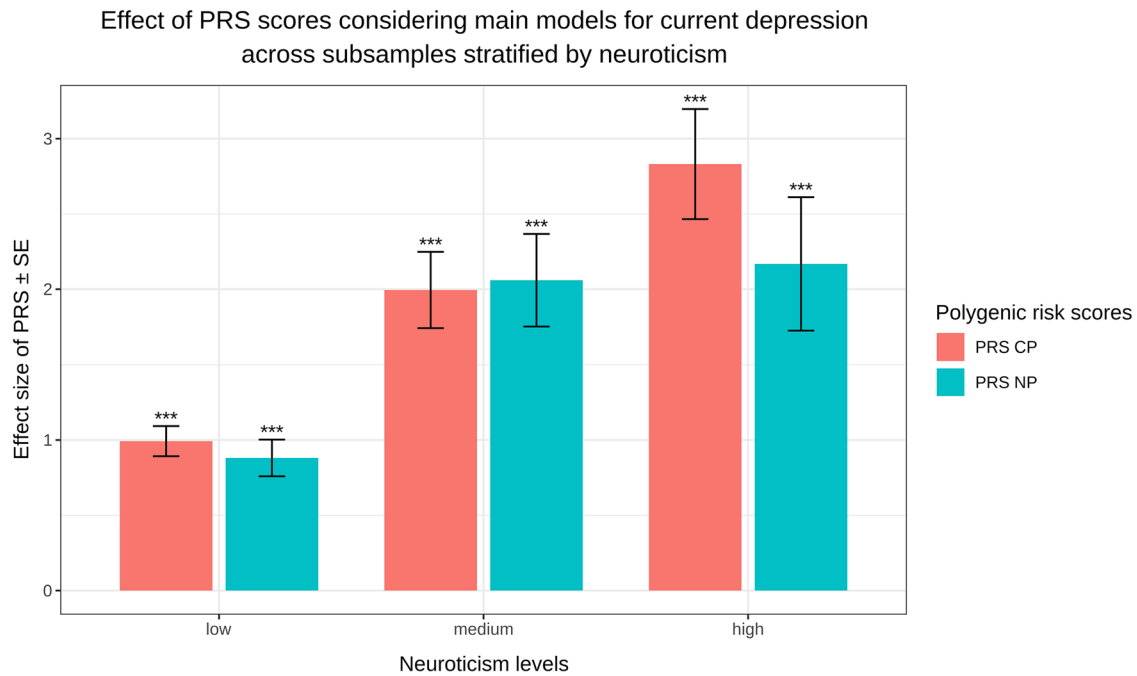


Fig. 3 A clustered bar chart illustrating the predictive effect of polygenic risk scores (PRS) representing an aggregated genetic risk for depression based on data from the no-pain (PRS_{NP}) and chronic pain (PRS_{CP}) groups on current depression in the total target sample stratified by neuroticism. The x-axis ranks the categories from low neuroticism to high neuroticism. The y-axis shows the effect sizes of PRS scores derived from our main regression models, with an error bar indicating standard error (SE). Following conventions, asterisks on top of the bars indicate the level of significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$). The colours specified in the legend represent results for PRS scores calculated in the chronic pain (CP) and no-pain (NP) groups, respectively.

(CP target subsample: $\beta = 1.2077$, $p_{\text{adj}} = 1.9597 \times 10^{-3}$, $\Delta R^2 = 3.2 \times 10^{-4}$; NP target subsample: $\beta = 0.9614$, $p_{\text{adj}} = 1.3734 \times 10^{-1}$, $\Delta R^2 = 1.686 \times 10^{-4}$). In contrast, PRS_{NP} did not contribute to the explained phenotypic variance in either subsample. Looking at the interaction models, we found no significant effect on lifetime depression for either PRS_{NP} or PRS_{CP} in interaction with neuroticism. The predictive effect of PRS scores on lifetime depression is illustrated in Fig. 4. For more details, see Supplementary Table 9.

DISCUSSION

In our study, we investigated the genetic background of depression occurring in the presence or absence of chronic pain, and how trait neuroticism interacts with genetic variation on current depression in these two depression subtypes. Finally, we also examined if PRS scores calculated in the above analyses significantly predicted current depression or depression diagnoses in an independent target sample, either independently or in interaction with neuroticism.

Investigation of the genetic architecture of depression in interaction with neuroticism may suggest distinct subtypes in the presence and absence of chronic pain

In the CP group, we observed higher means for depression and neuroticism scores, and the SNP heritability estimate of current depression was double in the CP group, which can be attributed to the fact that the CP group is more homogeneous, and depression scores, which tend to change dynamically in general, may be more persistent here, representing a continuous depressive state as consequence to CP. Although we found a high genetic correlation between the two groups, this was not unexpected given that we were making a comparison based on the same trait. In comparison, Meng et al. reported genetic

correlations across eight pain phenotypes, with estimates ranging from 0.36 to 0.83 [6].

In our main effect analyses we could not identify significant variants, genes, or gene-sets associated with current depression either in the CP or NP groups after correcting for multiple testing.

However, when the interaction effect of neuroticism was considered, a large number of significant variants were identified, especially in the NP group, which, besides confirming the importance of neuroticism in the manifestation of genetic vulnerability, contradicts our hypothesis that neuroticism plays a greater role in the appearance of depressive symptoms in the CP group. In the interaction analyses, regarding the lead SNPs of the risk loci, we did not find any overlapping variants between the CP and NP groups, which once again highlights differences between depression that manifests in the presence and absence of CP.

Of the lead SNPs identified, only one variant has so far been reported in the context of depression, namely rs17634917 in the CP group regarding SNP \times Neuroticism interaction, which appears to mediate a predisposing effect. It has been mapped to the *LOC105372465*, *RFPL4A* and *NLRP9* genes, the first two of which regulate transcription and the latter plays a role in the innate immune system. All other hits can be considered as new associations.

PRS analyses in an independent target sample also support differences between depression in the presence and absence of chronic pain

In addition to the above genome-wide by trait interaction studies suggesting that CP-associated depression may be a distinct depression subtype, our PRS studies pointed in a similar direction. Whereas for current depression, we did not find differences in the predictive power of the two PRS scores, we discovered a meaningful difference in the prediction of lifetime depression: PRS_{CP} predicted depression diagnoses in the CP target subsample,

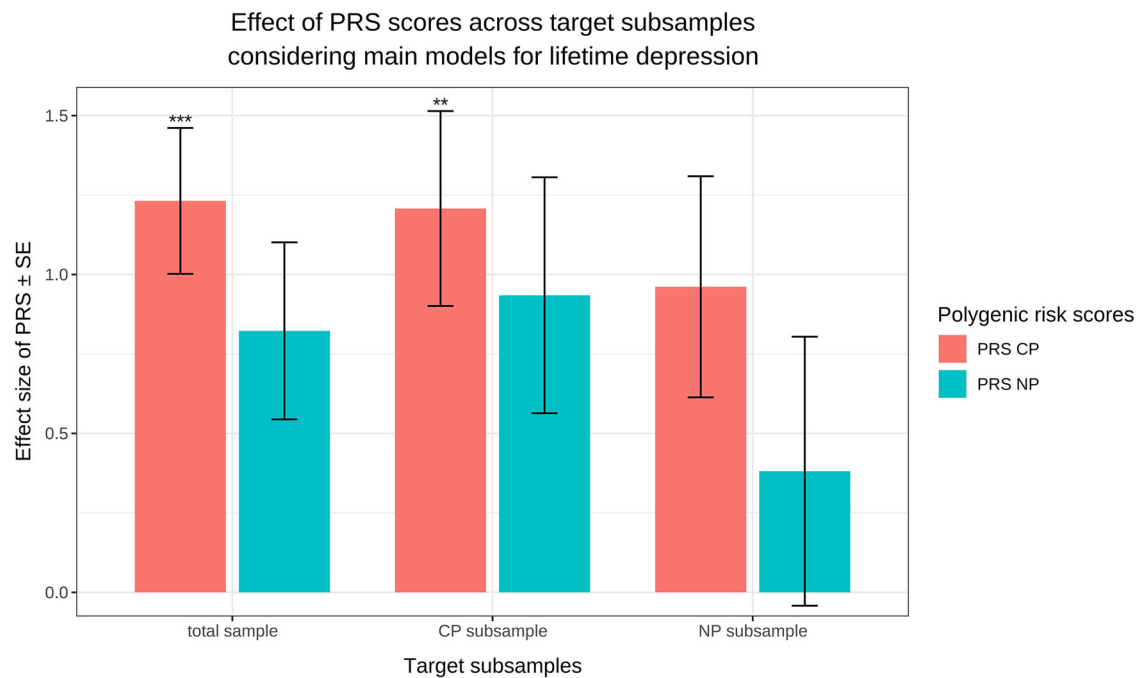


Fig. 4 A clustered bar chart illustrating the predictive effect of polygenic risk scores (PRS) representing an aggregated genetic risk for depression based on data from the no-pain (PRS_{NP}) and chronic pain (PRS_{CP}) groups on lifetime depression in the target sample stratified by chronic pain status. The x-axis lists the total target sample and two target subsamples based on the presence and absence of chronic pain. The y-axis shows the effect sizes of PRS scores derived from our main regression models, with an error bar indicating standard error (SE). Following conventions, asterisks on top of the bars indicate the level of significance (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$). The colours specified in the legend represent results for PRS scores calculated in the chronic pain (CP) and no-pain (NP) groups, respectively.

contrary to the NP group, supporting the concept of a distinct depression subtype.

The role of neuroticism in the manifestation of depressive symptoms is independent of chronic pain status

In line with previous knowledge, our results show that neuroticism has a great influence on the genetic risk of developing depressive symptoms, however, contrary to our expectations, this effect was largely independent of CP status. This former finding is also reflected in the effect of PRS scores on current depression, which is best seen in our interaction models, where the previously significant effect of PRS scores is mediated entirely through the $PRS \times$ Neuroticism effect.

Our analyses confirm variants previously associated with depression

For the main effect analysis in the NP group, we detected one suggestively significant genetic variant mapped to the *DCC* gene. *DCC* is the receptor of Netrin-1, together they play an important role in guiding the growth cones during adolescent neurodevelopment. There is converging evidence that this signalling pathway may be affected in multiple mental disorders, including MDD [51].

Furthermore, rs72762916, the sole overlapping variant identified with a suggestive significance in the NP group in both the main effect and interaction analysis results, was mapped to *ADARB2*. The *ADARB2* gene encodes an RNA-editing enzyme expressed primarily in inhibitory neurons. Interestingly, in addition to its polymorphisms that predispose to depression [52, 53], it has recently been suggested as a potential contributor to protection from headache [54].

In the CP group, we found two suggestive significant polymorphisms mapped to the genes *ERC2* and *CDH13* respectively, the former playing a role in circadian regulation and neuroplasticity [55], the latter encoding a protein localised to the cell surface that protects against oxidative stress and regulates axon growth during

neural development [56]. Both genes have been linked to depression [52, 57, 58], but a polymorphism of the *CDH13* gene has also been associated with an increased risk of migraine [59].

Several identified variants and genes are in line with previous results concerning the interaction of stress in the background of depression

From our results in the CP group, we highlight two genes: *SORCS2* and *SDC2*. In interaction with neuroticism, we identified two protective polymorphisms mapped to the former gene: rs4689765 with a suggestive significance and rs61096498, which was significant at the genome-wide level. *SORCS2* is known to play a role in the BDNF-dependent synaptic plasticity. It has been already linked to neuroticism [60], but more interestingly, Chen et al. [61] found that the hippocampal overexpression of *SORCS2* repressed chronic stress-induced depressive-like behaviours in mice. Our observations suggest that *SORCS2* might be affected in CP-induced depression as well.

SDC2, on the other hand, is a member of the syndecan family, involved in cell binding, cell signalling, and cytoskeletal organisation, which we identified in association with current depression via MAGMA gene-based tests, drawing on our results from the genome-wide by trait interaction analysis. It has been reported in the context of inflammation and blood-brain barrier disruption [62] and as a candidate gene for post-traumatic stress disorder (PTSD) [63] and suicidal ideation [64]. It is well documented that CP has a higher comorbidity rate with PTSD [65], moreover, prior research has demonstrated that targeting symptoms of PTSD may improve pain [66], implying common neurobiological mechanisms between the two phenotypes.

Among the results of the interaction analysis in the NP group, we identified three significant variants mapped to the *LOC102724945* RNA gene and two variants, one of which was only suggestive, mapped to the *TRPM3* gene. *TRPM3* encodes a

cation-selective channel and is involved in cellular calcium signalling and homeostasis. Coleman et al. [67] reported polymorphisms in the aforementioned genes which conveyed a susceptibility to MDD only in the subgroup that had not experienced trauma. Our results both originated from the NP group that had previously been pruned of CP, and the same genetic variants had no significant effect in the CP group. This again suggests, albeit indirectly, a link between the neurobiology underlying trauma and CP.

Our analyses identify genes with pleiotropic effects and underscore the importance of transcriptional regulation in the development of depression

We mapped 3 genetic variants among the top findings of the main effect analysis in the NP group to the *CSMD1*, *HDAC4* and *RGS6* genes, respectively. *CSMD1* is an integral membrane component affecting several processes, including tumour suppression, complement-mediated synaptic pruning, and immune-related signalling [68]. *RGS6* regulates G-protein signalling and is enriched in hippocampal and cortical neurons. It was also implicated as a regulator of 5-HT_{1A}R signalling, thus modulating anxiety and depression [69]. Lastly, *HDAC4* is a transcriptional regulator which is also involved in neuronal synaptic plasticity and memory formation [70].

What these genes have in common is that they have been all reported in the context of pleiotropy, affecting multiple mental disorders [71, 72]. In contrast, our CP group is a more homogenous and specially selected sample, where SNPs mediating general susceptibility may play a minor role.

HDAC4 was also reported by the MAGMA gene-based tests in the NP group, where 5 out of 16 genes were involved in transcriptional regulation (*SRSF10*, *TSHZ2*, *HDAC4*, *EXOSC5*, *PNRC2*). Their importance is reflected in the abundance of microRNA and transcription factor targets in the gene-set enrichment analysis results, which were not replicated in the CP group.

Potential importance of findings for clinical practice and future application

Given that depression is a highly heterogeneous disorder with a multifactorial background, where different symptomatic manifestations are underlain by highly divergent neurobiological processes and disturbances and are subject to the interacting effects of distal and proximal stressors, both a deeper understanding of its potential subtypes and the identification of biomarkers or fingerprints are urgently warranted. Furthermore, depression associated with somatic disorders or comorbidities is considered a clinically distinct entity. Such somatic comorbidities may have a bidirectional relationship with depression, may present as a chronic stressor increasing risk, and may complicate the treatment of both conditions. Therefore, understanding the relationship between CP and depression, as well as the role of personality in this relationship, may bring us closer to characterizing subtypes, providing important insight into risk prediction, biomarker signatures for screening, and potentially novel treatment targets. Based on the accumulating evidence supported by our findings, CP-associated depression should also be considered a distinct clinical entity in clinical studies on depression treatment. As the next steps, responses to existing treatments in distinct subgroups of depression, including the one associated with CP, should be separately investigated, paving the way towards a much-needed precision approach in psychiatry.

Limitations

Our results must be viewed and interpreted in the context of a number of limitations which may also influence the generalizability of our findings.

First, there is evidence that there are genetic differences between CP subtypes [6, 73], however, we chose not to take these

differences into account in order to include as many participants as possible in our analyses, increasing statistical power. Future studies should take these questions into consideration.

Second, our analyses were performed using the UK Biobank database. Although the UK Biobank is one of the most extensive available population health resource databases, it is also subject to several biases that must be kept in mind when generalizing our findings. The UK Biobank cohort is ethnically predominantly White European (approx. 94%), aged during initial recruitment 40–69 years, participants are healthier, wealthier, and better educated compared to the general population. In addition to the above limitations related to representativeness, there may also be significant selection bias related to recruitment methods and low enrollment rates in the originally invited participants. Finally, the majority of UK Biobank data are self-reported and cross-sectional, including our CP-related variables and both measures for current depressive symptoms. The above-mentioned limitations of the sample should be carefully considered when interpreting our findings and generalizing them to larger populations [74–77].

Third, in addition to the above potential biases related to the characteristics of the UK Biobank sample, we excluded individuals with several mental disorders, including bipolar and psychotic disorders, as well as those with mental retardation and substance use disorders, which should also be considered when applying our findings in the general population. Further studies addressing subject diversity and generalizability are required [78].

CONCLUSION

In summary, our findings support that the proposed depression subtype associated with CP has distinct genetic characteristics. Furthermore, our results demonstrated that neuroticism has a prominent role in modulating the genetic risk of current depression independently of CP status, which is less pronounced when considering more robust phenotypes such as lifetime depression. We identified novel genetic risk polymorphisms in association with current depression, including multiple variants involved in synaptic plasticity and transcriptional regulation. Our findings emphasise that the genetic background of complex and heterogeneous phenotypes such as depression is challenging to unravel without considering the context of personality traits and stress factors, given their intricate interrelationships. A deeper comprehension of their interactions and the genetic architecture of depression subtypes may facilitate the development of innovative therapeutic approaches and lead us towards personalised medicine. Further research is needed to replicate our findings within population samples of different ancestries and to investigate the potential differences across CP subtypes.

DATA AVAILABILITY

UKB data is available for further research upon application to the data owners: UK Biobank, application number 1602.

REFERENCES

1. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. *Arch Intern Med*. 2003;163:2433–45.
2. Humo M, Lu H, Yalcin I. The molecular neurobiology of chronic pain-induced depression. *Cell Tissue Res*. 2019;377:21–43.
3. Farrell SF, Kho P-F, Lundberg M, Campos AI, Renteria ME, de Zoete RMJ, et al. A shared genetic signature for common chronic pain conditions and its impact on biopsychosocial traits. *J Pain*. 2023;24:369–86.
4. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. *Pain*. 2019;160:1361–73.
5. McIntosh AM, Hall LS, Zeng Y, Adams MJ, Gibson J, Wigmore E, et al. 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.* 2016;13:e1002090.
6. Meng W, Adams MJ, Reel P, Rajendrakumar A, Huang Y, Deary IJ, et al. Genetic correlations between pain phenotypes and depression and neuroticism. *Eur J Hum Genet.* 2020;28:358–66.
 7. Johnston KJA, Adams MJ, Nicholl BJ, Ward J, Strawbridge RJ, Ferguson A, et al. Genome-wide association study of multisite chronic pain in UK biobank. *PLoS Genet.* 2019;15:e1008164.
 8. Campos AI, Ngo TT, Medland SE, Wray NR, Hickie IB, Byrne EM, et al. Genetic risk for chronic pain is associated with lower antidepressant effectiveness: converging evidence for a depression subtype. *Aust N Z J Psychiatry.* 2022;56:1177–86.
 9. Shen X, Howard DM, Adams MJ, Hill WD, Clarke T-K, Deary IJ, et al. A phenome-wide association and mendelian randomisation study of polygenic risk for depression in UK biobank. *Nat Commun.* 2020;11:2301.
 10. Hullam G, Antal P, Petschner P, Gonda X, Bagdy G, Deakin B, et al. The UKB envirome of depression: from interactions to synergistic effects. *Sci Rep.* 2019;9:9723.
 11. Galvez-Sánchez CM, Montoro Aguilar CI. Migraine and neuroticism: a scoping review. *Behav Sci Basel Switz.* 2022;12:30.
 12. Galvez-Sánchez CM, Montoro CI, Duschek S, Del Paso GAR. Pain catastrophizing mediates the negative influence of pain and trait-anxiety on health-related quality of life in fibromyalgia. *Qual Life Res.* 2020;29:1871–81.
 13. Bond MR. The relation of pain to the Eysenck personality inventory, Cornell medical index and Whiteley index of hypochondriasis. *Br J Psychiatry J Ment Sci.* 1971;119:671–8.
 14. Goubert L, Crombez G, Van Damme S. The role of neuroticism, pain catastrophizing and pain-related fear in vigilance to pain: a structural equations approach. *Pain.* 2004;107:234–41.
 15. Ligthart L, Boomsma DI. Causes of comorbidity: pleiotropy or causality? Shared genetic and environmental influences on migraine and neuroticism. *Twin Res Hum Genet.* 2012;15:158–65.
 16. Eysenck HJ, Eysenck SBG. Manual of the Eysenck personality questionnaire. London: Hodder and Stoughton; 1975.
 17. John OP, Donahue EM, Kentle RL. Big five inventory. *J Pers Soc Psychol.* 1991. <https://doi.org/10.1037/107550-000>.
 18. Kotov R, Gamez W, Schmidt F, Watson D. Linking 'big' personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychol Bull.* 2010;136:768–821.
 19. Kendler KS, Gatz M, Gardner CO, Pedersen NL. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry.* 2006;63:1113–20.
 20. Ormel J, Jeronimus BF, Kotov R, Riese H, Bos EH, Hankin B, et al. Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clin Psychol Rev.* 2013;33:686–97.
 21. Jeronimus BF, Kotov R, Riese H, Ormel J. Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: a meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. *Psychol Med.* 2016;46:2883–906.
 22. Williams AL, Craske MG, Mineka S, Zinbarg RE. Neuroticism and the longitudinal trajectories of anxiety and depressive symptoms in older adolescents. *J Abnorm Psychol.* 2021;130:126–40.
 23. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet.* 2018;50:920–7.
 24. Tian R, Ge T, Kweon H, Rocha DB, Lam M, Liu JZ, et al. Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression. *Nat Commun.* 2024;15:1755.
 25. Smith DJ, Nicholl BJ, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS ONE.* 2013;8:e75362.
 26. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–9.
 27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606–13.
 28. Eszlari N, Brunicscs B, Millinghoff A, Hullam G, Petschner P, Gonda X, et al. Biology of Perseverative negative thinking: the role of timing and folate intake. *Nutrients.* 2021;13:4396.
 29. Privé F, Albiñana C, Arbel J, Pasaniuc B, Vilhjálmsson BJ. Inferring disease architecture and predictive ability with LDpred2-auto. *Am J Hum Genet.* 2023;110:2042–55.
 30. Purcell S, Chang C. PLINK 2.0. Available from: www.cog-genomics.org/plink/2.0/. Accessed Jan. 8, 2024.
 31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
 32. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet.* 2019;51:277–84.
 33. Zhang Q, Privé F, Vilhjálmsson B, Speed D. Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat Commun.* 2021;12:4192.
 34. Watanabe K, Umičević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun.* 2019;10:3222.
 35. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8:1826.
 36. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015;11:e1004219.
 37. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinforma Oxf Engl.* 2011;27:1739–40.
 38. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The gene ontology knowledgebase in 2023. *Genetics.* 2023;224:iyad031.
 39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–29.
 40. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics.* 2021;36:5424–31.
 41. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291–5.
 42. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika.* 1991;78:691–2.
 43. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2021. Available from: <https://www.r-project.org/>. Accessed Dec. 16, 2024.
 44. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4:1686.
 45. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics.* 2018;34:2781–7.
 46. Juliusdottir T, Stefánsson A. topR: create custom plots for viewing genetic association results. 2024. Available from: <https://CRAN.R-project.org/package=topR>. Accessed Dec. 16, 2024.
 47. Juliusdottir T. topR: an R package for viewing and annotating genetic association results. *BMC Bioinformatics.* 2023;24:268.
 48. Wickham H. ggplot2: elegant graphics for data analysis. Springer, New York, 2016.
 49. Nakazawa M. fmsb: functions for medical statistics book with some demographic data. 2024. Available from: <https://CRAN.R-project.org/package=fmsb>. Accessed Dec. 16, 2024.
 50. Nakazawa M. R practices of medical and health data analysis using R. Japan: Pearson Education; 2007.
 51. Vosberg DE, Leyton M, Flores C. The Netrin-1/DCC guidance system: dopamine pathway maturation and psychiatric disorders emerging in adolescence. *Mol Psychiatry.* 2020;25:297–307.
 52. Als TD, Kurki MI, Grove J, Voloudakis G, Therrien K, Tasanko E, et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat Med.* 2023;29:1832–44.
 53. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirihi M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.* 2019;22:343–52.
 54. Olofsson IA, Kristjánsson RP, Callesen I, Davidsson O, Winsvold B, Hjalgrim H, et al. Genome-wide association study reveals a locus in ADARB2 for complete freedom from headache in Danish blood donors. *Commun Biol.* 2024;7:646.
 55. Lahtinen A, Puttonen S, Vanttola P, Viitasalo K, Sulkava S, Pervjakova N, et al. A distinctive DNA methylation pattern in insufficient sleep. *Sci Rep.* 2019;9:1193.
 56. Killen AC, Barber M, Paulin JJW, Ranscht B, Parnavelas JG, Andrews WD. Protective role of Cadherin 13 in interneuron development. *Brain Struct Funct.* 2017;222:3567–85.
 57. Hawi Z, Tong J, Dark C, Yates H, Johnson B, Bellgrove MA. The role of cadherin genes in five major psychiatric disorders: a literature update. *Am J Med Genet B Neuropsychiatr Genet.* 2018;177:168–80.
 58. Howard DM, Adams MJ, Shirihi M, Clarke T-K, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun.* 2018;9:1470.
 59. Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, et al. Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat Genet.* 2013;45:912–7.
 60. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms,

- and neuroticism identified through genome-wide analyses. *Nat Genet.* 2016;48: 624–33.
61. Chen W-J, Zhu B-L, Qian J-J, Zhao J, Zhang F, Jiang B, et al. Hippocampal SorCS2 overexpression represses chronic stress-induced depressive-like behaviors by promoting the BDNF-TrkB system. *Pharmacol Biochem Behav.* 2024;242:173820.
 62. Corti F, Ristori E, Rivera-Molina F, Toomre D, Zhang J, Mihailovic J, et al. Syndecan-2 selectively regulates VEGF-induced vascular permeability. *Nat Cardiovasc Res.* 2022;1:518–28.
 63. Ashley-Koch AE, Garrett ME, Gibson J, Liu Y, Dennis MF, Kimbrel NA, et al. Genome-wide association study of posttraumatic stress disorder in a cohort of Iraq-Afghanistan era veterans. *J Affect Disord.* 2015;184:225–34.
 64. Gonda X, Krause S, Erdelyi-Hamza B, Sutori S, Gal Z, Eszlari N, et al. GWAS implicates SDC2 gene encoding Syndecan 2, previously associated with PTSD, insomnia, and autism, in suicidal ideation. *Neurosci Appl.* 2023;2:103339.
 65. Outcalt SD, Kroenke K, Krebs EE, Chumbler NR, Wu J, Yu Z, et al. Chronic pain and comorbid mental health conditions: independent associations of posttraumatic stress disorder and depression with pain, disability, and quality of life. *J Behav Med.* 2015;38:535–43.
 66. Sterling M, Smeets R, Keijzers G, Warren J, Kenardy J. Physiotherapist-delivered stress inoculation training integrated with exercise versus physiotherapy exercise alone for acute whiplash-associated disorder (StressModex): a randomised controlled trial of a combined psychological/physical intervention. *Br J Sports Med.* 2019;53:1240–7.
 67. Coleman JRI, Peyrot WJ, Purves KL, Davis KAS, Rayner C, Choi SW, et al. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK biobank. *Mol Psychiatry.* 2020;25:1430–46.
 68. Ermiş Akyuz E, Bell SM. The diverse role of CUB and sushi multiple domains 1 (CSMD1) in human diseases. *Genes.* 2022;13:2332.
 69. Stewart A, Maity B, Wunsch AM, Meng F, Wu Q, Wemmie JA, et al. Regulator of G-protein signaling 6 (RGS6) promotes anxiety and depression by attenuating serotonin-mediated activation of the 5-HT(1A) receptor-adenylyl cyclase axis. *FASEB J.* 2014;28:1735–44.
 70. Milazzo G, Mercatelli D, Di Muzio G, Triboli L, De Rosa P, Perini G, et al. Histone deacetylases (HDACs): evolution, specificity, role in transcriptional complexes, and pharmacological actionability. *Genes.* 2020;11:556.
 71. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell.* 2019;179:1469–82.e11.
 72. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet Lond Engl.* 2013;381:1371–9.
 73. Nicholl BI, Mackay D, Cullen B, Martin DJ, Ul-Haq Z, Mair FS, et al. Chronic multisite pain in major depression and bipolar disorder: cross-sectional study of 149,611 participants in UK biobank. *BMC Psychiatry.* 2014;14:350.
 74. Allen NE, Lacey B, Lawlor DA, Pell JP, Gallacher J, Smeeth L, et al. Prospective study design and data analysis in UK biobank. *Sci Transl Med.* 2024;16:eadf4428.
 75. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol.* 2017;186: 1026–34.
 76. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault J-B, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav.* 2023;7:1216–27.
 77. van Alten S, Domingue BW, Faul J, Galama T, Marees AT, Reweighting UK. Biobank corrects for pervasive selection bias due to volunteering. *Int J Epidemiol.* 2024;53:dyae054.
 78. Cruz LA, Cooke Bailey JN, Crawford DC. Importance of diversity in precision medicine: generalizability of genetic associations across ancestry groups toward better identification of disease susceptibility variants. *Annu Rev Biomed Data Sci.* 2023;6:339–56.

ACKNOWLEDGEMENTS

This project was supported by the National Research, Development and Innovation Office, Hungary (2019-2.1.7-ERA-NET-2020-00005), under the frame of ERA PerMed (ERAPERMED2019-108); by the Hungarian National Research, Development, and Innovation Office (grant number: K 143391); by the Hungarian Brain Research Program (Grant: 2017-1.2.1-NKP-2017-00002; NAP2022-I-4/2022); KTIA_13_NAPA-II/14; KTIA_NAP_13-1-2013- 0001; KTIA_NAP_13-2- 2015-0001; NAP2022-I-4/2022; by the Ministry of Innovation and Technology of Hungary, Development and Innovation Fund, under TKP2021-EGA-25; Sandor Krause was supported by the ÚNKP-23-3-I-SE-73 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. Dora Torok is supported by EKÖP-2024-68.

AUTHOR CONTRIBUTIONS

SK participated in the conceptualisation and design of the study, data analysis, interpretation of results, writing the original draft, and reviewing and editing further versions of the manuscript. DT participated in conceptualization and design of the study, developing the methodology, interpreting the results, writing the original draft, and reviewing and editing further versions of the manuscript. GB participated in funding acquisition and reviewing and editing the manuscript. GJ participated in funding acquisition, conceptualisation, and reviewing and editing the manuscript. XG participated in the conceptualization and design of the study, supervising the work, interpreting the data and results, writing the original draft, reviewing and editing further versions of the manuscript.

FUNDING

Open access funding provided by Semmelweis University.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-025-03331-5>.

Correspondence and requests for materials should be addressed to Xenia Gonda.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025