Impact of gene-by-trauma interaction in MDD-related multimorbidity clusters

Sarah Bonk
Nora Eszlari
Kevin Kirchner
Andras Gezsi
Linda Garvert
Mikko Koukkanen
Isaac Cano
Hans J. Grabe
Peter Antal
Gabriella Juhasz
Sandra Van der Auwera
auweras@uni-greifswald.de

Research Article

Keywords: major depression, childhood stress, dopamine beta-hydroxylase, dopamine receptor D2, tryptophan hydroxylase 1, methylenetetrahydrofolate reductase

Posted Date: October 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3456781/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Journal of Affective Disorders on May 1st, 2024. See the published version at https://doi.org/10.1016/j.jad.2024.05.126.
Abstract

Background:

Major depressive disorder (MDD) is considerably heterogeneous in terms of comorbidities, which may hamper the disentanglement of its biological mechanism. In a previous study, we classified the lifetime trajectories of MDD-related multimorbidities into seven distinct clusters, each characterized by unique genetic and environmental risk-factor profiles. The current objective was to investigate genome-wide gene-by-environment (G×E) interactions with childhood trauma burden, within the context of these clusters.

Methods:

We analyzed 76,856 participants and 3,875,386 single-nucleotide polymorphisms (SNPs) of the UK Biobank database. Childhood trauma burden was assessed using the Childhood Trauma Screener (CTS). For each cluster, Plink 2.0 was used to calculate SNP×CTS interaction effects on the participants’ cluster membership probabilities. We especially focused on the effects of 31 candidate genes and associated SNPs selected from previous G×E studies for childhood maltreatment’s association with depression.

Results:

At SNP-level, only the high-multimorbidity Cluster 6 revealed a genome-wide significant SNP rs145772219. At gene-level, LDLRAD4 was genome-wide significant for the low-multimorbidity Cluster 1 and C6orf89 and TAAR2 for the high-multimorbidity Cluster 7. Regarding candidate SNPs for G×E interactions, individual SNP results could be replicated for specific clusters. The candidate genes DRD2 (Cluster 1), and DBH and MTHFR (both Cluster 5), and TPH1 (Cluster 6) survived multiple testing correction.

Limitations:

CTS is a short retrospective self-reported measurement. Clusters could be influenced by genetics of individual disorders.

Conclusions:

The first G×E GWAS for MDD-related multimorbidity trajectories successfully replicated findings from previous G×E studies related to depression, and revealed risk clusters for the contribution of childhood trauma.

Highlights

· First G×E GWAS with childhood trauma for MDD-related multimorbidity trajectories

· Most G×E findings for high MDD-multimorbidity Clusters 5-6
· Significant correlation between cluster probability and CTS (high vs low risk clusters)

· Genetic findings correspond to the burden of childhood trauma

· Differential effects of candidate genes in different comorbidity clusters

**Introduction**

Major depressive disorder (MDD) is the most common psychiatric disorder in Western countries and caused by a combination of genetic predisposition and various non-genetic factors (Flint, 2023). The heritability of MDD, based on additive genetic variation, has been estimated to lie around 25% (Flint, 2023). Additionally, numerous factors have been identified that increase the risk of developing depressive episodes throughout lifetime. These are factors primarily associated with an unfavorable lifestyle, such as smoking, alcohol consumption, physical inactivity, and poor diet, or associated with stressful living conditions, including childhood adversity, stressful life events, and unemployment (Sarris et al., 2020). Unfortunately, MDD is highly heterogeneous on the clinical level, with variations in individual symptoms and the development of multiple co-occurring disorders. This often leads to negative effects caused by polypharmacy and a decreased quality of life for patients and makes it even more challenging to identify biological causes of MDD (Rush et al., 2006; Fried and Nesse, 2015; Fratelli et al., 2020).

Explaining this complex relationship between predisposing genetic factors and external risk factors on a biological level is difficult. Essentially, this means that the impact of a specific genetic variation on an individual’s phenotype depends on the exposure to external risk factors (known as gene-by-environment (G×E) interaction) (Karg and Sen, 2012). Childhood adversities have been identified as a significant risk factor for the development of various psychiatric disorders, including MDD (Mandelli and Serretti, 2013). In genetic studies, previous attempts to uncover such genetic interaction effects for childhood adversity have mostly concentrated on specific candidate genes involved in neurotransmitter systems (Mandelli and Serretti, 2013; Culverhouse et al., 2018; Border et al., 2019; Li et al., 2020a), or combined results of large genome-wide association studies (GWAS) into polygenic scores (PGS) (Peyrot et al., 2018; Coleman et al., 2020). In a genome-wide interaction approach using data from the Psychiatric Genomics Consortium (van der Auwera et al., 2018) the moderation effect of prominent candidate genes and childhood adversities on MDD could not be supported. To date, interaction analyses have not yielded new biological models, and the underlying mechanisms still remain elusive. Several factors could account for this, including limited sample sizes, as well as strong heterogeneity within the diagnosis of MDD regarding individual symptoms and co-occurring multimorbidities, which may share both genetic and non-genetic risk factors (Flint, 2023).

In the TRAJECTOME (*Temporal disease map-based stratification of depression-related multimorbidities*, Juhasz et al., 2023) project, we generated temporal disease trajectories of 86 pre-selected multimorbidities related to MDD in over 1.2 million individuals with the aim to identify clusters that represent temporal courses of MDD-related multimorbidity burden over an individual’s lifetime. These seven clusters are characterized by a unique genetic and non-genetic risk factor profile (Juhasz et al.,...
showing a clear differentiation between high- and low-risk clusters in terms of MDD and MDD-related disorders. The significant contributions of both genetic predisposition and external risk factors to these clusters suggest a possible G×E interaction. In our current analysis, we aim to identify specific clusters where the interaction between genetic factors and childhood trauma burden may influence the assignment to the MDD-related multimorbidity clusters. We expect that the contribution of G×E will vary across clusters, not only in terms of strength of the G×E signal but also in the specific genes and biological mechanisms involved.

**Methods**

**UK Biobank study population and measurements**

For our analysis, we used data from the UK Biobank (UKB) under the application number 1602, which contains comprehensive medical, phenotypic and genotypic information from participants recruited based on the NHS patient registers of people aged 40–69 years (Smith et al., 2013). All participants gave written informed consent, and ethical approval was obtained from the National Research Ethics Service Committee North West–Haydock (Nagel et al., 2018). All procedures were in accordance with the Declaration of Helsinki.

**Cluster membership outcome variable**

Within the TRAJECTOME project, seven clusters were identified that reflect an individual’s MDD-related multimorbidity burden throughout lifetime (Juhasz et al., 2023).

To construct these clusters, we utilized temporal disease information from a total of 502,504 participants from the UK Biobank, along with 687,005 participants from two other general population cohorts (THL cohorts from Finland (Sund, 2012) and CHSS from catalonia (Farré et al., 2016)). The disease trajectories were categorized into four different cumulative time intervals (aged [0–20], [0–40], [0–60], and [0–70]). We included 86 diseases with a minimum prevalence of 1% and significant relevance to MDD within any of the time intervals to compute MDD-related multimorbidity scores for each time interval and each participant, respectively. Participant clustering of these scores identified seven distinct clusters reflecting temporal courses of the MDD-related multimorbidity burden throughout the lifespan (Fig. 1). The clusters correspond to specific clinical subtypes with high and low MDD-related multimorbidity burden (see Table 3 for the five diseases with the most increased and decreased prevalence within each cluster). The clusters are characterized as follows: Participants with a high probability for Clusters 1–4 have a decreased depression prevalence and a lower MDD-related multimorbidity burden whereas participants with a higher probability for Clusters 5–7 show an increased prevalence for depression and a higher overall MDD-related multimorbidity burden. For more details on the clustering procedure and clinical characteristics of the clusters see Geszi et al. (Juhasz et al., 2023). To account for incomplete and uncertain participant trajectories in the analyses, we excluded individuals who were under 60 years of age and had a maximum posterior probability of less than 0.25 for membership in any cluster (Juhasz et al., 2023). Our outcome variables were the posterior probabilities
(log-odds) of the individuals’ cluster membership for each of the identified seven MDD-related multimorbidity clusters (Juhasz et al., 2023).

**Childhood trauma assessment**

Childhood trauma was assessed within the UKB mental health online follow-up (Davis et al., 2020) using the Childhood Trauma Screener (CTS) (Walker et al., 1999; Glaesmer et al., 2013), the short form of the more comprehensive Childhood Trauma Questionnaire (CTQ) (Glaesmer et al., 2013; Klinger-König et al., 2022). The five-item CTS includes one question for each dimension of childhood trauma: emotional abuse (UKB data field 20487), physical abuse (data field 20488), sexual abuse (data field 20490), emotional neglect (data field 20489), and physical neglect (data field 20491). Answers are given on a five-point Likert-scale from “never true” to “very often true”. A summary score was calculated that reflects the burden of childhood trauma experienced, with a range of 0–20. The score was only calculated for participants that answered all of the five questions.

**Genetic data and GWAS**

The genomic quality control has been detailed elsewhere (Juhasz et al., 2023).

In brief, we selected White British participants (data field 22006) without putative sex chromosome aneuploidy (data field 22019) (Juhasz et al., 2023). We used UKB v3 genetic data of genotyped and imputed variants, positioned according to the GRCh37/hg19 genome assembly (Juhasz et al., 2023). Variants’ quality control included filtering out multiallelic variants and variants with a minor allele frequency (MAF) < 0.01, keeping only common biallelic single-nucleotide polymorphisms (SNPs) (Juhasz et al., 2023). For imputed SNPs, both info and certainty parameters had to be at least 0.9. Further, we excluded participants and SNPs according to missingness rate (iteratively, with cut-off points of 0.1, 0.05, and 0.01), as well as SNPs according to Hardy-Weinberg equilibrium violation ($p < 1 \times 10^{-5}$) (Juhasz et al., 2023). Before further steps of participant filtering, a linkage disequilibrium (LD) pruning was applied on SNPs with an $r^2$ of 0.2 (Juhasz et al., 2023). The maximal set of unrelated individuals (data-field 22020) was selected (Bycroft et al., 2018), and a king-cutoff “–kin 0.044” filtering step was done (Juhasz et al., 2023). Furthermore, a sex check and a heterozygosity outlier detection (Eszlari et al., 2019) was applied on participants (Juhasz et al., 2023). X chromosome and the pseudoautosomal regions of the two sex chromosomes were included in the analyses, in addition to autosomal chromosomes. Males’ haploid genotypes are coded as if they are homozygous.

For the GWAS analyses, we selected participants who passed the above GWAS quality control steps, had non-withdrawn consent in February 2022, and had non-missing data according to sex, age, all CTS questions, cluster memberships, and genotyping array. To control for population stratification, principal component analysis was run in the final set of participants ($N = 76,856$) and with the SNP subset after LD pruning detailed above (similarly to Juhasz et al., 2023).

Plink 2.0 (Chang et al., 2015) (https://www.cog-genomics.org/plink/2.0/) was used to run linear regression models and assess the moderation effect of each remaining SNP in interaction with the CTS
sum score on the seven MDD-related cluster probabilities as outcome. Additional predictors of each model were age, sex, the first ten genetic principal components, genotyping array, and main effects of the SNP and CTS score. Age was taken into the model as a non-linear variable using cubic splines with a knot at age 60 (calculated by R package *splines*, and function *bs*) (Juhasz et al., 2023). Continuous predictor and outcome variables were all standardized in the analyses, as in (Juhasz et al., 2023). Genetic and phenotypic data were available for 76,856 participants. Effect sizes and *p*-values of the interaction term were evaluated, and entered into post-GWAS analyses.

**Post-GWAS analysis**

To assess the impact of SNP-interaction results on biological processes, several post-GWAS analyses were applied that extract information on significant loci, genes and pathways.

The G×E GWAS summary results for all seven MDD-related clusters were first processed with FUMA (https://fuma.ctglab.nl/, Watanabe et al., 2017) to identify lead SNPs and significant loci. The maximum *p*-value of lead SNPs was set to 5×10^{-6}, r² ≥ 0.6 was set as threshold for independent significant SNPs and the maximum distance between LD blocks of independent significant SNPs was set to 250kb. Furthermore, MAGMA (Leeuw et al., 2015) gene-level analysis was performed to identify putative significant genes and gene sets. We defined the SNP set of each gene with an extended +/− 10 kb downstream or upstream of the gene. We used the UK Biobank Genome European panel data to evaluate the LD information between SNPs. Statistical comparison of results between the clusters were done in R (https://cran.r-project.org/).

**Candidate SNPs and gene selection**

To compare our results with previously identified and discussed candidate variants and genes for G×E interaction in the light of childhood maltreatment and depression, we selected two recent papers (Border et al., 2019; Li et al., 2020a) and included the following 31 genes: SLC6A4, BDNF, COMT, HTR2A, TPH1, TPH2, MAOA, DRD2, DRD3, DRD4, MTHFR, APOE, CLOCK, SLC6A3, ACE, ABCB1, DTNB1, DBH, CRHR1, FKBP5, CREB1, NTRK2, OXTR, IL1b, IL6, IL11, CRP, TNF, TNFRSF1A (TNFR1), TNFRSF1B (TNFR2), and GABRG2. These genes capture common biological mechanisms for G×E interaction in depression such as the serotonergic system, hypothalamic-pituitary-adrenal (HPA) axis or immune-related processes (Remes et al., 2021). To analyze these genes, a window of +/−1kb around the GRCh37/hg19 position was used, to mainly focus on SNPs within the coding region (compared to +/−10kb in the MAGMA analysis above). Within these genes, all SNPs available in the UKB genetic data and surviving the filtering process above were selected also including their putative candidate variants if available. These candidate variants are the most commonly investigated SNPs in the literature as listed by Border and Li (Border et al., 2019; Li et al., 2020a) (see Supplementary Table S4). In our study, we consider only candidate SNPs that are biallelic, which is the case in 20 out of the 31 candidate genes.

On the SNP-level *p*-values and effect estimates from the GWAS were analyzed, on the gene-level MAGMA-based *p*-values were investigated for each MDD-related cluster.
Candidate SNPs that showed a nominally significant G×E interaction for any cluster membership in our present results were investigated using the Oxford Brain Imaging Genetics Server (BIG40) to look up significant SNP associations with brain imaging phenotypes in the UK Biobank cohort itself (https://open.win.ox.ac.uk/ukbiobank/big40/pheweb33k/, Elliott et al., 2018; Smith et al., 2021).

Results

We analyzed data from 76,856 UKB participants with a mean age of 58.36 years, ranging from 40 to 72 years, of which 54% were female (characteristic of the sample see Supplementary Table S1). As these are data from the general population, the intensity of childhood trauma was relatively low with a mean CTS score of 1.67 and a range between 0 and 20 (see Supplementary Fig. S1).

On the phenotypic level, the correlations between posterior log-odds cluster membership probabilities and the CTS score (see Supplementary Table S2) reflected a pattern of high and low risk clusters that has already been observed for the clinical characteristics of the clusters (Juhasz et al., 2023). In detail, Clusters 1–4 showed a strongly significant but weak negative correlation with the CTS score, meaning that participants with a high probability to belong to the clusters tend to have a lower burden of childhood trauma. In contrast, the correlations with Clusters 5 and 6 were strongly significant but positive, reflecting a higher childhood trauma burden for subjects belonging to these clusters. Cluster 7, instead, revealed no significant correlation with the CTS score (see Supplementary Table S2).

Interaction GWAS for the seven MDD-related multimorbidity clusters

We performed G×E GWAS analyses in 76,856 UKB participants with complete phenotypic and genetic data using 3,875,386 genotyped or imputed SNPs (Table 1). Analyses for all seven clusters only revealed one genome-wide significant locus ($p < 5 \times 10^{-8}$) for Cluster 6 on the X chromosome with lead SNP rs145772219 (Supplementary Fig. S2 - S4). Thus, we set the level of suggestive significant SNPs to $p < 5 \times 10^{-6}$ to interpret their impact in interaction with the CTS score. Applying this new threshold, suggestive significant SNPs were found for all clusters with 291 distinct SNPs spanning 57 risk loci on all chromosomes except chromosomes 12, 22, and pseudoautosomal regions of sex chromosomes (Table 1, Fig. 2, Supplementary Table S3, Supplementary Fig. S2). The strongest genetic signal was found for Cluster 5 (132 significant SNPs, 19 significant loci) and Cluster 6 (103 significant SNPs, 28 significant loci) (Table 1, Fig. 2, Supplementary Table S3, Supplementary Fig. S2). In Clusters 3, 4, and 7 less than seven SNPs reached significance. The MAGMA gene-based analyses revealed three genome-wide significant genes after Benjamini-Hochberg correction for multiple testing using 18,937 protein-coding genes; one in Cluster 1 ($LDLRAD4 \ p = 9.8 \times 10^{-7}$) and two in Cluster 7 ($C6orf89 \ p = 2.1 \times 10^{-6}, \ TAAR2 \ p = 1.8 \times 10^{-6}$). Lowering the significance threshold to $p < 1 \times 10^{-4}$, additional genes emerged (see the genes in parentheses in Table 1, Supplementary Fig. S5). Even at this level no significant gene was found for Cluster 3. On the level of gene sets, MAGMA analysis identified two gene sets significant after multiple
testing correction using 15,486 gene-sets (GO\_cc: go\_mon1\_ccz1\_complex for Cluster 2 and Curated\_gene\_sets: borlak\_liver\_cancer\_egf\_up for Cluster 6). On the G\times E GWAS-level, the cluster-wise correlations of beta values again reflected the pattern of high- and low-risk clusters (Supplementary Fig. S6), which was also observed in the correlation between CTS score and the cluster membership probabilities (see Supplementary Table S2).

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
<th>Cluster 6</th>
<th>Cluster 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>N per SNP</td>
<td>75,990–76,856</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda</td>
<td>0.9374</td>
<td>0.9576</td>
<td>0.9898</td>
<td>0.9751</td>
<td>1.1019</td>
<td>1.2136</td>
<td>0.9927</td>
</tr>
<tr>
<td>SNPs</td>
<td>3,875,386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sign. SNPs*</td>
<td>46</td>
<td>34</td>
<td>1</td>
<td>2</td>
<td>132</td>
<td>103</td>
<td>6</td>
</tr>
<tr>
<td>sign. loci</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>28</td>
<td>5</td>
</tr>
<tr>
<td>sign. genes**</td>
<td>LDLRAD4, (MYOF)</td>
<td>-</td>
<td>(LDLRAD4)</td>
<td>-</td>
<td>(UGT2A3)</td>
<td>-</td>
<td>(MMP16, VGLL2, NRG3, DENND3)</td>
</tr>
<tr>
<td></td>
<td>- (UGT2A3)</td>
<td>-</td>
<td>-</td>
<td>- (MMP16, VGLL2, NRG3, DENND3)</td>
<td>- (FOXA3, DAW1, B3GNT7, TREX2, SBK2, FEM1A, HAUS7)</td>
<td>C6orf89, TAAR2, (PNLDC1, IL17A)</td>
<td></td>
</tr>
<tr>
<td>sign. candidate genes***</td>
<td>DRD2, (IL1B, CRP, OXTR)</td>
<td>- (NTRK2, DRD2, OXTR)</td>
<td>- (DRD4, IL6, MTHFR)</td>
<td>- (IL6, MTHFR)</td>
<td>DBH, MTHFR, (CREB1, APOE, DRD2, MAOA)</td>
<td>TPH1, (IL6, OXTR)</td>
<td>- (IL6, DRD2)</td>
</tr>
</tbody>
</table>

Lambda: genetic inflation factor; *significance of SNPs refers to a suggestive significance level \( p < 5 \times 10^{-6} \); **results from MAGMA analyses, significance based on Benjamini-Hochberg correction, using 18,937 protein-coding genes. In brackets additional genes with suggested significance \( p < 1 \times 10^{-4} \), ***results from MAGMA analyses, significance based on Benjamini-Hochberg correction, using 31 genes. In brackets additional genes with suggested significance \( p < 0.05 \).

### Evaluation of previously identified candidate genes and SNPs

To compare our findings with previous results and hypotheses on the interaction between SNPs and childhood trauma on depression, we selected 31 putative candidate genes from the literature (see methods section Candidate SNPs and gene selection). For these genes, 3,788 SNPs were available in the UKB genetic data and passed the filter procedure (Supplementary Table S4), with the highest number
situuated within the \textit{CLOCK} gene (N = 490) and the lowest within \textit{APOE, DRD4} and \textit{TPH1} (N = 8) before pruning. After gene-wise pruning ($r^2 = 0.8$), 937 independent SNPs were available, with the highest number in the \textit{NTRK2} gene (N = 110) and the lowest within \textit{DRD4} and \textit{TPH1} (N = 6). As candidate SNP studies often apply a lenient significance threshold of $p < 0.05$, we also applied this threshold in a first screening approach to not miss potential informative findings.

For each gene in the pruned list we found at least one nominal significant SNP in at least one of the seven cluster GWASes (Supplementary Table S5). The highest number of independent significant SNPs was found for the genes \textit{NTRK2, ABCB1}, and \textit{DBH}. Moreover, ten genes revealed a nominal significant effect for at least one SNP in all seven clusters (\textit{MTHFR, OXTR, SLC6A3, GABRG2, DTNBP1, FKBP5, IL6, NTRK2, DBH, DRD2}) whereas four genes (\textit{IL1B} in Cluster 1, \textit{ACE} in Cluster 5, \textit{APOE} in Cluster 5, \textit{CRHR1} in Cluster 6) showed significant SNPs in only one cluster. For three genes we observed individual SNPs that were significant in at least five of the seven clusters (rs74853132 of \textit{FKBP5}, rs4936274 of \textit{DRD2}, 4 SNPs of \textit{MTHFR}; Supplementary Tables S6 - S8, Supplementary Fig. S7 - S8). However, directions of effect were different among the clusters depending on the cluster correlation with high or low childhood trauma burden (Supplementary Fig. S9). Since our analysis was limited to common biallelic SNPs, some genetic candidate variants such as indels and variable number tandem repeats could not be included in this evaluation.

Looking at the clusters separately, the genes with the three highest numbers of independent significant SNPs were \textit{TPH2, FKBP5, DRD2, MTHFR} and \textit{OXTR} for Clusters 1, \textit{NTRK2, OXTR} and \textit{TPH2} for Cluster 2, \textit{NTRK2, FKBP5, MTHFR} and \textit{SLC6A4} for Cluster 3, \textit{NTRK2, MTHFR} and \textit{DBH} for Cluster 4, \textit{DBH, ABCB1} and \textit{MTHFR} for Cluster 5, \textit{NTRK2, DBH} and \textit{TNF} for Cluster 6 and \textit{DRD3, ABCB1} and \textit{FKBP5} for Cluster 7. Clusters 6 (N = 85) and 2 (N = 72) came up with the highest number of independent significant SNPs within candidate genes, whereas Cluster 4 revealed the lowest number (N = 51) (Supplementary Table S5).

Out of the 18 candidate SNPs that were available, either directly or as proxy ($r^2 > 0.8$), five exhibited a nominal significant association in at least one of the cluster GWASes (\textit{ABCB1} rs2235048, \textit{NTRK2} rs1147194, \textit{TNF} rs1041981, \textit{IL6} rs367801961 and rs60056354, \textit{TPH1} rs1800532). None of these SNPs remains significant after Benjamini-Hochberg $p$-value correction. Looking these SNPs up in the BIG40, all five SNPs revealed a significant association towards brain imaging phenotypes in the UKB cohort supporting their impact on psychiatric disorders (see Table 2 for specific parameters and effect alleles).
Table 2

Candidate SNPs that were nominally significant in an interaction term with CTS score for any cluster membership as outcome + our new suggested candidate SNP for DRD2. In addition, significant associations of these SNPs with brain imaging phenotypes in Oxford Brain Imaging Genetics Server are listed. Cluster memberships are the posterior probability (log-odds) of cluster membership to each of the identified seven MDD-related multimorbidity clusters. SNP: single-nucleotide polymorphism; CTS: Childhood Trauma Screener; MDD: major depressive disorder.

<table>
<thead>
<tr>
<th>Cluster membership as outcome</th>
<th>Significant interaction term in an additive genetic model (effect allele vs. reference allele)</th>
<th>Gene</th>
<th>Beta of the interaction term</th>
<th>P-value (nominal) of the interaction term</th>
<th>Oxford Brain Imaging Genetics Server results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>rs4936274 x CTS score (A vs. G)</td>
<td>DRD2</td>
<td>-0.020775</td>
<td>0.00237824</td>
<td>G allele: increased cortical thickness in the left middle temporal area</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
<td>-0.020644</td>
<td>0.0030806</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td></td>
<td></td>
<td>-0.020480</td>
<td>0.0060594</td>
<td></td>
</tr>
<tr>
<td>Cluster 4</td>
<td></td>
<td></td>
<td>-0.018032</td>
<td>0.0138477</td>
<td></td>
</tr>
<tr>
<td>Cluster 5</td>
<td></td>
<td></td>
<td>0.015057</td>
<td>0.0367074</td>
<td></td>
</tr>
<tr>
<td>Cluster 7</td>
<td></td>
<td></td>
<td>0.021306</td>
<td>0.0076216</td>
<td></td>
</tr>
<tr>
<td>Cluster 5</td>
<td>rs2235048 x CTS score (A vs. G)</td>
<td>ABCB1</td>
<td>-0.010444</td>
<td>0.0173438</td>
<td>A allele: decreased intensity of right choroid plexus</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>rs1147194 x CTS score (T vs. C)</td>
<td>NTRK2</td>
<td>0.013339</td>
<td>0.0162281</td>
<td>C allele: lower volume parameters of left nucleus accumbens and ventral striatum (suppl. Fig. S10)</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>rs1041981 x CTS score (A vs. C)</td>
<td>TNF</td>
<td>-0.009796</td>
<td>0.0252289</td>
<td>A allele: right hemisphere reduced volumes of several thalamic nuclei and reduced area of insula; left hemisphere reduced volume and cortical area of the superior temporal cortex (suppl. Fig. S11)</td>
</tr>
<tr>
<td>Cluster 6</td>
<td></td>
<td></td>
<td>0.013509</td>
<td>0.0093099</td>
<td></td>
</tr>
<tr>
<td>Cluster 4</td>
<td>rs367801961 x CTS score (A vs. G)</td>
<td>IL6</td>
<td>-0.012857</td>
<td>0.0079681</td>
<td>A allele: right hemisphere, increased volume of CA3 head in hippocampus, increased cortical thickness in orbital lateral sulcus and in the superior part of precentral sulcus; left</td>
</tr>
</tbody>
</table>


On the level of the 31 candidate genes, we evaluated the MAGMA gene based results for the seven Cluster GWASes (Supplementary Table S9). Applying a Benjamini-Hochberg p-value correction in each cluster separately, three clusters came up with significant genes: DRD2 in Cluster 1, DBH and MTHFR in Cluster 5 and TPH1 in Cluster 6 and 13 genes reached nominal significance in any of the clusters (Supplementary Table S9). A significance heatmap (Fig. 3) revealed that Clusters 1 and 2 as well as Clusters 4 and 7 were most similar regarding their significance pattern across the candidate genes, which was not consistent with the pattern observed on the genome-wide level or the association pattern with the CTS score (Supplementary Fig. S6). The strongest genetic signal throughout all clusters was found for the genes TPH1, OXTR, DRD2, MTHFR, and IL6. On cluster-level, the highest number of nominal significant candidate genes was found for Clusters 5 (N = 6) and 1 (N = 4) with the overlapping gene DRD2. The two genes ACE and CRHR1 revealed barely a significant SNP and were not significant in any of the clusters.

**Discussion**

In the current analysis we investigated the interaction effect between SNP-based genetic variation and childhood trauma on seven MDD-related multimorbidity clusters. These clusters reflect the temporal courses of MDD-related multimorbidity burden throughout life and could initially be associated with a unique clinical, genetic and modifiable risk-factor profile (Juhasz et al., 2023). Here, we extend this direct genetic characterization by moderation effects investigating childhood trauma, one of the strongest risk
factors for MDD and other psychiatric diseases in general. In seven G×E GWAS including 76,856 UK Biobank participants, we replicated the pattern of high- and low-multimorbidity clusters concerning childhood trauma burden which has already been found on the level of genetic and non-genetic factors (Juhasz et al., 2023). The strongest genetic findings in the G×E analyses could be observed for the high CTS burden Clusters 5 and 6 with more than 19 independent loci exceeding suggestive significance. These clusters were also associated with a high MDD-related multimorbidity burden. On the genome level, the correlation pattern showed a strong similarity between Clusters 1–4 which were all associated with a lower CTS burden. In contrast, Clusters 5–7 seemed to exhibit three individual but contrary genetic profiles that contribute to the high multimorbidity load in individuals with high CTS burden. Thus, childhood trauma might promote the development of certain diseases by altering biological pathways and metabolic processes that might be traced back to the identified genes (Table 1, Table 3). From a clinical point of view, especially the Clusters 5, 6 and 7 are of interest, as they give insights into the genetic risk-profiles for the development of certain diseases depending on their childhood trauma burden. As our clusters are based on depression-related multimorbidity trajectories, we selected the five diseases with the most increased and decreased prevalences within each cluster (Table 3) to draw biological connections towards the suggestive genes identified in our analysis.
Table 3
Biological connections toward the suggested genes, which are genome-wide significant in the GxE analysis, and the five diseases with the most increased and decreased prevalence within each cluster plus effect direction for depression.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>suggested genes in GxE results</th>
<th>top five positive and negative associated disorders*</th>
<th>link between genes and disorders in the literature</th>
</tr>
</thead>
</table>
| 1       | **LDLRAD4, MYOF**              | negative: depression, schizophrenia, reaction to severe stress, tonsillitis, allergic rhinitis, dorsalgia | **LDLRAD4**: schizophrenia (Kikuchi et al., 2003);  
**MYOF**: none |
| 2       | **LDLRAD4**                    | negative: depression, schizophrenia, reaction to severe stress, tonsillitis, allergic rhinitis, migraine | schizophrenia (Kikuchi et al., 2003) |
| 3       | -                              | negative: depression, tonsillitis, allergic rhinitis, migraine, asthma, pain (female genital organs)  
positive: hypertension, cerebral infarction, cerebrovascular disease, acute kidney failure, chronic kidney disease | - |
| 4       | **UGT2A3**                     | negative: depression, tonsillitis, allergic rhinitis, migraine, asthma, pain (female genital organs)  
positive: hypothyroidism, lipidemia, hypertension, benign prostatic hyperplasia | None |
| 5       | **MMP16, VGLL2, NRG3, DENND3** | negative: -  
positive: depression, schizophrenia, allergic rhinitis, intervertebral disc disorder, dorsalgia, pain (female genital organs) | **MMP16**: pain (and pain associated disorders like migraine) (Wotton et al., 2022);  
**NRG3**: depression (Paterson et al., 2017), schizophrenia (Avramopoulos, 2018; Li et al., 2020b);  
**VGLL2, DENND3**: none |
| 6       | **FOXA3, DAW1, B3GNT7, TREX2, SBK2** | negative: allergic rhinitis, asthma, migraine | **FOXA3**: asthma (Park et al., 2009);  
**DAW1**: allergy (Waage et al., 2018); |

*based on weighted Cox proportional hazards regression model (Juhasz et al., 2023)
<table>
<thead>
<tr>
<th>Cluster</th>
<th>suggested genes in G×E results</th>
<th>top five positive and negative associated disorders*</th>
<th>link between genes and disorders in the literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td><em>FEM1A, HAUS7</em></td>
<td>positive: depression, reaction to severe stress, somatoform disorders, nasopharyngitis, bronchitis, soft tissue disorders</td>
<td><em>B3GNT7, TREX2, SBK2, FEM1A, HAUS7: none</em></td>
</tr>
</tbody>
</table>
|         |                                | negative: alcohol related disorders, nicotine dependence, hypertension, acute kidney failure, chronic kidney disease | *C6orf89: maybe involved in allergic rhinitis and asthma (Liu et al., 2016; Xu et al., 2017) (regulation of human airway epithelium);*  
|         |                                | positive: depression, allergic rhinitis, asthma, tonsillitis, migraine, dermatitis | *TAAR2: involved in dopamine regulation (Efimova et al., 2022) which is associated with affective disorders (Grace, 2016);*  
|         |                                |                                                      | *IL17A: kidney disease and transplantation (Kim et al., 2012; Romanowski et al., 2015) and asthma (Holster et al., 2018); inconsistent results for allergic rhinitis (Wang et al., 2012; Fatahi et al., 2016) and atopic dermatitis (Narbutt et al., 2015; Klonowska et al., 2022);*  
|         |                                |                                                      | *PNLDC1: none*                                  |

*based on weighted Cox proportional hazards regression model (Juhasz et al., 2023)

The genetic risk profile for Cluster 5 includes the genes *MMP16, NRG3, VGLL2, DENND3*. Previous associations for *MMP16* with nociception can be linked to intervertebral disc disorder, dorsalgia and pain (female genital organs) (Wotton et al., 2022), three high prevalence diseases in Cluster 5 (Juhasz et al., 2023). The already known psychiatric risk gene *NRG3*, in turn, might have a potential developmental role in schizophrenia (Avramopoulos, 2018; Li et al., 2020b), bipolar disorder and major depression (Paterson et al., 2017). Pain and psychiatric disorders were associated with childhood trauma by overlapping brain mechanism: It has been shown that regions of the brain involved in the pain matrix (such as the anterior cingulate cortex, the amygdala, or the hippocampus) are altered after experiences of childhood abuse and trauma (Teicher et al., 2003; Brown et al., 2018). However, neither *VGLL2* nor *DENND3* showed any associations with the top five diseases with increased prevalence for Cluster 5, while *DENND3* is associated with the volumes of different brain regions (e.g. cortical surface area (Shadrin et al., 2021) cortical thickness (Shadrin et al., 2021; van der Meer et al., 2021), cerebellum (Zhao et al., 2019; Chambers et al., 2022)) and Alzheimer’s disease (Chung et al., 2022).

For Cluster 6, our G×E analysis identified *FOXA3, DAW1, B3GNT7, TREX2, SBK2, FEM1A, HAUS7* as suggestive genes. *FOXA3* might be involved in the regulation of allergic airway diseases and asthma (Park et al., 2009). As asthma is a potential risk factor for migraine and vice versa, also possible connections between *FOXA3* and migraine are conceivable (Wang et al., 2020). Asthma as well as
migraine are usually triggered by stress. *DAW1* is associated with allergy (Waage et al., 2018), which aligns with allergic rhinitis being one of the high prevalence diseases in Cluster 6. For the remaining genes no associations were reported. However, *B3GNT7* may play a role in the formation of neurophils and perineuronal nets in the adult brain (Takeda-Uchimura et al., 2022).

The genetic risk profile for Cluster 7 includes the genes *C6orf89, TAAR2, PNLDC1* and *IL17A*. The gene *C6orf89* encodes the bombesin receptor-activated protein (BRAP) which might be involved in the stress response of lung epithelia (Liu et al., 2016; Xu et al., 2017) and can be linked to allergic rhinitis and asthma. In mice, the BRAP homologous protein may have a protective effect on the behavioral response to stress via regulating dendritic spine formation and synaptic plasticity in the hippocampus (Yao et al., 2023). As a consequence, it was concluded that chronic stress might cause damage to hippocampus function (Yao et al., 2023). *TAAR2*, in turn, expresses a protein which is involved in dopamine regulation and adult neurogenesis (Efimova et al., 2022). As dysregulations in the dopamine system are related to affective disorders like MDD (Grace, 2016), *TAAR2* represents a promising target for treating neuropsychiatric disorders (Efimova et al., 2022). *IL17A* is, among others, associated with chronic kidney disease (Kim et al., 2012) and kidney transplantation (Romanowski et al., 2015) as well as asthma (Holster et al., 2018). Regarding allergic rhinitis (Wang et al., 2012; Fatahi et al., 2016) and atopic dermatitis (Narbutt et al., 2015; Klonowska et al., 2022) studies revealed inconsistent results. Especially, asthma and allergic rhinitis appear to have different genetic risk profiles for *IL17A* (Resende et al., 2017). A link towards the role of *PNLDC1* instead could not be found.

With respect to the candidate genes, we did not confirm their impact in G×E analyses on our MDD-related multimorbidity clusters as a whole. However, some genes suggest a biological connection towards specific individual clusters, underscoring biological heterogeneity stemming from distinct temporal patterns of MDD-related multimorbidity.

From the 18 available candidate SNPs, 5 (located in *ABCB1, NTRK2, TNF, IL6, TPH1*) showed a nominally significant interaction in at least one cluster (Supplementary Table S4). From these 5 SNPs, no one had a significant interaction in more than three clusters. Hence, our observation that they are not significant in all clusters is in line with Border et al. (Border et al., 2019) and Li et al. (Li et al., 2020a) as they were also not able to confirm the impact of these SNP – Childhood Trauma interactions on MDD in general. However, the effect direction of the candidate SNP for *NTRK2* in Cluster 4 was in line with previous findings (Juhasz et al., 2011; van der Auwera et al., 2018; Li et al., 2020a).

*IL6* delivered the strongest genetic signal with significant results in both candidate SNP and gene-wise MAGMA analyses in several clusters (Cluster 4 and Cluster 6 on SNP-level, Cluster 3, 4, 6 and Cluster 7 on gene-level). The GWAS catalog lists an association between *IL6* and asthma which is among the top five diseases in Clusters 3, 4, 6, 7. Conversely, asthma is not among the top five diseases in Cluster 1, 2, and 5, where the interaction effect with *IL6* is not significant. The strong genetic signal of *IL6* might be due to the close interrelation between inflammation, stress and depression (Ting et al., 2020). There, it
was shown that psychosocial stress acts as a trigger for depression development by initiating changes in HPA-axis and immune/inflammatory system (Ting et al., 2020).

Also, results for *TPH1*, which catalyzes the first and rate limiting step in the biosynthesis of serotonin, revealed significant interactions on SNP (Cluster 1, Cluster 2 and Cluster 6) as well as on gene level (Cluster 6) in clusters that tend to have a low as well as high CTS burden. As *TPH1* is associated with a broad range of psychiatric conditions (Shnayder et al., 2022) this might explain the link toward these clusters. Different effect directions for *TPH1* on SNP level are reflected by the decreased prevalence rates of psychiatric conditions in the low-risk Clusters 1 and 2 in contrast to the increased prevalence rates for the high-risk Cluster 6.

On a gene-based level several candidates might be of special interest as they either survive multiple testing within a cluster or show a strong association pattern towards several clusters.

The gene *DRD2* (which encodes a D2 subtype of the dopamine receptor) showed a significant interaction effect in several clusters (Clusters 1, 2, 5, and 7; Fig. 3, Supplementary Table S9) and was the only gene that could be confirmed on the gene-level by Border et al. (2019). *DRD2* is associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Shioda, 2017), which is among the top five associated disorders in all of these clusters. Interestingly, similar to Border et al. (2019), we found no significant interaction effect of the candidate SNP (rs1800497). This SNP was initially assigned to the *DRD2* gene, but was later found to be located in the *ANKK1* gene (https://www.ncbi.nlm.nih.gov/gene/1813). Instead, we found a significant interaction for rs4936274 in Clusters 1–5, and 7 (Supplementary Table S6) which is not in LD with the historical candidate SNP.

Hence, we propose that rs4936274 might be an alternative candidate SNP for the *DRD2* x CTQ interaction on depression, as we observe a clear reversal of the effect direction when comparing clusters with high vs low MDD burden.

Furthermore, the *DBH* and *MTHFR* genes revealed significant interaction results with CTS in the gene-based analyses in Cluster 5 (Supplementary Table S9), although the proposed candidate SNP was not available in our analyses. In addition, *MTHFR* had a nominally significant interaction in Clusters 3 and 4 (Supplementary Table S9) and four independent *MTHFR* SNPs ($r^2 < 0.8$; Supplementary Fig. S7) showed a significant interaction in five clusters which might also be due to the broad association of *MTHFR* with neurological and psychiatric disorders (Liew and Gupta, 2015; Zhang et al., 2022).

The oxytocin receptor gene (*OXTR*) was significant in Clusters 1, 2 and 6 (Supplementary Table S9). Our dataset does not contain the candidate SNP, but in other studies that SNP had no significant interaction effect (Tollenaar et al., 2017; van der Auwera et al., 2018). Instead, we found two independent SNPs (rs60345038, rs62243375) that showed a nominally significant interaction in Cluster 1, 2 and 6. One of them (rs60345038) also had a significant association with social cognitive performance in individuals with schizophrenia (Davis et al., 2014) and could also be a novel risk variant that is possibly linked to and associated with familial type 2 diabetes (Amin et al., 2023).
Our study has several limitations: The CTS, being a retrospective self-reported measurement, is likely to be influenced by recall bias (Baldwin et al., 2019). Further, the cluster probabilities are based on MDD-related multimorbidities, which makes it difficult to compare our results with previous G×E findings for depression, although we present results for the first analysis on G×E interaction for temporal MDD-multimorbidity clusters. In addition, the cluster assignment strongly depends on the reliability of the data from the healthcare system where missassignments can lead to wrong results for the G×E analysis. Due to strong correlations among our parameters (MDD, diseases, environmental factors and CTS), interpreting the correlations between GWASes in terms of causes and mechanisms may prove challenging. Results could be biased by the direct GWAS results for the clusters.

To conclude, our results underscore that some of the former candidate SNPs exert their effects on MDD-related multimorbidity patterns depending on the level of childhood trauma. Such multimorbidity patterns may explain previously inconclusive results on G×E analyses. This genetically based susceptibility for early trauma effects may root in differences in brain phenotypes. Each SNP can have its distributed impacts across numerous brain regions (van der Meer et al., 2020), and these brain-wide differences may establish inter-individual differences in sensitivity to environmental (e.g., traumatic) factors, and thus in multimorbidity patterns, as suggested by our present results. Furthermore, our findings indicate that the moderation of SNP effects by CTS may exert a more prominent influence on the high multimorbidity clusters compared to the low multimorbidity clusters. However, future studies are to reveal the exact etiopathological mechanisms from G×E SNPs through brain phenotypes towards multimorbidity patterns. Regarding the role of candidate SNPs, we conclude that rs4936274 is likely a better candidate SNP for the $DRD2 \times CTQ$ interaction on depression than the former candidate SNP and should be tested in further analysis. Investigating MDD-related multimorbidity patterns may be a promising approach in G×E analyses.

Declarations

Conflicts of Interest:

HJG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care.

Acknowledgment:

This work uses data provided by patients and collected by the NHS as part of their care and support. Copyright © (2019), NHS England. Re-used with the permission of the UK Biobank (Application Number 1602). All rights reserved.

Funding:
This study was supported by 2019-2.1.7-ERA-NET-2020-00005 under the frame of ERA PerMed (ERAPERMED2019-108); by the Hungarian Brain Research Program 3.0 (NAP2022-I-4/2022); by TKP2021-EGA-25, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-EGA funding scheme; by the Hungarian National Research, Development, and Innovation Office, with grants K 143391 and PD 134449. N. E. was supported by the ÚNKP-22-4-II-SE-1 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund; and is supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

This study was supported by the Federal Ministry of Education and Research (BMBF, grant no. 01KU2004) under the frame of ERA PerMed (ERAPERMED2019-108).

**Author Contribution:**

Conceptualization: SV, NE, SB, KK; formal analysis: NE, SB, AG, SV; writing original draft: SB, KK, NE, SV; writing review: LG, HJG; resources: IC, MK, HJG, PA, GJ; supervision: SV, GJ, PA; all authors critically reviewed the manuscript and approved the final version.

**References**


Figures
**Figure 1**

**Trajectories of the weighted direct MDD-related multimorbidity score over time using 7 clusters.** Each box corresponds to a cluster of participants in which the trajectories of the scores are similar. Each colored line corresponds to the trajectory of a single individual. The red lines show the mean trajectory in a cluster. The x-axis corresponds to the discrete cumulative time intervals (1: 0-20, 2: 0-40, 3: 0-60, and 4: 0-70), and the y-axis shows the value of the weighted direct MDD-related multimorbidity score. (This figure was created using the same code as for the figures in (Juhasz et al., 2023), but with a smaller subset with available data on childhood.)
Figure 2

Graphical representation of the significant loci (p < 5x10^{-6}) for the MDD-related clusters G×E GWAS analysis in UKB. Each colour represents a Cluster. The coloured knots represent the genomic location of the significant loci.
Figure 3

Heatmap of gene-based results derived by MAGMA based on candidate genes. -log10 $p$-values for each gene and each cluster are displayed.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTables.xlsx