

# Genome-by-Trauma Exposure Interaction Effects in Depression

Chuong, T.M.; Adams, M.J.; Kwong, A.S.F.; Haley, C.S.; Amador, C; McIntosh, A.M

## Summary

### Background

Self-reported trauma exposure has consistently been found to be a risk factor for Major Depressive Disorder (MDD) and several studies have reported interactions with genetic liability. To date, most studies have examined interaction effects with trauma exposure using genome-wide variants (single nucleotide polymorphisms SNPs) or polygenic scores, both typically capturing less than 3% of phenotypic risk variance. We sought to re-examine genome-by-trauma interaction effects using genetic measures utilising all available genotyped data and thus, maximising accounted variance.

### Methods

Measures of self-reported depression, neuroticism and trauma exposure for 148 129 participants with whole genome SNP data are available from the UK Biobank study. Here, we used a mixed-model statistical approach utilising genetic, trauma exposure and genome-by-trauma exposure interaction similarity matrices to explore sources of variation in depression and neuroticism.

### Findings

Our approach estimated the heritability of MDD to be approximately 0.160 [SE 0.016]. Subtypes of self-reported trauma exposure (catastrophic, adult, childhood and full trauma) accounted for a significant proportion of the variance of each trait, ranging from 0.056 [SE 0.013] to 0.176 [SE 0.025]. The proportion of MDD risk variance accounted for by significant genome-by-trauma interaction ranged from 0.074 [SE 0.006] to 0.201 [SE 0.009]. Results from sex-specific analyses found genome-by-trauma interaction variance estimates approximately 5-fold greater for MDD in males than in females.

### Interpretation

This is the first study to utilise an approach combining all genome-wide SNP data when exploring genome-by-trauma interaction effects in MDD and present evidence that interaction effects are influential in depression manifestation. This effect accounts for greater trait variance within males which points to potential differences in depression aetiology between the sexes. The methodology utilised in this study can be extrapolated to other environmental factors to identify modifiable risk environments and at-risk groups to target with interventions.

## Research In Context

### Evidence before this study

We searched PubMed up to January 30<sup>th</sup> 2022, with the following terms: (“gene environment interaction” OR “gene environment” OR “genome wide by environment” OR “GWEIS” OR “polygenic environment” OR (“gene” AND “environment”)) AND (“polygenic risk score” OR “polygenic score” OR “genomic relationship matrix” OR “GRM”) AND (“trauma” OR “environmental adversity” OR “stressful life events”) AND (“depression” OR “major depressive disorder” OR “MDD” OR “depressive symptoms”). Date or language restrictions were not applied. We further reviewed the reference lists of identified articles. This search was supplemented by reviewing related articles identified by Google Scholar. We identified 12 relevant articles. Studies to date have not explored genome-by-environment interaction effects in depression using genomic similarity matrices, however, these effects have been explored using individual single nucleotide polymorphisms (SNPs) from genome-wide studies and polygenic scores (PGSs). Some findings suggest genome-by-environment interaction effects increase risk of depression. However, replication attempts have produced either inconsistent or null findings. Taken together, it is evident that findings have failed to provide consistent evidence of substantial interaction effects. These findings may be a result of limited statistical power in analyses due to genome-wide variants and PGSs failing to capture the polygenic nature of depression with sufficient precision.

### Added value of this study

This study is the first to explore genome-by-trauma interaction effects on MDD through the estimation of variance components using relationship matrices. Genomic relationship matrices (GRMs) utilise all available genotyped variants, thus, capturing a greater proportion of the trait variance and potentially providing greater power to detect genetic effects in comparison to PGSs. Additional relationship matrices capturing trauma exposure, and genome-by-trauma exposure similarity are computed and included into mixed linear models. We found evidence for substantial genome-by-trauma (including subtypes of trauma) exposure interaction effects on depression manifestation. Estimated genome-by-trauma interaction effects were larger in males than in females.

### Implications of all the available evidence

Our findings are the first to show substantial genome-by-trauma effects on depression using whole genome methods. These findings highlight that the role of trauma exposure on depression manifestation may be non-additive and different between sexes. Exploring these effects in depth may yield important insight into various mechanisms, which may explain prevalence differences observed between males and females. Future work can build upon the framework we propose to explore genome-by-trauma interaction effects and the underlying molecular sites and mechanisms which are involved in depression manifestation.

## Introduction

Depression is a highly prevalent psychiatric disorder with a lifetime risk of ~16%<sup>1, 2</sup> and is a leading cause of disability worldwide.<sup>3</sup> Twin studies have produced heritability estimates ranging between 30-40%<sup>4</sup> suggesting both genetic and environmental factors are influential. In fact, genome-wide association studies (GWAS) have uncovered many single nucleotide polymorphisms (SNPs) associated with depression, however, SNP-based heritability estimates of 5-10%<sup>5, 6</sup> are much lower than estimates obtained from twin studies which has highlighted the issue of missing heritability. Explanations for missing heritability include inflation of twin-based heritability estimates attributable to shared environmental effects as well as gene-environment interplay.<sup>7-10</sup>

Some environmental factors such as self-reported trauma exposure have been found to play a role in depression; with case-control studies suggesting individuals diagnosed with Major Depressive Disorder (MDD) report higher levels of trauma exposure.<sup>11-13</sup> In turn, trauma exposure in childhood has been consistently associated with adverse outcomes including increased risk of MDD in adolescence and adulthood.<sup>14, 15</sup> These findings strongly suggest a causal role for trauma in depression's aetiology. Moreover, research has shown self-reported trauma exposure and MDD to be genetically correlated, suggesting shared genetic risk factors for both<sup>16</sup>, or potentially more complex interplay effects of genetics and trauma exposure on depression manifestation.

One key form of interplay that has been explored is the interaction effect of genetics and trauma exposure on depression manifestation. Gene-by-environment interaction effects refer to the differential effects of an environmental exposure on traits in individuals with differing genotypes. This can be conceptualised as an individual's genetic sensitivity to certain environments, which may result in an exacerbated risk of a disorder.<sup>17, 18</sup> Minimal evidence of interaction effects has been yielded from studies exploring SNP-by-trauma effects.<sup>16, 19-22</sup>

Moreover, research utilising polygenic scores (PGSs); genetic measures that can be calculated for each individual by identifying, weighting, and summing genotyped risk variants found to be associated with depression<sup>23, 24</sup>, have yielded inconsistent findings. Whilst some studies have highlighted sex differences and found significant interaction effects associated with MDD outcomes<sup>18, 25-27</sup>, some replication attempts reported null findings<sup>28, 29</sup> and follow up meta-analyses have suggested that reported findings were likely to be false positives.<sup>30</sup> An explanation for the inconsistent findings may lie in the predictive accuracy and validity of PGSs.<sup>10</sup> PGSs build upon the information provided by GWAS, which still have limited statistical power for detecting trait associated genetic variants and their effect sizes.<sup>24, 31, 32</sup> Power limitations of GWAS, and thus PGSs, are greater for traits that also have a substantial environmental component, such as depression, as opposed to traits with higher genetic aetiology.<sup>33</sup> This is reflected in the fact that current PGSs capture less than 3%<sup>5, 6</sup> of the phenotypic variance proposed for depression.

One way to circumvent the issues associated with PGSs, is to make use of genetic measures that are capturing greater variance. Genomic relationship matrices (GRMs) have been able to do this by utilising all genotyped SNPs.<sup>34</sup> Including matrices representing genetic, environmental and interaction effect similarity within a sample, in mixed linear models, can provide estimates of the genetic, environmental, and genome-by-environment interaction components of trait variance.<sup>35</sup>

Here, we estimated the contribution of trauma exposure and its interaction with genetic variation to depression as well as neuroticism. We chose to explore neuroticism as this trait has been shown to have a greater genetic component<sup>36, 37</sup> and has a strong phenotypic link with MDD suggesting the exploration of neuroticism to be useful in understanding the genetic aetiology of both of these

traits.<sup>38, 39</sup> As the existing literature has highlighted the role of trauma exposure in sex differences observed in MDD, we also explored these effects in males and females separately.<sup>40</sup> Here we show that utilising the entirety of genotyped genetic variants can improve statistical power in the exploration of genome-by-trauma interaction effects. More importantly, our findings suggest that genome-by-trauma interaction effects may play a much larger role in depression manifestation than previously thought.

## Methods

### PARTICIPANTS

We used data from the UK Biobank (UKB), a national study exploring genetic and environmental determinants of health using individuals recruited from 22 different centres across the United Kingdom.<sup>41, 42</sup> A follow up Mental Health Questionnaire (MHQ) was administered assessing common mental health disorders including trauma experience.<sup>43</sup> In this study we had access to individual level data from 148 129 participants who completed the MHQ with available genetic, trauma experience, depressive symptoms and/or neuroticism information.

The UKB study received ethical approval from the NHS National Research Ethics Service (reference: 11/NW/0382) and all participants provided written informed consent. This study has been approved by the UKB Access Committee (Project #4844).

### GENOTYPES

Two separate array chips with 95% SNP overlap, were used to genotype UKB participants; the Applied Biosystems™ UK BiLEVE Axiom™ Array by Affymetrix1 captures 807 411 SNPs and the Applied Biosystems™ UK Biobank Axiom™ Array captures 825 927 SNPs.<sup>44</sup> Quality control of genotyped SNPs consisted of the exclusion of SNPs with missingness > 2% and a Hardy-Weinberg Equilibrium test  $p < 10^{-6}$ . SNPs with minor allele frequency (MAF) < 0.05 and individuals with > 2% missing genotypes were excluded from analyses. A total of 414 584 common SNPs across the 22 autosomes were included in analyses.

### PHENOTYPES

All phenotypes were defined using retrospective self-reported responses to questions assessing help-seeking behaviour, depressive symptoms, and trauma experience. Field codes and more information on classifications are available in supplementary table 1.

#### Depression & Neuroticism Phenotypes

CIDI depression was defined using questions from the Composite International Diagnostic Interview Short Form (CIDI).<sup>45</sup> These were administered within the online follow-up Mental Health Questionnaire (MHQ), directly assessing the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)<sup>46</sup> symptomatic and functional criteria for MDD, and are associated with greater MDD specificity.<sup>47</sup>

Supplementary analyses were conducted using broad depression and neuroticism phenotypes. Broad depression was defined using self-reported help seeking behaviour for mental health difficulties. Case and control status was determined from the response to Touchscreen Questionnaire questions administered during initial recruitment; 'Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?' or 'Have you ever seen a psychiatrist

for nerves, anxiety, tension or depression? Individuals responding ‘Yes’ to either and ‘No’ to both questions were classified as cases and controls, respectively.

Neuroticism was measured using the 12-item Neuroticism scale of the revised short form Eysenck Personality Questionnaire (EPQ-R).<sup>48</sup> Response options were ‘Yes’, ‘No’, ‘Do not know’, ‘Prefer not to answer’. Participant summed scores of ‘Yes’ responses (ranging from 0-12) were used in downstream analyses.

### Trauma Phenotypes

Participants were administered a 16-item questionnaire relating to traumatic experiences as a part of the MHQ. Five items explored childhood trauma using the Childhood Trauma Screener<sup>49, 50</sup>; five items explored adult trauma using an equivalent screener developed by the UKB Mental Health steering group<sup>51</sup>; and the remaining six items explored catastrophic trauma using questions related to events that often trigger post-traumatic stress disorder.<sup>16</sup> More information on questions, distribution of responses can be found in supplementary figure 1 and table 2. **Table 1** shows sample sizes of individuals included in statistical analyses with complete trauma exposure, depression and neuroticism information.

**Table 1.** Sample sizes of UKB participants with complete trauma, depression and neuroticism information.

Trauma Measure	CIDI		Broad		Neuroticism
	Case	Control	Case	Control	
<b>FULL TRAUMA</b>	32 015	81 809	45 093	82 370	116 995
<b>CHILDHOOD</b>	33 532	85 405	47 583	85 894	121 670
<b>ADULT</b>	33 262	83 722	46 886	84 234	119 914
<b>CATASTROPHIC</b>	33 657	85 704	47 850	86 127	122 108

*Abbreviations.* UKB, UK Biobank; CIDI, Composite International Diagnostic Inventory diagnosis criteria of Major Depressive Disorder (MDD); Broad, diagnosis of MDD based on self-reported diagnosis of nerves, anxiety, tensions or depression from healthcare professional.

Individual regression analyses exploring associations between each trauma item and depression/neuroticism outcomes with age and sex as covariates (supplementary table 2), suggested most trauma items to be significantly associated with all of the three phenotypes. Hence, all items were utilised in the formation of trauma phenotypes.

In order to capture separate and independent dimensions of trauma, principal components (PCs) of complete responses to all trauma items (full trauma) as well as the three sub-categories of trauma items (childhood, adult, catastrophic trauma) were obtained. Trauma items were rescaled to have a mean of one and unit variance before PCs were extracted using the ‘prcomp’ function in R 4.0.2.<sup>52</sup> For more information on depression/neuroticism associations with trauma PCs and PC loadings see supplementary table 3-4.

### COVARIANCE MATRICES

Covariance matrices were used in the exploration of genetic, environmental and genome-by-environment components of variation. As the number of participants with available depression/neuroticism and trauma information is large (~150k), analysing the pairwise covariance matrices jointly for the whole cohort was computationally intractable. To work around these computational issues, participants were split into five different clusters based on the geographical location of their recruitment centres (north, midnorth, midsouth, southwest and southeast regions

in the UK) and all subsequent analyses were replicated across the five clusters; for more information on how clusters were formed and their demographic characteristics see supplementary Figure 2 and table 5-6.

Genomic Relationship Matrices (GRMs): represent the expected genetic similarity between individuals. We computed one GRM for all individuals in each of the geographical clusters using genetic variants that passed quality control (see section 'GENOTYPES') using GCTA v1.91.4beta.<sup>34</sup> Supplementary GRMs were computed using *only* unrelated individuals (individuals with genomic similarity values < 0.05).

Trauma Environmental Relationship Matrices (E): represents similarity between individuals based on their trauma eigenvectors (PCs). Separate Es were computed by calculating participant similarity based on their full trauma, childhood, adult and catastrophic trauma eigenvectors. Supplementary Es were computed using the first eigenvector (PC1) of full trauma, as this accounted for the greatest variance in our outcome phenotypes. To control for genetic covariance between trauma exposure and our phenotypes of interest (depression/neuroticism), supplementary Es were computed using trauma eigenvectors pre-corrected for the full GRM. We used OSCA (v0.45) default algorithm 1<sup>53</sup> to compute these variables, for more information on available algorithms see supplementary materials section A.

Gene-Environment Interaction Matrices (GxE): represent shared genome-by-trauma interaction effects. These were computed by multiplying GRM and ERMs using a cell-by-cell (Hadamard) product.<sup>35, 54, 55</sup>

## ANALYSES

### Genetic Correlations

Using the first eigenvector (PC1) for the subtypes of trauma, the SNP heritability of each trauma variable was explored. Heritability estimates of the trauma variables were obtained by fitting the trauma variables as the dependent variable and Gs as random effects within a mixed linear model framework (estimates of trauma PC1 variance attributable to G), see model 1 below. The genetic correlations between trauma PC1 and depression variables were explored using the moment-based method, Haseman-Elston regression analyses, where standard errors were calculated using a leave-one-individual-out jackknife technique. Age, sex, genotyping array and the first 15 principal components of the GRM were included as covariates.

### Variance Components Analyses

Variance components of depression/neuroticism were explored within mixed linear model frameworks. Four models were explored with varying levels of complexity:

1.  $y = X\beta + g + \varepsilon$
2.  $y = X\beta + e + \varepsilon$
3.  $y = X\beta + g + e + \varepsilon$
4.  $y = X\beta + g + e + g \times e + \varepsilon$

Where  $y$  is a  $n \times 1$  vector of observed depression/neuroticism phenotypes;  $\beta$  is a vector of fixed effects (which include age, sex, genotyping array and the first 15 principal components of the full sample GRM) and  $X$  is its design matrix;  $g$  is a  $n \times 1$  vector of SNP effects (representing additive genetic effects) with  $g \sim \text{Norm}(0, G\sigma_G^2)$ ;  $e$  is a  $n \times 1$  vector representing common environmental effects of childhood, adult, catastrophic or all trauma with  $e \sim \text{Norm}(0, E\sigma_E^2)$ ;  $g \times e$  is a  $n \times 1$

vector representing interactions between genetic and trauma effects with  $g \times e \sim \text{Norm}(0, G \circ E \sigma_{G \circ E}^2)$ ; and  $\varepsilon$  is a  $n \times 1$  vector of residual effects.

Estimates of variance attributable to the G, E and GxE components are obtained from analyses using CIDI and Broad depression as dependent variables and are converted to the liability scale within GCTA.<sup>34</sup> As UKB is a population-based study, we utilised the prevalence rates observed within the whole sample. Prevalence rates used were 0.28 and 0.35 for joint sex analyses; 0.35 and 0.43 for female analyses; 0.19 and 0.27 for male analyses for CIDI and broad depression, respectively.

Analyses were repeated using only unrelated individuals. All analyses were replicated across the five geographical cluster samples using GCTA v1.91.4beta<sup>34</sup> and results (estimates of variance components) were meta-analysed using R package `metafor`.

## Results

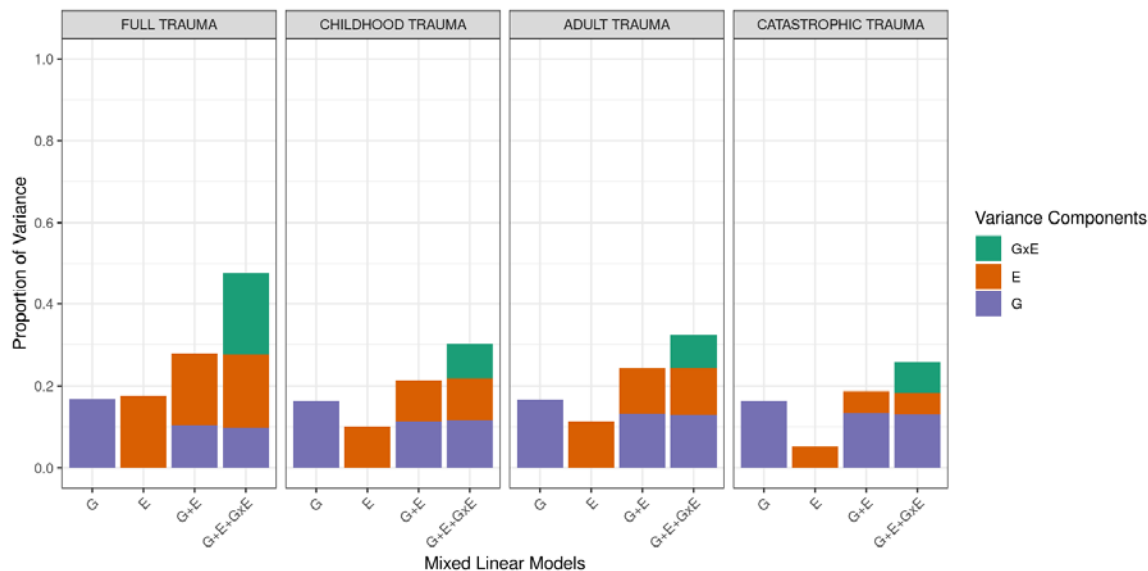
Initial analyses explored the SNP heritability of the first PC of full trauma and sub-categories of trauma. Trauma variables meta-analysed SNP heritability estimates were; full trauma 0.17 [SE 0.008], childhood trauma 0.15 [SE 0.008], adult trauma 0.063 [SE 0.008] and catastrophic trauma 0.11 [SE 0.008] (see supplementary tables 8). Similar results were obtained when using only unrelated individuals. All genetic correlations between the first PC of trauma variables and depression/neuroticism phenotypes are shown in Table 2. Genetic correlations between the first PC of trauma variables and broad depression/neuroticism phenotypes were modest; in contrast, we observed stronger genetic correlations between trauma variables and the CIDI depression phenotype. Results from each cluster can be found in supplementary table 9.

**Table 2.** Genetic Correlations Between Trauma and Depression/Neuroticism Phenotypes

Phenotype	Full Trauma		Childhood Trauma		Adult Trauma		Catastrophic Trauma	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
<b>CIDI</b>								
<b>Depression</b>	0.632	0.085	0.605	0.091	0.647	0.134	0.536	0.104
<b>Broad Depression</b>	0.390	0.082	0.337	0.090	0.358	0.132	0.310	0.098
<b>Neuroticism</b>	0.333	0.064	0.332	0.071	0.274	0.105	0.204	0.081

*Abbreviations:* SE, Standard Error; CIDI, Composite International Diagnostic Interview.

**Figure 1** shows the estimates for the proportion of CIDI depression variance explained by the different sources included in the mixed linear models (results are the meta-analysis of the five UKB sub-samples). All estimates for proportion of variance explained by all components were statistically significant. Log-likelihood ratio tests (LRTs) suggested that the inclusion of trauma (E) and genome-by-trauma (GxE) interaction components improve model fit. Full details of these analyses, including estimates, standard errors (SEs), LRT values as well as results using broad depression and neuroticism as the dependent variable can be found in supplementary materials (supplementary tables 10-13).



**Figure 1.** Proportion of CIDI Depression Variance Explained by Genetic, Environmental and Interaction Sources in UK Biobank. The proportion of variance explained is specified on the y-axis. The sources of variation: G, Genomic; E, Environmental; GxE, Gene-by-Environment Interactions and their corresponding models are specified on the x-axis. E represent trauma exposure. Facets show the different trauma exposures explored.

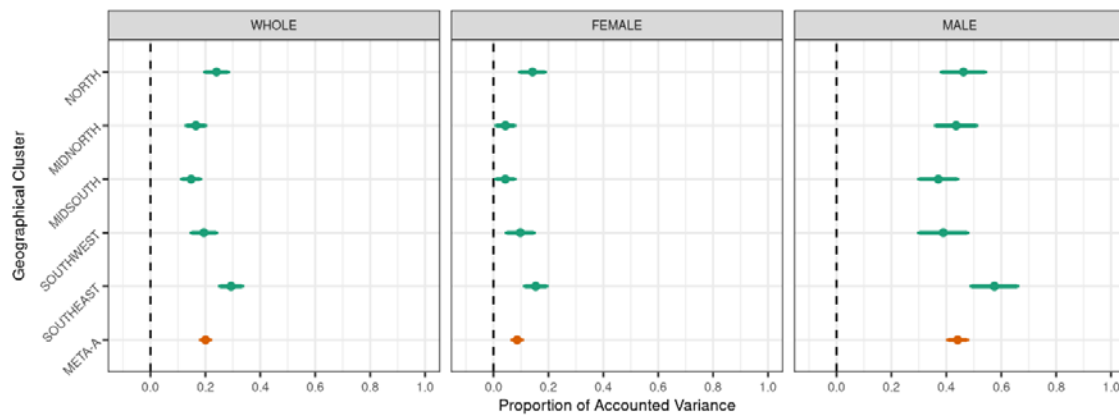
Heritability estimates (i.e. proportion of phenotypic variance accounted for by the G) of CIDI depression were stable across the different UKB sub-samples  $\sim 0.16$  [SE 0.016]. The meta-analysed estimates for proportion of variance attributable to the trauma relationship matrices (Es) were 0.18 [SE 0.025]=full trauma, 0.101 [SE 0.027]=childhood trauma, 0.113 [SE 0.03]=adult trauma, 0.05 [SE 0.013]=catastrophic trauma. The meta-analyses estimates for proportion of variance attributable to the interaction effect (GxE) were highest when exploring full trauma 0.201 [SE 0.009] and were 0.084 [SE 0.006], 0.081 [SE 0.005], 0.074 [SE 0.006] when respectively exploring childhood, adult and catastrophic trauma separately.

Similar results were obtained when using only unrelated individuals as well as when mixed linear models utilise Es computed from trauma eigenvectors pre-corrected for the full sample G (supplementary tables 14-15). In contrast, whilst model fit, compared to models excluding Es, are significantly improved, we observed smaller estimates and LRT values when mixed linear models utilised Es computed from only the first eigenvector (PC1) of full trauma items (supplementary table 1).

Significant, yet, smaller estimates of variance components were observed for broad depression and neuroticism (supplementary tables 10-13).

**Figure 2** shows the estimates for the proportion of CIDI depression variance explained by the interaction (GxE) included in the mixed linear models. Here, the interaction effect used Es capturing full trauma exposure. Results for each geographical cluster as well as within female/male only samples are presented. Full details of these analyses, including estimates, SEs, LRTs values as well as results using broad depression and neuroticism as the dependant variable can be found in supplementary materials (supplementary tables 10,17-18).





**Figure 2.** Proportion of CIDI Depression Variation explained by the Genome-by-Trauma Interactions. Forest plot x-axis shows proportion of CIDI Depression variance explained by the genome-by-trauma interaction effect with bars representing standard errors. Results for geographic clusters (green) and meta-analysed estimates (orange) are shown on the y-axis. Facets represent the analysis results using different samples; WHOLE, joint male and female sample; FEMALE, female only sample; MALE, male only sample.

Meta-analysed estimates for proportion of CIDI depression variance explained by the interaction effect across the clusters were statistically significant within the whole (joint female and male), female and male samples. Compared to the analyses of the full sample (joint male and female), the meta-analysed interaction variance was smaller when explored within the female sample and increased when explored within the male sample. Similar results were observed when using only unrelated individuals (supplementary tables 19-20).

## Discussion

This study sought to provide a comprehensive update to the literature on genome-by-trauma interaction effects within depression and depression related traits by exploring variance components as opposed to fixed effect estimates of genome-by-trauma interactions. We compute genetic (using all genotyped variants), trauma exposure and genome-by-trauma interaction effect similarity to explore trait variance attributable to these effects. By utilising all available genotyped data, the mixed linear models implemented have greater statistical power to identify phenotypic variation attributable to genetic and genome-by-trauma interaction effects. We explore general (full trauma) and specific (childhood, adult, catastrophic) measures of trauma.

Our heritability estimates of the trauma measures support findings from the literature.<sup>56</sup> Our results suggest statistically significant and modest genetic correlations between the trauma and depression/neuroticism variables. Genetic correlations are 2-fold greater with CIDI depression as opposed to broad depression and neuroticism. A perception component of these traits i.e. the more extreme one is on the depression scale the more likely one is to either remember or perceive an event as traumatic, may explain these genetic correlations.

Our results provide depression and neuroticism heritability (proportion of variance captured by the common additive genetic variants) estimates and estimates for proportion of variance attributable to self-reported trauma exposure in line with previous literature.<sup>57-59</sup> Findings also suggest the sub-categories; childhood, adult and catastrophic trauma explain a substantial proportion of CIDI depression variance.

This study is the first to show contributions of genome-by-trauma interaction effects to depression (and related traits) phenotypic variation that are of relatively large magnitude (7-20%), approximately the same magnitude as the variance captured by self-reported trauma exposure itself (5-18%). These are consistent (although estimates are lower) across the sub-categories (childhood, adult and catastrophic) of trauma exposure (see **Figure 1**). Lower, yet still significant, interaction estimates are observed for broad depression and neuroticism phenotypes, except for the non-significant genome-by-catastrophic trauma interaction estimate within broad depression (supplementary table 13).

Model fit is also significantly improved when mixed linear models utilise environmental relationship matrices ( $E_s$ ) computed using only the first eigenvector (principal component (PC) 1) of trauma items. However, these estimates are substantially attenuated when compared to results from models including  $E_s$  computed using all PCs of trauma items. This suggests that important trauma exposure and genome-by-trauma interaction effects may be distributed across the different dimensions of self-reported trauma exposure. The inclusion of more self-reported trauma exposure PCs, may additionally capture the differential impact of subtypes of trauma exposure. Our research design and analyses are repeated across five geographical cluster samples. These within-sample replications, while not independent samples, yielded relatively consistent estimates and standard errors, increasing confidence in our results.

Exploring variance components of depression/neuroticism within males and females separately indicates that the proportion of CIDI depression variation captured by genome-by-trauma interaction effects differ substantially between the sexes, with estimates being approximately 5-fold greater in males. This pattern, although estimates were lower, was also observed for broad depression and neuroticism phenotypes. These results, alongside the evident prevalence differences, highlight the need to explore these effects within the sexes separately. Our findings suggest that trauma exposure and sensitivity to trauma exposure accounts for greater variance in depression/neuroticism outcomes for males. Whilst utilising principal components (PCs) enables the use of all trauma exposure variables, it is difficult to interpret directions of associations as higher PC values do not necessarily mean higher levels of trauma exposure. Further research can be conducted to explore the direction of these associations. Exploring individual trauma (e.g. neglect, physical abuse etc.) measures may provide a better understanding of the effect of specific trauma and genome-by-trauma experiences.

Our findings provide evidence of differential effects of trauma exposure dependent on differences in individual genetic liability to depression. Results can highlight both modifiable environments for preventative measures as well as at-risk groups that should be prioritised for these interventions. It is evident that the use of all genotyped variants yield substantial findings when exploring genome-by-trauma exposure interaction effects as opposed to much of the literature making use of PGSs.<sup>16, 20, 26, 30</sup> It is evident that the method employed here has major advantages over polygenic scores (PGSs), when exploring genome-by-environment interaction effects. However, there are limitations to this study that need to be considered when interpreting results.

Whilst, the GRMs computed utilise all genotyped SNPs, and subsequently account for greater variance than PGSs; discrepancies between twin study heritability and SNP heritability estimates of depression are still apparent. Twin study estimates may be biased upward due to the presence of gene-environment interplay effects, and thus, real heritability estimates are likely to fall between SNP heritability and twin study heritability estimates.<sup>60</sup> Moreover, results show that our environmental variables, full trauma and the sub-categories of trauma, have moderate heritability estimates and these are genetically correlated with our outcome phenotypes

(depression/neuroticism). This highlights that our environmental measures capture both genetic and environmental variances. As genomic relationship matrices are not capturing the entirety of the genetic variance within depression/neuroticism outcomes, the variance captured by the trauma (and sub-trauma) measures may capture residual genetic variance.

Here, to control for genetic covariance between our environmental and outcome variables, we also explored measures of trauma pre-corrected for the available genetic measure (GRMs). The differences in estimates of variance components were negligible. However, similar to the aforementioned limitation, this effect can be more accurately controlled for with an improved genetic measure (e.g. GRM utilising imputed or whole-genome sequenced data). Simulation findings from the literature suggests that making use of imputed or whole-genome sequencing genetic data for GRMs can uncover a further substantial proportion of genetic variance<sup>60</sup>, which would be useful in addressing the limitations outlined above (although would increase time and computational resources needed).

Our trauma exposure, depression and neuroticism variables were measured using retrospective self-report. Furthermore, measures of trauma exposure and CIDI depression were obtained later than measures of broad depression and neuroticism with the follow up UKB mental health questionnaire. This indicates potential measurement error within our variables.<sup>16</sup> However, it is important to note that self-report is likely the most effective method to explore these phenotypes in studies the size of UKB. Incorporating more objective measures of trauma exposure, e.g. omics measures (DNA methylation) may be able to provide a measure of trauma exposure that is less susceptible to reporting bias and thus, measurement error. For instance, the availability of methylation data has and continues to increase substantially, and can be used as good, and in some cases far better proxy measures of environments as seen with smoking.<sup>35, 61</sup> Evidence suggests there may be a methylation profile observed in individuals with trauma exposure.<sup>62, 63</sup> Genome-by-environment interaction effects using methylation data, can then be dissected to explore biological pathways with non-additive effects on outcomes that can be directly targeted. Findings could also further clarify the relationship between genetic liability and trauma exposure.

In conclusion, our findings provide empirical evidence of depression/neuroticism variation attributable to genome-by-trauma interaction effects. We further highlight that the magnitude of these effects are much larger for males in comparison to females. These findings can be further explored to identify both risk groups and modifiable environments/biological pathways which yield greater risk of depression manifestation, which would be useful in personalised/preventative interventions.

## References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602.
2. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys. *Epidemiologia e Psichiatria Sociale*. 2009.
3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012.

4. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* 2015;47(7):702-9.
5. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics.* 2018.
6. Howard DM, Adams MJ, Clarke TK, 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. *Nature Neuroscience.* 2019.
7. Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics (Fourth Edition)* 1996.
8. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - concepts and misconceptions. *Nat Rev Genet.* 2008;9(4):255-66.
9. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. 2009.
10. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. 2010.
11. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *American Journal of Psychiatry.* 2012.
12. Negele A, Kaufhold J, Kallenbach L, Leuzinger-Bohleber M. Childhood Trauma and Its Relation to Chronic Depression in Adulthood. *Depression Research and Treatment.* 2015.
13. Moskvina V, Farmer A, Swainson V, O'Leary J, Gunasinghe C, Owen M, et al. Interrelationship of childhood trauma, neuroticism, and depressive phenotype. *Depression and Anxiety.* 2007.
14. Hopfinger L, Berking M, Bockting CLH, Ebert DD. Emotion regulation mediates the effect of childhood trauma on depression. *Journal of Affective Disorders.* 2016.
15. Huh HJ, Kim KH, Lee HK, Chae JH. The relationship between childhood trauma and the severity of adulthood depression and anxiety symptoms in a clinical sample: The mediating role of cognitive emotion regulation strategies. *Journal of Affective Disorders.* 2017.
16. 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. *Molecular Psychiatry.* 2020.
17. Uher R. Gene-environment interactions in common mental disorders: An update and strategy for a genome-wide search. 2014.
18. Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological Medicine.* 2016.
19. Arnau-Soler A, Macdonald-Dunlop E, Adams MJ, Clarke TK, MacIntyre DJ, Milburn K, et al. Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Transl Psychiatry.* 2019;9(1):14.
20. Arnau-Soler A, Adams MJ, Hayward C, Thomson PA, Porteous D, Campbell A, et al. Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS ONE.* 2018.
21. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. Correction to: Genome-wide gene-environment interactions in neuroticism: an exploratory study across 25 environments. *Transl Psychiatry.* 2021;11(1):207.
22. Dunn EC, Wiste A, Radmanesh F, Almlil LM, Gogarten SM, Sofer T, et al. Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. *Depress Anxiety.* 2016;33(4):265-80.
23. Demirkan A, Penninx BWJH, Hek K, Wray NR, Amin N, Aulchenko YS, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular Psychiatry.* 2011.
24. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. *Plos Genet.* 2013.

25. Colodro-Conde L, Couvy-Duchesne B, Zhu G, Coventry WL, Byrne EM, Gordon S, et al. A direct test of the diathesis–stress model for depression. *Molecular Psychiatry*. 2018.
26. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. *British Journal of Psychiatry*. 2014.
27. Arnau-Soler A, Adams MJ, Clarke TK, MacIntyre DJ, Milburn K, Navrady L, et al. A validation of the diathesis-stress model for depression in Generation Scotland. *Translational Psychiatry*. 2019.
28. Shao N, Gong Y, Wang X, Wei J, Shi J, Ding H, et al. Effects of polygenic risk score, childhood trauma and resilience on depressive symptoms in Chinese adolescents in a three-year cohort study. *Journal of Affective Disorders*. 2021;282:627-36.
29. Gillett AC, Jermy B, Lee SH, Pain O, Howard DM, Hagenaars S, et al. Exploring polygenic-environment and residual-environment interactions for depressive symptoms within the UK Biobank. *medRxiv*. 2021.
30. Peyrot WJ, Van der Auwera S, Milaneschi Y, Dolan CV, Madden PAF, Sullivan PF, et al. Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biological Psychiatry*. 2018.
31. Nicholls HL, John CR, Watson DS, Munroe PB, Barnes MR, Cabrera CP. Reaching the End-Game for GWAS: Machine Learning Approaches for the Prioritization of Complex Disease Loci. *Front Genet*. 2020;11:350.
32. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*. 2020;9.
33. Mullins N, Lewis CM. Genetics of Depression: Progress at Last. 2017.
34. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*. 2011;88(1):76-82.
35. Amador C, Zeng Y, Barber M, Walker RM, Campbell A, McIntosh AM, et al. Genome-wide methylation data improves dissection of the effect of smoking on body mass index. *Plos Genet*. 2021;17(9):e1009750.
36. Docherty AR, Moscati A, Peterson R, Edwards AC, Adkins DE, Bacanu SA, et al. SNP-based heritability estimates of the personality dimensions and polygenic prediction of both neuroticism and major depression: findings from CONVERGE. *Translational Psychiatry*. 2016;6.
37. Vukasovic T, Bratko D. Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol Bull*. 2015;141(4):769-85.
38. Kendler KS, Gatz M, Gardner CO, Pedersen NL. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry*. 2006;63(10):1113-20.
39. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*. 2018.
40. Kendler KS, Gardner CO. Sex differences in the pathways to major depression: a study of opposite-sex twin pairs. *Am J Psychiatry*. 2014;171(4):426-35.
41. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: Come and get it. 2014.
42. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
43. Davis K, Hotopf M. Mental health phenotyping in UK Biobank. 2019.
44. 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.
45. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):93-121.
46. American Psychiatric Association DSAAPA. Diagnostic and statistical manual of mental disorders: DSM-5. 2013.

47. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*. 2020.
48. Eysenck SB, Eysenck H. J., Barrett P. A revised version of the psychoticism scale. *Personality and individual differences*. 1985;6(1):pp.21-9.
49. Glaesmer H, Brähler E, Schulz A, Freyberger H, Grabe HJ, Häuser W. The childhood trauma screener (CTS) - Development and validation of cut-off-scores for classificatory diagnostics. *Psychiatrische Praxis*. 2013.
50. Grabe HJ, Schulz A, Schmidt CO, Appel K, Driessen M, Wingenfeld K, et al. [A brief instrument for the assessment of childhood abuse and neglect: the childhood trauma screener (CTS)]. *Psychiatr Prax*. 2012;39(3):109-15.
51. Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. *BJPsych Open*. 2020;6(2):e18.
52. Team RC. R: A language and environment for statistical computing. 2020.
53. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, et al. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*. 2019;20(1):107.
54. Nivard MG, Middeldorp CM, Lubke G, Hottenga JJ, Abdellaoui A, Boomsma DI, et al. Detection of gene-environment interaction in pedigree data using genome-wide genotypes. *European Journal of Human Genetics*. 2016;24(12):1803-9.
55. Heckerman D, Gurdasani D, Kadie C, Pomilla C, Carstensen T, Martin H, et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(27):7377-82.
56. Warrier V, Kwong ASF, Luo M, Dalvie S, Croft J, Sallis HM, et al. Gene-environment correlations and causal effects of childhood maltreatment on physical and mental health: a genetically informed approach. *Lancet Psychiatry*. 2021;8(5):373-86.
57. Jermy BS, Glanville KP, Coleman JRI, Lewis CM, Vassos E. Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Mol Psychiatry*. 2021.
58. Realo A, van der Most PJ, Allik J, Esko T, Jeronimus BF, Koots-Ausmees L, et al. SNP-Based Heritability Estimates of Common and Specific Variance in Self- and Informant-Reported Neuroticism Scales. *J Pers*. 2017;85(6):906-19.
59. Schneider G, Kohnke C, Teismann H, Berger K. Childhood trauma and personality explain more variance in depression scores than sociodemographic and lifestyle factors-Results from the BiDirect Study. *Journal of Psychosomatic Research*. 2021;147.
60. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*. 2015;47(10):1114-20.
61. Andersen AM, Philibert RA, Gibbons FX, Simons RL, Long J. Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report. *Am J Med Genet B Neuropsychiatr Genet*. 2017;174(6):641-50.
62. Matosin N, Cruceanu C, Binder EB. Preclinical and Clinical Evidence of DNA Methylation Changes in Response to Trauma and Chronic Stress. *Chronic Stress (Thousand Oaks)*. 2017;1.
63. Mehta D, Bruenig D, Carrillo-Roa T, Lawford B, Harvey W, Morris CP, et al. Genomewide DNA methylation analysis in combat veterans reveals a novel locus for PTSD. *Acta Psychiatr Scand*. 2017;136(5):493-505.