bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank

Jonathan R.I. Coleman[1,2], Wouter J. Peyrot[3], Kirstin L. Purves[1], Katrina A.S. Davis[2,4], Christopher Rayner[1], Shing Wan Choi[1], Christopher Hübel[1,2], Héléna A. Gaspar[1,2], Carol Kan[4], Sandra Van der Auwera[5], Mark James Adams[6], Donald M. Lyall[7], Karmel W. Choi[8,9,10,11], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[12], Erin C. Dunn[10,11,13], Evangelos Vassos[1,2], Andrea Danese[1,14,15], Barbara Maughan[1], Hans J. Grabe[5], Cathryn M. Lewis[1,2], Paul F. O'Reilly[1], Andrew M. McIntosh[6], Daniel J. Smith[7], Naomi R. Wray[16,17], Matthew Hotopf[2,4], Thalia C. Eley[1,2*], Gerome Breen[1,2*]

[1] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[2] NIHR Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK

[3] Amsterdam UMC, Vrije Universiteit Medical Center, Department of Psychiatry, Amsterdam, The Netherlands

[4] Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[5] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

[6] Division of Psychiatry, University of Edinburgh, Edinburgh, UK

[7] Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

[8] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[9] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[10] Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, USA.

[11] Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.

[12] Consortium members listed in Supplementary Materials

[13] Department of Psychiatry, Harvard Medical School, Boston, MA, USA.

[14] Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[15] National and Specialist CAMHS Trauma and Anxiety Clinic, South London and Maudsley NHS Foundation Trust, London, UK

[16] Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

[17] Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia

* Address correspondence to Dr Gerome Breen (gerome.breen@kcl.ac.uk, +442078480409) or Prof. Thalia Eley thalia.eley@kcl.ac.uk, +442078480863), Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, SE5 8AF, UK.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Depression is more frequent among individuals exposed to traumatic events. However, the relationship between trauma exposure and depression, including the role of genetic risk factors, is complex and poorly understood. The UK Biobank concurrently assessed Major Depressive Disorder (MDD) and self-reported lifetime exposure to traumatic events in 126,522 genotyped individuals of European ancestry. We compared MDD cases and healthy individuals reporting trauma exposure with those who did not (final sample size range: 24,094-92,957). The SNP-heritability of MDD was greater in participants reporting trauma exposure than in individuals not reporting trauma exposure. Genetic correlations between MDD and psychiatric traits were strong regardless of reported trauma exposure, whereas genetic correlations between MDD and both BMI (and related traits) and educational attainment were observed only in individuals reporting trauma exposure. The differing SNP-heritability and genetic correlations with MDD when stratified by reported trauma exposure suggests the genetic contribution to MDD is greater when additional risk factors are present.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Depression is among the most common mental illnesses worldwide and accounts for 5.5% of all years lost through disability globally [1]. In England approximately 28% of individuals self-report depression during their lifetime [2]. The most common clinically recognised form of depression is called Major Depressive Disorder (MDD). Both environmental and genetic factors influence MDD. In particular, MDD is more commonly observed among individuals reporting exposure to stressful life events and early-life traumas [3–6]. In turn, reported trauma exposure has been robustly correlated with a range of adverse life outcomes including MDD [6–9]. The relationship between MDD and reported trauma exposure is complex, with studies showing both that reported trauma exposure is associated with subsequent MDD, and that MDD is associated with subsequent reported trauma exposure [10,11]. However, the majority of people reporting exposure to traumatic experiences do not report MDD [6–9].

Twin studies show that MDD is moderately heritable, with 30-40% of the variance in MDD attributable to genetic factors [12]. The proportion of heritability captured by common genetic variants, also known as single nucleotide polymorphism or SNP-heritability, can be estimated from genome-wide association study (GWAS) data. Such estimates tend to be lower than those obtained from twin approaches, due to the incomplete capture of genetic information in GWAS data among other reasons [13]. The most recent major depression GWAS from the Psychiatric Genomics Consortium was anchored in 35 cohorts (including the 23andMe discovery cohort [14]) recruited with a variety of methods [15]. This meta-analysis identified 44 loci significantly associated with major depression, and estimated a SNP-heritability of 9-10% [15]. GWAS results strongly suggest both the

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mild and more severe forms of depression are polygenic, with potentially thousands of variants with very small individual effects contributing to risk.

There are far fewer genetic studies of reported trauma exposure than of MDD. However, the available studies have demonstrated that reported trauma exposure is heritable, with twin heritability estimates of 20-50% [16–18] and SNP-heritability estimates of 30% [19]. Combining measures of trauma exposure and depression on a large scale is difficult and, as with many environmental measures, requires careful phenotyping [20]. Potential confounds include the (often unavoidable) use of retrospective self-reported measures of trauma exposure, which can be weakly correlated with objective measures of traumatic experiences [9]. Furthermore, current (i.e. state) low mood or MDD can increase self-reporting of previous trauma exposure [9,21]. Previous individual study cohorts have generally been too small for effective GWAS, while meta-analyses have contained considerable heterogeneity due to the use of different phenotyping instruments in the included studies.

However, some notable genome-wide analyses of MDD and trauma exposure have been performed. A genome-wide by environment interaction study of depressive symptoms and stressful life events in 7,179 African American women identified a genome-wide association near the *CEP350* gene (although this did not replicate in a smaller cohort) [22]. A recent investigation in 9,599 Han Chinese women with severe MDD identified three variants associated with MDD specifically in individuals who did not report exposure to traumatic events prior to MDD onset [23].

Several attempts have been made to estimate the interaction of overall genetic risk and trauma by using polygenic risk scores (PRS) for MDD to perform PRS-by-trauma interaction analyses. Such studies test whether there are departures from additivity (where the combined effect of PRS and trauma differs from the sum of

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the individual effects) or from multiplicativity (where the combined effect differs from the product of the individual effects). Results from these have been highly varied, with reports of significant additive and multiplicative interactions [24]; significant multiplicative interactions only [25]; and, in the largest previous study published (a meta-analysis of 5,765 individuals), no interactions [26].

The release of mental health questionnaire data from the UK Biobank resource provides an opportunity to assess the relationship between genetic variation, risk for MDD, and reported trauma exposure in a single large cohort. We performed GWAS of probable MDD with and without reported lifetime trauma exposure in UK Biobank European ancestry individuals. We used GWAS results to estimate the SNP-heritability of MDD in individuals with and without reported lifetime trauma exposure. We then estimated the genetic correlation between MDD in the two groups, and compared the patterns of genetic correlation between MDD in these groups and a wide range of physical and psychiatric traits. Finally, we performed polygenic risk scoring, using MDD-relevant external traits, and sought to extend previous analyses of PRS-by-trauma interactions in MDD.

Methods

Phenotype definitions

The UK Biobank assessed a range of health-related phenotypes and biological measures including genome-wide genotype data in approximately 500,000 British individuals aged between 40 and 70 [27]. This includes 157,366 participants who completed an online follow-up questionnaire assessing common mental health disorders, including MDD symptoms, and 16 items assessing traumatic events (Resource 22 on http://biobank.ctsu.ox.ac.uk) [28]. Phenotypes were derived from this

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

questionnaire, using definitions from a recent publication describing its phenotypic structure [28].

Individuals with probable MDD met lifetime criteria based on their responses to questions derived from the Composite International Diagnostic Interview (CIDI; Supplementary Table 1). We excluded cases if they self-reported previous diagnoses of schizophrenia, other psychoses, or bipolar disorder. Controls were excluded if they self-reported any mental illness, reported taking any drug with an anti-depressant indication, had previously been hospitalised with a mood disorder or met previously-defined criteria for a mood disorder (Supplementary Table 1) [29].

Participants were asked questions relating to traumatic experiences in childhood using the Childhood Trauma Screener (a shortened version of the Childhood Trauma Questionnaire [30–32]) and an equivalent screener for adulthood developed by the UK Biobank Mental Health steering group to mirror the childhood items [28]. In addition, participants were asked questions related to events that commonly trigger post-traumatic stress-disorder (PTSD). Responses to individual questions were dichotomised and assessed for enrichment in MDD (Supplementary Table 2a).

We selected reported events with an odds ratio > 2.5 with MDD, to obtain a single binary variable for stratification that captured exposure to the traumatic events most associated with MDD. Items from all three areas of reported trauma assessment were enriched in MDD cases. Three items referred to events in childhood (did not feel loved, felt hated by a family member, sexually abused), three to events in adulthood (physical violence, belittlement, sexual interference) and one PTSD-relevant event (ever a victim of sexual assault). In order to capture increased severity of exposure, only individuals reporting two or more enriched items were

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

included as reporting trauma exposure, while those reporting none of the items were included as not reporting trauma exposure. Individuals reporting a single trauma item, or who did not provide an answer were excluded from the analyses (Supplementary Table 1). A breakdown of reported traumatic experiences by sex and MDD status is provided in Supplementary Table 2b. Further discussion of the definition of trauma exposure is included in the Supplementary Note.

Phenotype preparation for analyses

Three sets of analyses were performed (i) comparing MDD cases and controls overall, (ii) limited to individuals reporting trauma exposure, and (iii) limited to individuals not reporting trauma exposure (Table 1). In addition, sensitivity analyses were performed on reported trauma exposure (overall and stratified by MDD diagnosis; see Supplementary Methods and Results, and Supplementary Table 3). For each analysis, phenotypes were first residualised on 6 ancestry principal components from the genetic data of the European samples as well as factors capturing initial assessment centre and genotyping batch. More details on phenotype preparation can be found in the Supplementary Methods.

Phenotype distribution

Previous analyses have shown that, compared to the participants in the UK Biobank as a whole, those who completed the mental health questionnaire were more likely to have a university degree, came from a higher socioeconomic background, and reported fewer longstanding illnesses or disabilities [28]. Accordingly, participants were compared across a number of standard demographic variables and common correlates of MDD: sex, age (at questionnaire), education (university

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

degree vs. not), neighbourhood socioeconomic status (SES, as Townsend deprivation index [33]) and BMI (recorded from measurements taken at the initial recruitment of the participants into the biobank). For further details on these analyses, see Supplementary Methods.

Genetic data

Genetic data for GWAS analyses came from the full release of the UK Biobank data (N=487,410; [34]). Autosomal genotype data from two highly-overlapping custom genotyping arrays (covering ~800,000 markers) underwent centralised quality control before being imputed in a two-stage imputation to the Haplotype Reference Consortium (HRC) and UK10K (for rarer variants not present in the HRC) reference panels [34–36]. In addition to this central quality control, variants for analysis were limited to common variants (minor allele frequency > 0.01) that were either directly genotyped or imputed from the HRC with high confidence (IMPUTE INFO metric > 0.4) [35].

Individuals were excluded where recommended by the UK Biobank core analysis team for unusual levels of missingness or heterozygosity, or if they had withdrawn consent for analysis. Using the genotyped SNPs, individuals with call rate < 98%, who were related to another individual in the dataset (KING r < 0.044, equivalent to removing up third-degree relatives and closer [37]) or whose phenotypic and genotypic gender information was discordant (X-chromosome homozygosity (F_X) < 0.9 for phenotypic males, F_X > 0.5 for phenotypic females) were also excluded. Removal of relatives was performed using a "greedy" algorithm, which minimises exclusions (for example, by excluding the child in a mother-father-child trio). All analyses were limited to individuals of European ancestry, as defined by 4-means

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

clustering on the first two genetic principal components provided by the UK Biobank [38]. This ancestry group included 95% of the respondents to the mental health questionnaire - as such, the non-European ancestry groups were considered too small to analyse informatively. Principal components analysis was also performed on the European-only subset of the data using the software flashpca2 [39]. After quality control, individuals with high-quality genotype data and who had completed the online mental health questionnaire were retained for analysis (N=126,522).

Polygenic risk score analyses and SNP-heritability analyses in BOLT-LMM used the genotyped variants [34]. Variants for these analyses were limited to common variants (minor allele frequency > 0.01) with call rate >98% that were in approximate Hardy-Weinberg equilibrium (HWE test $p > 10^{-8}$). The same individuals were used for analyses using the imputed and the genotyped data.

Analyses

Genome Wide Association Studies (GWAS)

GWAS were performed to assess the association of individual variants with MDD (overall and stratified by reported trauma exposure). GWAS were performed using linear regressions on imputed genotype dosages in BGenie v1.2 [34], with residualised phenotypes as described above. Phenotypes and genotypes were mean-centred and standardised. Genome-wide significance was defined at the conventional level $p < 5 \times 10^{-8}$ [40]. Results from each GWAS were clumped to define genetic loci in PLINK2 [41]. Loci were defined following established protocols (Supplementary Methods) [15].

Betas from the GWAS were converted to odds ratios (OR) using LMOR (http://cnsgenomics.com/shiny/LMOR/) and observed sample prevalences [42].

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Standard errors were calculated from the p-value and estimated OR [43]. Performing GWAS on residuals, rather than including covariates in the analysis, is a restriction imposed by the BGenie software (which was used because it is specifically designed for analysing the UK Biobank genetic data). Sensitivity analyses were performed to test for biases resulting from this method. Specifically, for each GWAS, each variant with nominal significance ($p<0.0001$) was also tested using logistic regression including covariates in R 3.4.1, in order to confirm the results from BGenie [44].

SNP-heritability

Results from GWAS were combined to assess the proportion of variance due to the additive effect of common genetic variants (SNP-heritability). SNP-heritability was calculated on the observed scale using BOLT-LMM v2.3 [45]. The estimate for MDD in the cohort was converted to the liability scale in R 3.4.1, assuming a population prevalence of 28% [2,46]. Converting estimates of SNP-heritability for a case-control trait from the observed scale to the liability scale requires accurate estimates of the prevalence of the trait in the (sub)population. When comparing a trait stratified by a correlated variable (as is the case when we compare the SNP-heritability of MDD stratified by reported trauma exposure), the population prevalence in each stratum is unknown. To address this, we approximated the expected prevalence of MDD in individuals either reporting or not reporting trauma exposure (Supplementary Methods). This allowed us to convert the observed scale SNP-heritability of MDD to the liability scale in both strata. A second challenge is that trauma exposure is itself a heritable trait that is genetically correlated with MDD in this study. The potential impact of this on SNP-heritability estimation is not intuitive. To benchmark our findings, we performed simulations of SNP-level data to explore

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the expected SNP-heritability of MDD in individuals reporting and not reporting trauma exposure, assuming differences in SNP-heritability resulted only from the genetic correlation between MDD and reported trauma exposure. Further details of these analyses are provided in the Supplementary Methods.

Genetic correlations

Genetic correlations were calculated to assess shared genetic influences between MDD and other phenotypes. Genetic correlations (r_g) were calculated in LD Score v1.0.0 [47] using the default HapMap LD reference. Two sets of genetic correlations were calculated: between the stratified GWAS results from this analysis (internal phenotypes), and between each GWAS from this analysis and a curated list of 392 publically-available phenotypes (external phenotypes) [47,48].

Genetic correlations were tested for difference from 0 (default in LD Score), and for difference from 1 (in Microsoft Excel, converting r_g to a chi-square as $[(r_g-1)/se]^2$) [47,48]. Genetic correlations were considered significant if they passed the Bonferroni-adjusted threshold for each analysis (internal: $p < 3x10^{-3}$; external: $p < 1.3x10^{-4}$), which is conservative in this instance as tests were not strictly independent.

The genetic correlation of MDD with each external phenotype was compared between individuals reporting trauma exposure and individuals not reporting trauma exposure using a two-stage method. First, differences were assessed using two sample z-tests [49]. Nominally-significant differences ($p < 0.05$) by this method were then compared using the block-jackknife (Supplementary Methods) [48,50,51]. Results using the jackknife were considered significant if they passed the Bonferroni-adjusted threshold ($p < 1.3x10^{-4}$).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Polygenic Risk Scoring

Polygenic risk scores (PRS) were calculated to further assess shared genetic influences between MDD and other traits, and to estimate the effect on MDD of interactions between the whole genome (modelled as PRS) and reported trauma exposure. Polygenic risk scores (PRS) were calculated using PRSice v2 (https://github.com/choishingwan/PRSice [41,52]). Specifically, PRS from analyses of major depression (PGC MDD + 23andMe) [15], schizophrenia (SCZ) [53], bipolar disorder (BIP) [54], body mass index (BMI) [55] and glycated haemoglobin (HbA1c; used as a negative control) [56] were calculated and compared in all participants and stratifying by reported trauma exposure. The PGC major depression GWAS contained participants from UK Biobank, so to derive the PGC MDD + 23andMe PRS we used a restricted set of summary statistics without these individuals (but including individuals from 23andMe [14]) - for further discussion of this overlap, see Supplementary Note [15]. Analyses used logistic regression, including all covariates used in creating the residuals for GWAS. PRS were calculated at seven thresholds (external GWAS $p < 0.001$, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5) to allow assessment of the spread of association between PRS and MDD. In total, five external phenotypes were used to produce PRS for three target phenotypes (MDD overall and stratified by reported trauma exposure), resulting in 15 analyses. A conservative Bonferroni adjustment for multiple testing was used, correcting for 105 tests (seven thresholds and 15 analyses), giving a final threshold for significance of $p < 0.0004$.

In addition to these stratified analyses, we performed formal PRS-by-environment analyses, calculating interactions between reported trauma exposure and the PRS capturing the most variance from each of the main-effect PRS

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analyses. These analyses included the same covariates used in the GWAS, and all PRS-by-covariate and reported trauma exposure-by-covariate interactions [57,58]. Both multiplicative and additive interactions were tested. Multiplicative interactions test the null that the PRS and reported trauma exposure combine as the products of their individual effects, and were tested using logistic regression [25,26]. Additive interactions test the null that the PRS and reported trauma exposure combine as the sum of their individual effects.

Sensitivity analyses

Differences in phenotypic variables were observed between cases and controls. To assess the impact of including these variables as covariates, all analyses were rerun retaining all previous covariates and including as further covariates: age (at questionnaire), neighbourhood socioeconomic status (SES, as Townsend deprivation index [33]), BMI (at baseline assessment), and a binary variable of education (university degree vs. not). The same covariates were also included in PRS and SNP-heritability analyses. Sensitivity analyses focussing on reported trauma exposure as an outcome were similarly rerun (Supplementary Methods).

The majority of the cohort with data on both MDD symptoms and reported trauma status were controls who did not report trauma (Table 1). To assess whether this disbalance in sample status affected our results, genetic correlation analyses with external phenotypes were rerun on ten downsampled cohorts, each with 9,487 participants (the number of cases not reporting trauma exposure; see Supplementary Methods).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Phenotype distribution

Phenotypic and genetic data were available on 24,094 to 92,957 individuals (Table 1). Overall, 36% of individuals met our definition of MDD-relevant trauma exposure, and were more frequently cases (45%) than controls (17%; OR = 5.23; $p < 10^{-50}$, chi-square test). Cases differed significantly from controls, both overall and when taking reported trauma exposure into account. Individuals with MDD were mostly females, significantly younger, less likely to have a university degree, came from more deprived neighbourhoods, and had higher BMI at recruitment (all $p \leq 0.001$; Supplementary Table 4).

		Participants with genomic data			
		Reported trauma exposure	No reported trauma exposure	Excluded	Total
MDD	Cases	**13,393** [b]	**9,487** [c]	6,595	**29,475** [a]
	Controls	**10,701** [b]	**39,677** [c]	13,104	**63,482** [a]

Table 1: Participants available for analysis.
Groups of individuals used in each of the three analyses are in bold.
The superscripts denote the groups used in each of the three main analyses:
a) MDD in all participants
(29,475 cases, 63,482 controls, N = 92,957)
b) MDD in participants reporting trauma exposure
(13,393 cases, 10,701 controls, N = 24,094)
c) MDD in participants not reporting trauma exposure
(9,487 cases, 39,677 controls, N = 49,164).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-wide association studies

We performed GWAS for MDD overall and stratified by reported trauma exposure (Supplementary Table 6; Supplementary Figures 1-3). No analysis showed evidence of genome-wide inflation that was attributable to confounding (95% confidence intervals of all regression intercepts from LD Score included 1; Supplementary Table 7). One genome-wide significant locus (rs11515172, Chr 9:11Mb, $p = 3.82 \times 10^{-8}$) was identified in the analysis of MDD overall, and remained significant when using logistic regression ($p = 4.69 \times 10^{-8}$, OR = 0.96, SE = 0.007; Supplementary Table 6). This locus has been repeatedly associated with depression [15,59,60] and with neuroticism [61–64]; however, it should be noted that all of these studies included UK Biobank. The locus is intergenic, and is not annotated to any currently known biological feature of interest (Supplementary Table 8).

SNP-heritability

The liability-scale estimate of MDD SNP-heritability overall was 20% (95% confidence interval: [18-22%]; Supplementary Table 7). The liability scale SNP-heritability of MDD in those reporting trauma exposure was 24% [18-31%], with approximated prevalence of MDD in the trauma-exposed population of 52%. In those not reporting trauma exposure, the liability scale SNP-heritability was 12% [7-16%], and the approximated prevalence of MDD in the non-exposed population was 17% (Supplementary Table 7). This difference was significant ($p = 0.0021$, Z-test).

These estimated SNP-heritabilities could be confounded by the genetic correlation between MDD and reported trauma exposure in the absence of interaction. We thus conducted simulations of SNP-level data, which yielded expected estimates for the liability scale SNP-heritability of MDD 14-15% in those

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reporting trauma exposure and 15-16% in those not reporting trauma exposure (Supplementary Methods). Our simulations thus suggest that our empirical findings were not confounded by the heritable component of reported trauma exposure, nor by the transformation from the observed scale to the liability scale.

Genetic correlations

Genetic correlations were calculated between all internal phenotypes (those assessed in the UK Biobank in this analysis; Supplementary Table 10). The genetic correlation between MDD in individuals reporting trauma exposure and MDD in individuals not reporting trauma exposure was high and did not differ significantly from 1 (r_g = 0.77 [0.48-1.05]; difference from 0: $p = 1.8 \times 10^{-7}$; difference from 1: p = 0.11).

We observed a significant r_g between MDD and reported trauma exposure in the full cohort (0.62 [95% CI: 0.76-0.94], $p < 10^{-50}$). Given that trauma items were selected for association with MDD, we also calculated the genetic correlation between MDD in the full cohort and reported trauma exposure in the controls, which was also significantly greater than 0 (0.31 [0.18-0.45], $p = 4\times10^{-6}$; Supplementary Table 10). This correlation persisted when using independent major depression GWAS summary statistics, as reported trauma exposure was significantly correlated with the recent PGC MDD + 23andMe PRS (Spearman's rho = 0.0675, $p < 10^{-50}$) [15].

Genetic correlations between MDD and external psychiatric traits were significant ($p < 1.3\times10^{-4}$) regardless of trauma exposure, and did not differ substantially in magnitude between the groups (z-test for comparisons of r_g ranged from p = 0.104 - 0.992; Figure 1). In contrast, correlations between MDD and body

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

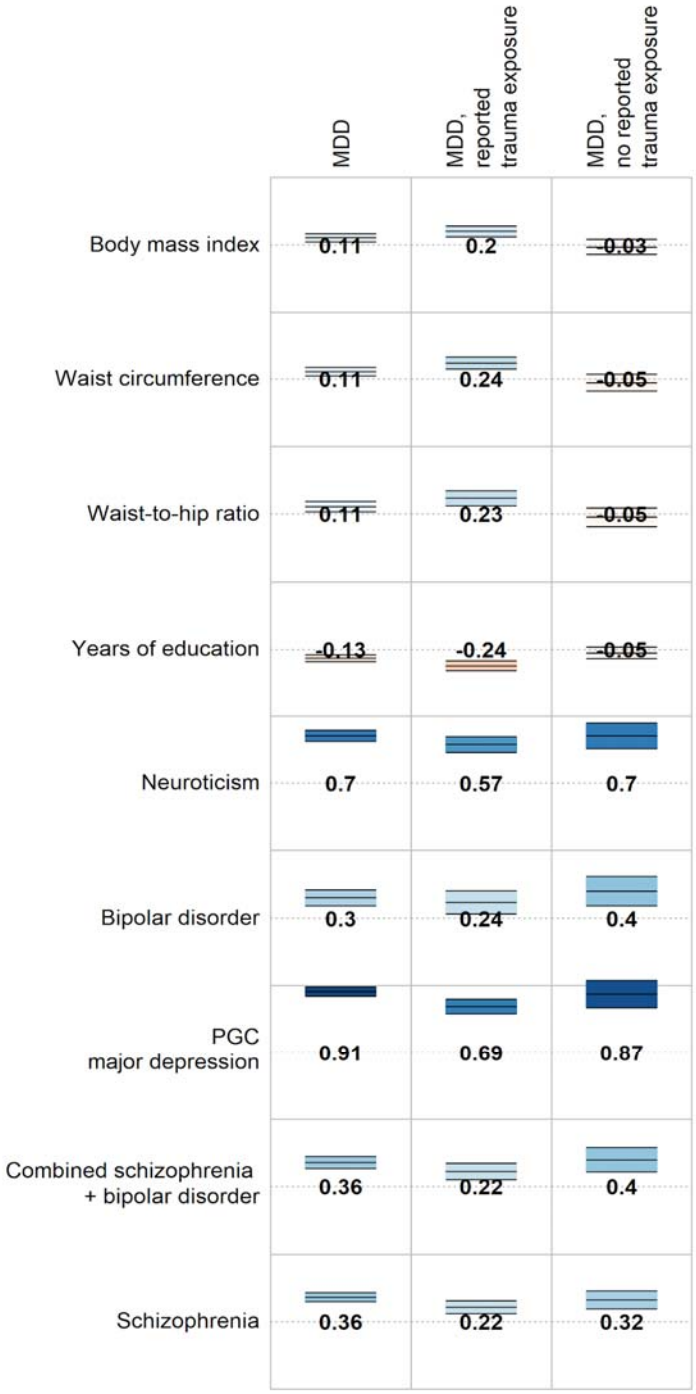

Figure 1: Genetic correlations between MDD (overall and stratified by reported trauma exposure) and selected traits and disorders. Full genetic correlation results are available in Supplementary Table 11. Numbers = genetic correlations. Colour = direction of effect (blue = positive, red = negative). Colour intensity = size of correlation. Upper and lower bars are 95% confidence interval of genetic correlation.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

composition, reproductive, and socioeconomic phenotypes were only significant in individuals reporting trauma exposure. These correlations were also notably larger in this group reporting trauma exposure compared to individuals not reporting trauma exposure, although no differences remained significant following multiple testing correction (all jackknife $p > 1.3x10^{-4}$; Figure 1, Supplementary Table 11).

Polygenic risk scores across strata

Individuals with high genetic risk scores for PGC MDD + 23andMe were more likely to be cases than controls, in all three analyses (Table 2; see Supplementary Table 12 for full details of all risk score analyses, including the number of SNPs in each score). A similar pattern was observed with SCZ risk scores. Individuals with higher BIP risk scores were more likely to be cases overall (Table 2 upper left panel) and when considering individuals not reporting trauma exposure (Table 2 lower panel), but not when considering individuals reporting trauma exposure (Table 2 middle panel, Supplementary Table 12). Conversely, those with higher BMI risk scores were more likely to be cases than controls overall and in individuals reporting trauma exposure, but not in individuals not reporting trauma exposure. No significant differences were observed in the negative control analysis with HbA1c.

Polygenic risk score by trauma exposure interactions

In participants with data on MDD and on reported trauma, significant additive interaction terms were observed from linear regression when using the PGC MDD + 23andMe PRS and when using the BMI PRS - the combined effect of PRS and reported trauma exposure was greater than the sum of the individual effects (beta > 0, Table 2 upper right panel). A significant multiplicative interaction term was also

Base	Base N	Best Threshold	Analysis	PRS			PRS x Reported Trauma					
							Multiplicative			Additive		
				OR	95% CI	p	OR	95% CI	p	Beta	95% CI	p
PGC MDD + 23andMe	116,404 // 314,990	**0.5**	*MDD*	**1.26**	**1.24-1.28**	**< 10^{-50}**	1.01	1.00 - 1.03	0.132	**0.011**	**0.008-0.014**	**2.69×10^{-11}**
SCZ	36,989 // 113,075	**0.3**		**1.11**	**1.09-1.12**	**1.11×10^{-41}**	0.99	0.97-1.00	0.158	0.008	-0.003-0.004	0.659
BIP	7,481 // 9,250	**0.2**		**1.07**	**1.05-1.08**	**4.57×10^{-19}**	0.99	0.97-1.00	0.165	-0.000	-0.003-0.003	0.961
BMI	339,224	**0.3**		**1.04**	**1.03-1.06**	**2.60×10^{-8}**	**1.02**	**1.01-1.04**	**0.0074**	**0.006**	**0.003-0.009**	**1.13×10^{-4}**
HB1Ac	46,368	0.001		1.01	1.00-1.02	0.163	1.01	0.99-1.02	0.391	0.002	-0.001-0.005	0.186
PGC MDD + 23andMe	116,404 // 314,990	**0.4**	*MDD, reported trauma exposure*	**1.24**	**1.21-1.27**	**< 10^{-50}**						
SCZ	36,989 // 113,075	**0.5**		**1.06**	**1.03-1.09**	**2.96×10^{-5}**						
BIP	7,481 // 9,250	0.5		1.04	1.01-1.07	0.00329						
BMI	339,224	**0.5**		**1.07**	**1.04-1.10**	**1.85×10^{-7}**						
HB1Ac	46,368	0.001		1.02	1.00-1.05	0.0863						
PGC MDD + 23andMe	116,404 // 314,990	**0.4**	*MDD, no reported trauma exposure*	**1.21**	**1.18-1.23**	**< 10^{-50}**						
SCZ	36,989 // 113,075	**0.5**		**1.09**	**1.06-1.11**	**4.34×10^{-12}**						
BIP	7,481 // 9,250	**0.2**		**1.07**	**1.04-1.09**	**4.05×10^{-8}**						
BMI	339,224	0.3		1.02	1.00-1.04	0.0980						
HB1Ac	46,368	0.001		1.01	0.99-1.03	0.492						

Table 2: Main effect and interaction effects for polygenic risk scores (PRS) associated with MDD overall and in stratified analyses. Interaction effects are on the multiplicative scale (OR) and the additive scale (Beta). Bold = significant associations (main analyses: $p < 0.000143$; interactions: $p < 0.01$). Base N = Cases // Controls. OR/Beta = Increase with 1 SD increase in PRS or trauma exposure. Results are reported at the "best" threshold (that with the lowest p-value in main effect analyses) - results across all thresholds are reported in Supplementary Table 12.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed for the BMI PRS - the combined risk of MDD from BMI PRS and reported trauma exposure was greater than expected from the product of the individual risks (OR > 1). The interaction results between PGC MDD + 23andMe PRS and reported trauma exposure were in line with the difference in MDD SNP-heritability between individuals reporting (SNP-h^2 =24%) and not reporting trauma exposure (SNP-h^2 = 12%) described above.

Sensitivity analyses

Sensitivity analyses were performed: i) focussed on reported trauma exposure (overall and stratified by MDD), ii) to assess the impact of controlling for age, neighbourhood socioeconomic status, BMI, and education, and iii) downsampling such that each group had 9,487 participants. Results from parallel analyses focussed on reported trauma exposure (overall and stratified by MDD) were broadly similar to the results from analyses of MDD (Supplementary Results). Controlling for age, neighbourhood socioeconomic status, BMI, and education did not alter the conclusions drawn from the GWAS, SNP-heritability analyses, nor the genetic correlations observed between the internal phenotypes (Supplementary Results). Genetic correlations between MDD and external phenotypes did not differ significantly from the main analysis (all z-test $p < 0.05$), but were sufficiently attenuated that the genetic correlations of MDD with body composition phenotypes were no longer significant in individuals reporting trauma exposure (although significant correlations with sociodemographic and educational phenotypes remained). Differences in the PRS analyses were limited to analyses involving the BMI PRS. BMI PRS was no longer associated with MDD in any analysis, and no interactions including the BMI PRS remained significant.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In analyses using equally-sized downsampled groups, genetic correlations between MDD and external phenotypes were attenuated across most phenotypes, but not significantly (two-sample z-tests, all $p > 0.05$). As such, the general pattern of genetic correlations observed in the main analysis was retained, although only the correlation between MDD and waist circumference would have remained significant after multiple testing correction in this downsampled cohort.

Full results for all three sensitivity analyses are included in the Supplementary Material.

Discussion

We investigated the relationship between MDD and self-reported trauma exposure in the largest single cohort available to date (N=73,258 with MDD and reported trauma data), examining individual genetic variants, SNP-heritability, genetic correlations, and polygenic risk scores. Our analyses suggest higher SNP-heritability of MDD in individuals reporting trauma exposure, with a wider range of genetic correlations observed, compared to individuals not reporting trauma exposure.

Significant differences in SNP heritability, combined with the high genetic correlation between MDD in individuals reporting trauma exposure and in individuals not reporting trauma exposure (which was not statistically different from 1), suggest that the same set of variants are associated with MDD in both strata, but that their total effect is increased in those reporting trauma exposure. Importantly, this could reflect a smaller environmental component of variance in these individuals, reflecting their shared exposure to MDD-relevant traumatic experiences.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our estimate of the SNP-heritability of MDD (20%) is higher than that reported in previous studies of major depression (~9%) [15]. This may be explained by the relative homogeneity of the UK Biobank compared to previous meta-analyses. The UK Biobank is a single-country cohort ascertained using a consistent protocol. The same questionnaire was used to gather symptom data, and the samples were stored, extracted, and genotyped using a single method. In contrast, meta-analyses have needed to combine diverse ascertainment, sampling, and genotyping.

Simulations suggest the significantly higher SNP-heritability of MDD in individuals reporting trauma exposure was not attributable to gene-environment correlation between MDD and reported trauma exposure, nor to the transformation of the observed scale heritability to the liability scale. However, we did not address sources of potential bias from genetic architectures other than those simulated, from intrinsic challenges of heritability estimation in case-control data [65,66], or from potential collider bias resulting from selection bias [67].

We also assumed that the population prevalence of reported trauma exposure can be extrapolated from that observed in this sample (see Supplementary Methods). Although the UK Biobank allows us to integrate genetic and environmental data at scale, in a reasonably homogeneous cohort, it has a "healthy volunteer bias", whereby the participants tend to have better overall health and higher socioeconomic status compared to the equivalent overall population of this age [68]. It is possible that the depressive and traumatic experiences reported by these participants may not generalise to the whole population, or to clinically-ascertained cases. Furthermore, we focussed on European ancestry; further studies in non-European populations are required [69].

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The high genetic correlation between MDD in individuals reporting and not reporting trauma exposure was supported by significant genetic correlations between MDD and other psychiatric disorders regardless of reported trauma exposure. In individuals reporting trauma exposure, further significant genetic correlations were observed between MDD and anthropometric traits like greater body mass index, as well as with poorer educational attainment, both of which are known to be associated with MDD [70–72]. The genetic correlations with MDD observed only in individuals reporting trauma exposure are consistent with previous literature on traumatic experiences across the life course, and related phenomena such as Adverse Childhood Experiences (ACEs). This literature has found that such adversities are associated not only with psychiatric risk but also with wide-ranging impairments in social and health outcomes including obesity and education [73–76]. However, we stress that causal conclusions should not be drawn from these data, or that the reported trauma exposure is responsible for the observed differences.

In PRS-by-reported trauma exposure interaction analyses, we identified a significant deviation from additivity for the MDD PRS and reported trauma exposure on risk of MDD. These results are in line with the larger SNP-heritability of MDD in exposed compared to unexposed individuals. The simplest explanation for this result is that the effect of the MDD PRS and reported trauma exposure on MDD combine multiplicatively, rather than additively. For BMI PRS however, the interaction with reported trauma exposure appears to be more complex, being neither additive nor multiplicative. In sensitivity analyses controlling for BMI (obtained at recruitment, approximately five years before the mental health questionnaire), the BMI PRS-by-reported trauma exposure interaction was no longer significant, suggesting that the observed interaction may reflect differences in BMI. Further research, with

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

concurrent measurements of BMI, trauma exposure and MDD in a longitudinally-sampled cohort would offer further insight into the potentially causal relationship between these three variables.

There are a number of limitations to consider when interpreting our PRS-by-environment interaction analyses. PRS-by-environment interaction analyses test a specific hypothesis, namely that the overall association of common variants with the outcome (modelled as a PRS) varies dependent on the environmental exposure being tested. As such, the absence of a PRS-by-environment interaction does not preclude the existence of specific gene-by-environment interactions, including those featuring variants contributing to the PRS. Furthermore, we cannot exclude the possibility that the correlation between the MDD PRS and reported trauma exposure may alter the observed interaction. As such, caution is needed before drawing strong conclusions, especially given the small effect sizes, and limited predictive power, of PRS in this study (Supplementary Table 12).

Throughout this paper, we have referred to our depression phenotype as "MDD" rather than "major depression". We do this because our definition is based on the CIDI-SF, which has previously been shown to have good concordance with direct clinical assessments of MDD [77,78]. However, it should be noted that direct assessment was not performed, and our probable MDD cases may not have met criteria within a clinical setting. Nonetheless, genetic correlations between studies of MDD and of major depression are high, suggesting there is strong genetic continuity across different methods of assessing depression [15,60].

Our results differ in several respects from those of a study of MDD and adversity in Han Chinese women [23]. No difference in the SNP-heritability of MDD between individuals reporting and not reporting trauma exposure was observed in

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the previous study, and we did not replicate individual variant results. However, this is unsurprising, as there are a number of differences between the studies of which the primary one is sample size (this study: 73,258; CONVERGE: 9,599). Other differences included culture and ethnicity, and the deeper phenotyping methodology applied in CONVERGE, resulting in a severe MDD phenotype. Notably, the previous study did not report a genetic correlation between MDD and trauma exposure [23].

Sensitivity analyses focussed on trauma found that self-reported traumatic experience was significantly heritable, as has been previously observed [19]. We strongly emphasise that this does not necessarily imply that traumatic experiences themselves have a biological component - such experiences may be associated with other significantly heritable traits, and their biology would then be reflected in the observed heritability of trauma exposure. One potential set of heritable traits that may be associated with reporting traumatic experiences are personality traits such as risk-taking, and this might explain the observed genetic correlations with psychiatric traits. A similar phenomenon has been proposed to underlie observed genetic correlations with socioeconomic status [79]. Our trauma exposure measure relies on retrospective self-report, which is correlated with personality traits and mood at time of report [9]. This may also explain the genetic correlations we observe with reported trauma exposure (including in controls, who do not report previous psychiatric illness).

Retrospective self-report of this kind is not the ideal measure for this phenotype, and precludes robust measurement of the severity and timing of the reported trauma exposure. However, retrospective report is the only feasible option in large cohorts like the UK Biobank. The requirement for cohort sizes large enough

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to identify the small individual genetic effects typical of complex genetic traits such as MDD makes self-report the most practicable method of data-collection.

Retrospectively reported trauma and MDD data are not robust to reverse causation, and our results cannot strongly inform any temporal or causal hypotheses about the relationship between trauma and MDD. To test such hypotheses would require either longitudinal studies (with the inherent logistical difficulties in obtaining both environmental and genomic data) or more powerful genomic studies of trauma exposure in a larger cohort to generate sufficient robustly associated genetic variants to inform approaches such as Mendelian randomisation. In addition, future work may benefit from assessing the heritability of broader depression phenotypes that lie beyond our binary criteria, including reward sensitivity and negative valence traits [80].

In summary, we find that genetic associations with MDD in UK Biobank vary by context. Specifically, the SNP-heritability of MDD is larger in individuals reporting trauma exposure compared to those not doing so. Furthermore, the genetic correlation of MDD with body composition and with education was significant only in individuals reporting exposure to trauma. Together, these findings suggest the genetic contribution to MDD is greater when additional risk factors are present.

Acknowledgements

We thank the members of the UK Biobank Mental Health Genetics Group for their valuable discussion and feedback on this work. We are also deeply indebted to the scientists involved in the construction of the UK Biobank, and to the investigators who comprise the PGC. Finally, we thank the hundreds of thousands of subjects who have shared their life experiences with investigators in the UK Biobank and the PGC.

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This research has been conducted using the UK Biobank Resource, as an approved extension to application 16577 (Dr Breen). This study represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980). WJP was funded by NWO Veni grant 91619152. K.L.P acknowledges funding from the Alexander von Humboldt Foundation. N.R.W acknowledges funding from the Australian National Health and Medical Research Council (1078901 and 1087889). The PGC has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390,** 1211–1259 (2017).
2. McManus, S., Bebbington, P., Jenkins, R. & Brugha, T. *Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014: a Survey Carried Out for NHS Digital by NatCen Social Research and the Department of Health Sciences, University of Leicester*. (NHS Digital, 2016).
3. Green, J. G. *et al.* Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication I: associations with first onset of DSM-IV disorders. *Arch. Gen. Psychiatry* **67,** 113–123 (2010).
4. Nanni, V., Uher, R. & Danese, A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am. J. Psychiatry* **169,** 141–151 (2012).
5. Kessler, R. C. The effects of stressful life events on depression. *Annu. Rev. Psychol.* **48,** 191–214 (1997).
6. McLaughlin, K. A., Conron, K. J., Koenen, K. C. & Gilman, S. E. Childhood adversity, adult stressful life events, and risk of past-year psychiatric disorder: a test of the stress sensitization hypothesis in a population-based sample of adults. *Psychol. Med.* **40,** 1647–1658 (2010).
7. Kessler, R. C., Davis, C. G. & Kendler, K. S. Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychol. Med.* **27,** 1101–1119 (1997).
8. Collishaw, S. *et al.* Resilience to adult psychopathology following childhood

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

maltreatment: evidence from a community sample. *Child Abuse Negl.* **31,** 211–229 (2007).

9. Reuben, A. *et al.* Lest we forget: comparing retrospective and prospective assessments of adverse childhood experiences in the prediction of adult health. *J. Child Psychol. Psychiatry* **57,** 1103–1112 (2016).
10. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* **156,** 837–841 (1999).
11. Kendler, K. S. & Karkowski-Shuman, L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol. Med.* **27,** 539–547 (1997).
12. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).
13. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49,** 1304–1310 (2017).
14. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48,** 1031–1036 (2016).
15. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50,** 668–681 (2018).
16. Jang, K. L., Vernon, P. A., Livesley, W. J., Stein, M. B. & Wolf, H. Intra- and extra-familial influences on alcohol and drug misuse: a twin study of gene-environment correlation. *Addiction* **96,** 1307–1318 (2001).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Stein, M. B., Jang, K. L., Taylor, S., Vernon, P. A. & Livesley, W. J. Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: a twin study. *Am. J. Psychiatry* **159,** 1675–1681 (2002).
18. Lyons, M. J. *et al.* Do genes influence exposure to trauma? A twin study of combat. *Am. J. Med. Genet.* **48,** 22–27 (1993).
19. Power, R. A. *et al.* Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol. Med.* **43,** 1965–1971 (2013).
20. Dunn, E. C. *et al.* Genetic determinants of depression: recent findings and future directions. *Harv. Rev. Psychiatry* **23,** 1–18 (2015).
21. Schraedley, P. K., Turner, R. J. & Gotlib, I. H. Stability of retrospective reports in depression: traumatic events, past depressive episodes, and parental psychopathology. *J. Health Soc. Behav.* **43,** 307–316 (2002).
22. Dunn, E. C. *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* **33,** 265–280 (2016).
23. Peterson, R. E. *et al.* Molecular Genetic Analysis Subdivided by Adversity Exposure Suggests Etiologic Heterogeneity in Major Depression. *Am. J. Psychiatry* **175,** 545–554 (2018).
24. Peyrot, W. J. *et al.* Effect of polygenic risk scores on depression in childhood trauma. *Br. J. Psychiatry* **205,** 113–119 (2014).
25. Mullins, N. *et al.* Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol. Med.* **46,** 759–770 (2016).
26. Peyrot, W. J. *et al.* Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consortium. *Biol. Psychiatry* (2017). doi:10.1016/j.biopsych.2017.09.009

27. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK biobank data: come and get it. *Sci. Transl. Med.* **6,** 224ed4 (2014).
28. Davis, K. A. S. *et al.* Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* **4,** 83–90 (2018).
29. Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8,** e75362 (2013).
30. Bellis, M. A., Hughes, K., Leckenby, N., Perkins, C. & Lowey, H. National household survey of adverse childhood experiences and their relationship with resilience to health-harming behaviors in England. *BMC Med.* **12,** 72 (2014).
31. Bernstein, D. P. *et al.* Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am. J. Psychiatry* **151,** 1132–1136 (1994).
32. Glaesmer, H., Schulz, A. & Häuser, W. The childhood trauma screener (CTS)-development and validation of cut-off-scores for classificatory diagnostics. *Psychiatr. Prax.* **40,** 220–226 (2013).
33. Townsend, P., Phillimore, P. & Beattie, A. *Health and Deprivation: Inequality and the North*. (Croom Helm, 1988).
34. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
35. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).
36. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

37. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26,** 2867–2873 (2010).

38. Warren, H. R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49,** 403–415 (2017).

39. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33,** 2776–2778 (2017).

40. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32,** 227–234 (2008).

41. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).

42. Lloyd-Jones, L. R., Robinson, M. R., Yang, J. & Visscher, P. M. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics* (2018). doi:10.1534/genetics.117.300360

43. Marioni, R. E. *et al.* GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* 246223 (2018). doi:10.1038/s41398-018-0150-6

44. Team, R. C. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. (2014).

45. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50,** 906–908 (2018).

46. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36,** 214–224 (2012).

47. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2015).

48. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
49. Daniel. *Biostatistics: A Foundation For Analysis In Health Sciences, 7Th Ed*. (Wiley India Pvt. Limited, 2006).
50. Tukey, W. J. Bias and confidence in not-quite large samples. *Ann. Math. Stat.* **29,** 614 (1958).
51. Quenouille, M. H. Notes on Bias in Estimation. *Biometrika* **43,** 353–360 (1956).
52. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).
53. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
54. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43,** 977–983 (2011).
55. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).
56. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59,** 3229–3239 (2010).
57. Keller, M. C. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75,** 18–24 (2014).
58. Yzerbyt, V. Y., Muller, D. & Judd, C. M. Adjusting researchers' approach to

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

adjustment: On the use of covariates when testing interactions. *J. Exp. Soc. Psychol.* **40,** 424–431 (2004).

59. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 118810 (2018).
60. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9,** 1470 (2018).
61. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry* **21,** 749–757 (2016).
62. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50,** 6–11 (2018).
63. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).
64. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50,** 920–927 (2018).
65. Weissbrod, O., Flint, J. & Rosset, S. Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. *Am. J. Hum. Genet.* **103,** 89–99 (2018).
66. Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E5272–81 (2014).
67. Munafo, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Smith, G. D. Collider Scope: How selection bias can induce spurious associations. *bioRxiv* 079707

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2016). doi:10.1101/079707

68. Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N. & Allen, N. E. OP41 The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. *J. Epidemiol. Community Health* **70,** A26–A26 (2016).
69. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).
70. de Wit, L. *et al.* Depression and obesity: a meta-analysis of community-based studies. *Psychiatry Res.* **178,** 230–235 (2010).
71. Luppino, F. S. *et al.* Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch. Gen. Psychiatry* **67,** 220–229 (2010).
72. Afari, N. *et al.* Depression and obesity: do shared genes explain the relationship? *Depress. Anxiety* **27,** 799–806 (2010).
73. Fuemmeler, B. F., Dedert, E., McClernon, F. J. & Beckham, J. C. Adverse childhood events are associated with obesity and disordered eating: results from a U.S. population-based survey of young adults. *J. Trauma. Stress* **22,** 329–333 (2009).
74. Metzler, M., Merrick, M. T., Klevens, J., Ports, K. A. & Ford, D. C. Adverse childhood experiences and life opportunities: Shifting the narrative. *Child. Youth Serv. Rev.* **72,** 141–149 (2017).
75. Jaffee, S. R. *et al.* Childhood Maltreatment Predicts Poor Economic and Educational Outcomes in the Transition to Adulthood. *Am. J. Public Health* **108,** 1142–1147 (2018).
76. Danese, A. & Tan, M. Childhood maltreatment and obesity: systematic review

bioRxiv preprint doi: https://doi.org/10.1101/247353; this version posted November 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and meta-analysis. *Mol. Psychiatry* **19,** 544–554 (2014).

77. Haro, J. M. *et al.* Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *Int. J. Methods Psychiatr. Res.* **15,** 167–180 (2006).
78. Kessler, R. C. *et al.* Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US national comorbidity survey (NCS). *Int. J. Methods Psychiatr. Res.* **7,** 33–55 (1998).
79. Hill, W. D. *et al.* Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Curr. Biol.* **26,** 3083–3089 (2016).
80. Nusslock, R. & Alloy, L. B. Reward processing and mood-related symptoms: An RDoC and translational neuroscience perspective. *J. Affect. Disord.* **216,** 3–16 (2017).