Genome-wide gene-environment analyses of depression and reported lifetime traumatic experiences in UK Biobank

Jonathan Coleman[1,2], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[3], UK Biobank Mental Health Consortium[3], 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 SE5 8AF; UK.

[2] NIHR Biomedical Research Centre for Mental Health; South London and Maudsley NHS Trust; London SE5 8AF; UK.

[3] Consortium members listed in Supplementary Materials

* Address correspondence to Dr Gerome Breen; Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry, Psychology & Neuroscience; King's College London; London SE5 8AF; UK. gerome.breen@kcl.ac.uk, +442078480409.

Abstract

Depression is a debilitating common mental disorder more frequently observed among individuals exposed to traumatic events. Genetic factors explain approximately one third of the risk for depression, and multiple genetic loci have been identified. The UK Biobank concurrently assessed depression and reports of trauma exposure in 157,366 individuals with pre-existing genome-wide association study (GWAS) data. Using this dataset, we compared SNP-heritability, genetic correlations and polygenic risk scores across cases and controls, stratified by reported trauma exposure. Analyses used unrelated European ancestry individuals with high-quality genotype data (final N range: 24,094-98,720). Genetic correlations with depression were substantial between participants reporting trauma exposure and in unexposed individuals. However, estimates of heritability from common variants were greater in those reporting trauma exposure (SNP $h^2_{liability}$ = 24%) than in unexposed cases (SNP $h^2_{liability}$ = 10-14%). Genetic correlations between depression and psychiatric traits were strong regardless of reported trauma exposure, whereas body mass index (and related phenotypes) was genetically correlated with depression status in trauma exposed (rather than unexposed) individuals. As such, in those reporting trauma exposure, the common genetic component of depression also reflects phenotypic correlates of depression such as body mass index. Our findings suggest that the environment acts on genetic risk for depression to increase the SNP-heritability of depression. Further, the homogeneity of genetic correlations in trauma unexposed depression (with psychiatric traits) and lack of correlation with BMI echoes earlier ideas of endogenous depression.

Introduction

Depression is among the most common mental illnesses worldwide, affecting approximately 10-20% of individuals at some point in their lifetime [1,2] and accounting for 5.5% of all years lost through disability globally [3]. Both environmental and genetic factors influence depression. In particular, depression is more commonly observed among individuals exposed to stressful life events and early-life traumas [4,5]. However, depression has also been shown to correlate with an increased *subsequent* incidence of exposure to trauma, both dependent traumatic events, which are partly influenced by an individual, as well as independent traumatic events. This suggests that a bidirectional causal relationship between depression and traumatic events may exist [6].

Evidence from twin and family quantitative genetic studies suggests the depression is moderately heritable, with 30-40% of the phenotypic variance of depression attributable to genetic factors [7]. The progress of genome-wide association studies (GWAS) in the last decade has enabled the estimation of the proportion of the heritability captured by common genetic variants (with allele frequency >1% in the population), also known as the single nucleotide polymorphism or SNP-heritability. Typically, such approaches estimate lower heritabilities than can be obtained from twin approaches, both because the causal variants are not usually assayed themselves (meaning their effect is imperfectly captured by nearby correlated variants) and because the effects of rare genetic variation may not be captured [8]. In depression, SNP-heritability estimates are approximately 9-10% from the largest GWASs to date [9].

The most recent analysis of major depression from the Psychiatric Genomics Consortium combined multiple cohorts, including from 23andme, to identify 44 loci [9].

GWAS results strongly suggest depression is a complex polygenic disorder, meaning that it has a common variant component comprising potentially thousands of variants, each with very small individual effects.

Combining measures of trauma exposure and genetics in the study of depression may lead to faster, and more clinically relevant, progress in depression genetics, but has been limited by the difficulties of acquiring both of these data types in large cohorts [10]. Assessing trauma exposure, as with many environmental measures, requires careful phenotyping. However, there is often considerable heterogeneity between the instruments used in individual studies, which makes it difficult to pool studies for meta-analysis. Individual study cohorts tend to be too small for GWAS analyses. The recent release of mental health questionnaire data from the UK Biobank resource provides an opportunity to overcome these difficulties and assess these effects.

The UK Biobank is an epidemiological resource assessing a range of health-related phenotypes in approximately 500,000 British individuals aged between 40 and 70 [11]. In addition to a broad range of health phenotypes, genome-wide genotype data is available on all participants. This includes 157,366 participants who completed an online follow-up questionnaire assessing common mental health disorders, including depression, and a 26-item questionnaire component addressing traumatic events [Davis et al, Under Review - included as a supplementary file for review purposes]. The availability of psychiatric, genetic and trauma exposure data at this scale enables the assessment of the effects of these factors separately and in combination, either through examining statistical interactions between genetics and trauma exposure, or through using data on trauma exposure to stratify genetic analyses of depression. Adopting the latter approach, we performed genome-wide

association studies and estimated the SNP-heritability of depression with and without reported lifetime trauma exposure in UK Biobank. We used GWAS results to estimate genetic correlations and polygenic risk scores across strata, comparing patterns of genetic correlation across a wide range of external traits.

Methods

Genetic data

Genetic data for analyses was obtained from the full release of the UK Biobank data (N=487,410; for details see [12]). Briefly, autosomal genotype data from two highly-overlapping custom genotyping arrays (covering ~800,000 markers) underwent centralised quality control to remove genotyping errors 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 [12,15,16]. In addition to this central quality control, variants for analysis were limited to common variants (minor allele frequency > 0.01) imputed with higher confidence (IMPUTE INFO metric > 0.4), and which were either directly genotyped or were imputed from the HRC [15].

Using the genotyped SNPs, individuals were removed if: recommended by the UK Biobank core analysis team for unusual levels of missingness or heterozygosity; SNP genotype call rate < 98%; related to another individual in the dataset (KING r < 0.044, equivalent to removing up to third-degree relatives inclusive); phenotypic and genotypic gender information was discordant (X-chromosome homozygosity (F_X) < 0.9 for phenotypic males, F_X > 0.5 for phenotypic females). Removal of relatives was performed using a greedy algorithm, which minimise exclusions (for example, by excluding the child in a mother-father-child trio). All analyses were limited to

individuals of White Western European ancestry, as defined by 4-means clustering on the first two genetic principal components provided by the UK Biobank [13]. Principal components analysis was also performed on the European-only subset of the data using the software flashpca2 [14]. After quality control, individuals were excluded from analysis if they did not complete the mental health online questionnaire (N=126,522).

Polygenic risk score analyses used the genotyped variants [12]. Variants for this analysis 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.

Phenotype definitions

Phenotypes were derived from an online mental health questionnaire that UK Biobank participants were invited to complete (Resource 22 on http://biobank.ctsu.ox.ac.uk), and built on those defined in a recent publication describing the phenotypic structure of this resource [Davis et al, Under Review]. Individuals with depression were defined as those meeting lifetime criteria based on questions derived from the Composite International Diagnostic Interview (CIDI; Supplementary Table 1). Depression cases were excluded if they self-reported previous diagnoses of schizophrenia (or other psychoses) or bipolar disorder. Depression 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 [17] (Supplementary Table 1).

Reported trauma exposure was defined with the aim of obtaining a single binary variable for stratification that reflected an overall exposure to a severe and potentially depressogenic environment. Participants were asked questions relating to traumatic experiences in childhood (derived from the Childhood Trauma Screener, a shortened version of the Childhood Trauma Questionnaire [18–20]) and in adulthood (developed by the UK Biobank Mental Health steering group [Davies et al, Under Review]) to mirror the childhood items as well as common triggers of post-traumatic stress-disorder. Responses to individual questions were dichotomised (Supplementary Table 1) and assessed for enrichment in depression (Supplementary Table 2). Reported events with an odds ratio > 2.5 of depression were selected, to capture exposure to the traumas most associated with depression. The enriched items were drawn from questions concerning childhood (felt loved, felt hated by a family member, sexually abused), adulthood (physical violence, belittlement, sexual interference) and traumatic events (ever a victim of sexual assault). In order to capture increased severity of exposure, only individuals reporting two or more enriched items were treated as reporting trauma exposure, while those reporting none of the items were treated as unexposed. Individuals reporting a single trauma item, or who did not provide an answer were excluded from analyses (Supplementary Table 1).

Phenotype preparation for analyses

Three sets of analyses were performed comparing all depression cases and controls (referred to as [D+ vs D-]), depression cases and controls reporting trauma exposure [D+T+ vs D-T+] and depression cases and controls not reporting trauma exposure [D+T- vs D-T-] (Table 1). In addition, sensitivity analyses were performed

on reported trauma exposure (overall and stratified by depression diagnosis) and additionally stratifying analyses by sex (Supplementary Materials; Supplementary Table 3). For each analysis, phenotypes were residualised on 6 principal components from the genetic data of the European samples and factors capturing initial assessment centre and genotyping batch. Deviance residuals were obtained from generalised linear models with binomial link functions in R.3.4.1 [21].

Phenotype distribution

Participants were compared across a number of standard demographic variables and common correlates of depression - sex, age (at questionnaire), education (as % reporting receiving a degree), neighbourhood socioeconomic status (SES, as Townsend deprivation index [22]) and BMI (recorded from measurements taken at the initial recruitment of the participants into the biobank). For dichotomous variables (sex and education), comparisons were made using chi-square tests. For approximately continuous variables (age, SES and BMI), the skewness and kurtosis of the distribution was checked, and roughly normal variables (absolute values of skewness (as b_1) and kurtosis (as b_2) <= 2; [23]) were compared using Welch's t-tests. Non-normal continuous variables were compared using Mann-Whitney U tests. All comparisons were performed in R.3.4.1, using skewness and kurtosis calculations from the e1071 package [21,23,24]. An additional breakdown of individual trauma items by sex was performed.

Analyses

Genome Wide Association Studies (GWASs)

GWASs were performed using linear regressions on imputed genotype dosages in BGenie v1.2, software written for genetic analyses of UK Biobank [12], with residualised phenotypes as described above. Results from each analysis were clumped to define genetic loci in PLINK2 [25]. Loci were defined following established protocols [9]. Each locus comprised all variants with $p < 0.0001$ in linkage disequilibrium ($r^2 > 0.1$ in European subjects from the 1000 Genomes Phase 3 release [26]) with a nearby (< 3Mb) variant with a lower p-value. Neighbouring (< 50kb) or overlapping clumps were merged using bedtools [27]. Loci were annotated using RegionAnnotator v1.63 (https://github.com/ivankosmos/RegionAnnotator) to identify proximal (< 100kb from loci boundaries) features of interest using data from the EBI GWAS catalog; OMIM; GENCODE genes; genes previously implicated in autism and/or intellectual disability; copy-number variants previously implicated in psychiatric disorders; and mouse knockout phenotypes. Genome-wide significance was defined at the conventional level $p < 5 \times 10^{-8}$ [28].

Heritability and genetic correlations

SNP-heritability was calculated using BOLT-LMM v2.3 [29]. Trauma exposure may be under-reported in the general population, which impairs accurate estimation of the population prevalence [30–33]. This affects the depression phenotypes stratified by reported trauma exposure. Accordingly, estimates of SNP-heritability were calculated on the observed scale, and converted to the liability scale assuming a range of population prevalences +/- 10% of the observed sample prevalence. This was performed for all phenotypes for consistency.

Genetic correlations (r_g) between the stratified GWAS results were calculated in LD Score v1.0.0 [34] using the default HapMap LD reference. Genetic correlations

with external phenotypes were calculated using an online extension of LD Score, LD Hub v1.4.1, including all phenotypes in the LD Hub database [34–36]. Where multiple GWAS existed for a given trait, the most recent was retained. More recent GWAS of BMI [37], major depression [9] and heel bone mineral density [38] than those used in LD Hub were available. Genetic correlations with these more recent GWAS (removing the MHC region as in LD Hub) were calculated in LD Score to replace their equivalents from LD Hub.

Genetic correlations were tested for difference from 0 (default in LD Score/LD Hub), and for difference from 1 (in Microsoft Excel, converting r_g to a chi-square as $[(r_g -1)/se]*[(r_g -1)/se]$). Genetic correlations with external phenotypes were considered significant at $p < 0.0002$ (219 external phenotypes tested). Comparisons of genetic correlations with the same external phenotype between stratified analyses were tested post-hoc using two sample z-tests. As these tests were principally used to determine whether a given significant genetic correlation differed in magnitude between strata, no correction was made for multiple testing in this instance.

Polygenic Risk Scoring

Polygenic risk scores were calculated using PRSice v2 (https://github.com/choishingwan/PRSice [25,39]). Specifically, risk scores from analyses of major depression [9], schizophrenia [40], bipolar disorder [41], and body mass index [37] - all frequent comorbidities of depression - were calculated and compared across strata using linear regression implemented in PRSice. Risk scores 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 genetic risk scores and the phenotype. The major depression GWAS used to derive risk scores contained

participants from UK Biobank, so a restricted set of summary statistics without these individuals (but including individuals from 23andMe [42]) were used. Risk scores from an analysis of glycated haemoglobin (HbA1c [43]) were used as a negative control. In total, five external phenotypes were used to produce risk scores for three analyses, resulting in 15 analyses. A conservative Bonferroni correction for multiple testing was used, correcting for 105 tests (seven thresholds and 15 analyses), giving a final threshold for significance of $p < 0.0004$.

Results

Phenotype distribution

Phenotypic and genetic data were available on 24,094 (depression in individuals reporting trauma exposure [D+T+ vs D-T+]) to 92,957 individuals (depression in all individuals [D+ vs D-]; Table 1). Overall, 36% of individuals were classified as reporting trauma exposure, more frequently in depression cases (45%) than in controls (17%, $p < 10^{-50}$, chi-square test). Depression cases differed significantly from controls, both overall and when taking reported trauma exposure into account. Individuals with depression were mostly females, significantly younger, less likely to have a degree, came from more deprived neighbourhoods, and had higher BMI at recruitment (all $p \leq 0.001$; Supplementary Table 4).

		Reported trauma exposure			
		Exposed [T+]	Unexposed [T-]	Excluded	Total
Depression	Cases [D+]	**13,393[b]**	**9,487[c]**	6,595	**29,475[a]**
	Controls [D-]	**10,701[b]**	**39,677[c]**	13,104	**63,482[a]**
	Excluded	11,175	14,287	8,103	33,565
	Total	35,269	63,451	27,802	126,522

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) depression in all participants [D+ vs D-] (29,475 cases, 63,482 controls)
b) depression in trauma exposed participants [D+T+ vs D-T+] (13,393 cases, 10,701 controls)
c) depression in trauma unexposed participants [D+T- vs D-T-] (9,487 cases, 39,677 controls)

Genome-wide association studies

Genome-wide association studies identified one locus in the analysis of depression [D+ vs D-] ($p < 5x10^{-8}$; Table 2, see further results in 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 heritability estimation overlapped 1; Supplementary Table 7). The locus has been repeatedly associated with depression [9,44,45], is intergenic and not annotated to any biological feature of interest (Supplementary Table 8).

Chr	Locus BP	Index SNP	BP	A0	A1	Freq A1	B A1	SE A1	p	Genes +/- 100kb	Previous GWAS hits
9	11113891 - 11883299	rs11515172	11256041	C	T	0.216	-0.0337	0.0061	$3.82x10^{-8}$	-	Neuroticism, depression

Table 2: Genome-wide significant locus from depression GWAS [D+ vs D-].
FreqA1 = A1 frequency in non-Finnish samples from gnomAD [46]
BP = Base position of index SNP in hg19

Heritability

The estimated heritability of depression was significantly greater in individuals reporting trauma exposure [D+T+ vs D-T+] than in unexposed individuals [D+T- vs D-T-] when considering estimated population prevalences +/- 10% from the observed sample prevalence ($p = 0.00381$ from a z-test comparing minimum estimate of heritability of depression in individuals reporting trauma exposure versus maximum estimate in trauma unexposed individuals; Table 3; Supplementary Table 7).

Analysis	Sample Prevalence	N (+ // -)	Heritability calculated with BOLT-LMM			
			Observed Scale h^2	95% CI	Liability Scale h^2	95% CI
Depression [D+ vs D-]	31.7%	29,475 // 63,482	12%	11-13%	18-21%	17-23%
Depression in trauma exposed [D+T+ vs D-T+]	45.6%	13,393 // 10,701	15%	11-20%	24%	18-30%
Depression in trauma unexposed [D+T- vs D-T-]	19.3%	9,487 // 39,677	6%	4-8%	10-14%	8-17%

Table 3: Heritability estimates from BOLT-LMM. Liability-scale estimates are reported for population prevalences +/- 10% of sample prevalence.
95% CI = confidence interval (95%) for heritability estimates - for liability scale, bounds of the estimate are -/+ 1.96 SE from lowest and highest heritability estimates respectively (see also Supplementary Table 7).

Genetic correlations

The genetic correlation between depression in individuals reporting trauma exposure [D+T+ vs D-T+] and depression in unexposed individuals [D+T- vs D-T-] was high and did not differ significantly from 1 ($r_g = 0.766$ [95% CI: 0.478-1.05]; difference from 0: $p = 1.82 \times 10^{-7}$; difference from 1: $p = 0.112$). Genetic correlations were also calculated between all internal phenotypes (Supplementary Table 10).

In contrast, patterns of genetic correlations with external phenotypes differed between analyses (Figure 1; Supplementary Table 11). Depression in individuals reporting trauma exposure [D+T+ vs D-T+] showed significant ($p < 0.0002$) genetic correlations with a broad range of phenotypes previously shown to be genetically correlated with depression [9], including reproductive traits (age of first birth), educational achievement (years of schooling), and body composition (body mass index, obesity, triglyceride levels, waist circumference and waist-to-hip ratio), as well as with psychiatric traits (neuroticism, major depressive disorder, PGC cross-disorder analysis, schizophrenia, subjective well-being and insomnia). In contrast, there was a narrower range of significant genetic correlations with depression in unexposed individuals [D+T- vs D-T-], predominantly with psychiatric traits (neuroticism, bipolar disorder, major depressive disorder, PGC cross-disorder analysis, schizophrenia, subjective well-being and insomnia) - the sole exception to this pattern was a significant negative correlation with femoral neck bone mineral density.

The pattern of traits significantly genetically correlated with depression in those reporting trauma exposure [D+T+ vs D-T+] compared to the unexposed group [D+T- vs D-T-] can be broken into two sets. Associations with psychiatric traits were significant in both groups, and did not differ substantially in magnitude between the groups (z-test for comparison of r_g ranged from $p = 0.146$ - 0.949 for traits significantly correlated with depression in at least one analysis; Figure 1b). In contrast, the associations with body composition phenotypes were only significant in individuals reporting exposure to trauma [D+T+ vs D-T+], not in unexposed individuals [D+T- vs D-T-], and were also significantly larger (Figure 1a, Supplementary Table 11). For example, the genetic correlation between body mass

index and depression in individuals reporting trauma exposure [D+T+ vs D-T+] was 0.213 (95% CI: 0.132-0.293, p = 2.14×10^{-7}) compared with a correlation of -0.0231 (95% CI: -0.138 - 0.092, p = 0.693) with depression in unexposed individuals [D+T- vs D-T-] (p = 9.85×10^{-4}, z-test).

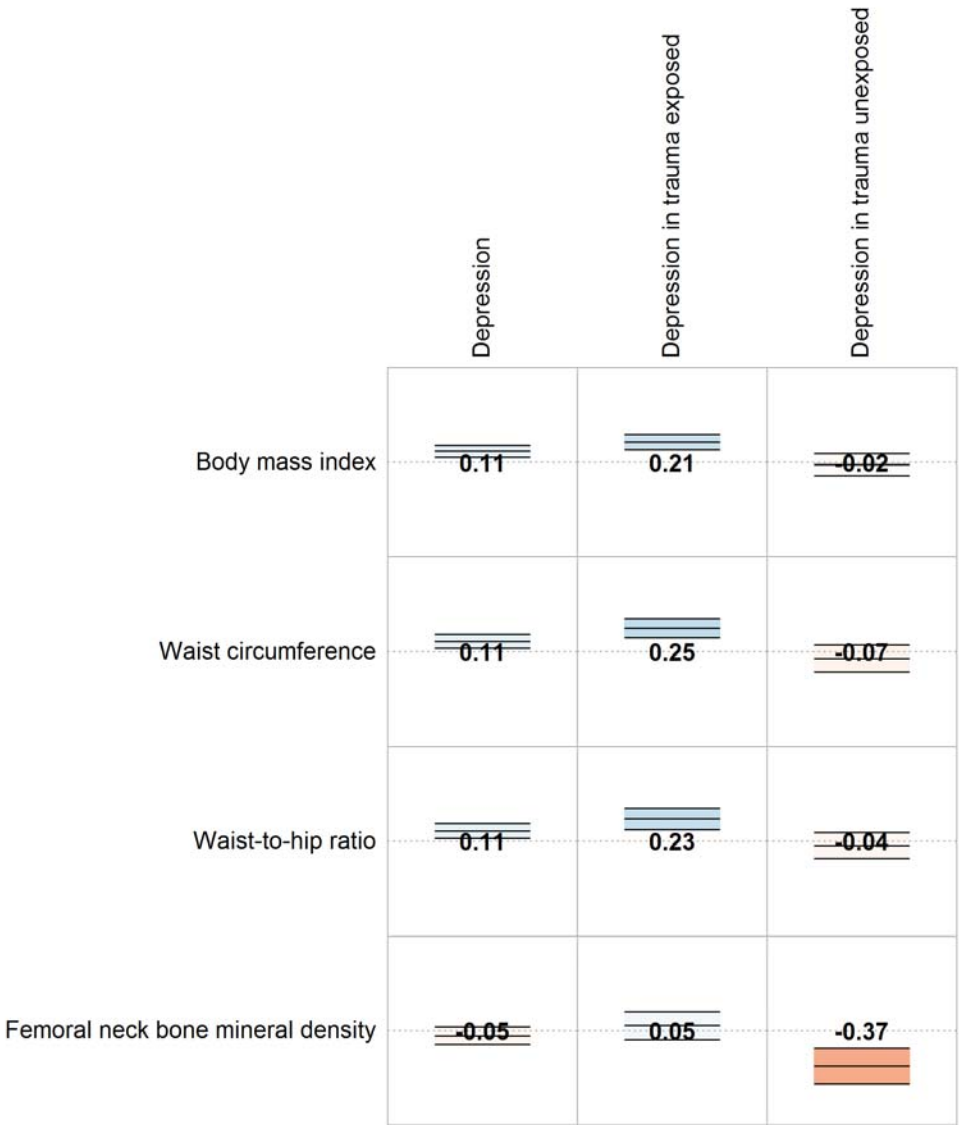

Figure 1a: Genetic correlations between body composition traits and depression, differing by reported trauma exposure (middle and right columns). Numbers = genetic correlations. Colour = direction of effect (blue = positive, red = negative). Colour intensity = size of correlation. Upper and lower bars are 95% confidence interval of genetic correlation.

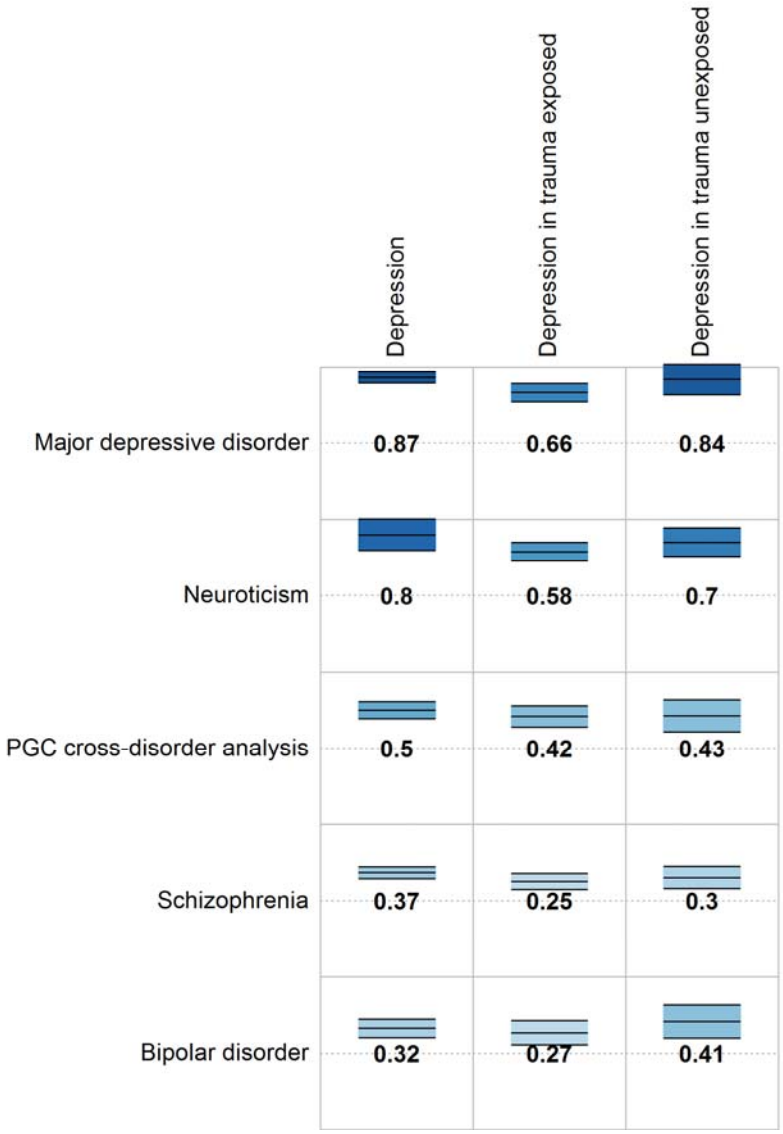

Figure 1b: Overlapping genetic correlations between depression and psychiatric disorders. Numbers = genetic correlations. Colour = direction of effect (blue = positive, red = negative). Colour intensity = size of correlation. Upper and lower bars are 95% confidence interval of genetic correlation.

Polygenic risk scores across strata

Genetic risk scores for major depression were higher in disorder cases than controls overall [D+ vs D-], when considering only individuals reporting trauma exposure [D+T+ vs D-T+], and when considering only individuals not reporting trauma exposure [D+T- vs D-T-] (Table 4; Supplementary Table 12). A similar pattern was observed with genetic risk scores for schizophrenia (Table 4;

Supplementary Table 12). For bipolar disorder risk scores, greater risk scores were observed in cases overall [D+ vs D-] and when considering unexposed individuals [D+T- vs D-T-], but when considering individuals reporting trauma exposure [D+T+ vs D-T+] (Table 4, Supplementary Table 12). Conversely, genetic risk scores for BMI were higher in depression cases than controls overall [D+ vs D-], and in individuals reporting trauma exposure [D+T+ vs D-T+], but did not differ significantly between cases and controls not reporting trauma exposure [D+T- vs D-T-] (Table 4; Supplementary Table 12). No significant differences were observed in the negative control analysis with HB1Ac (Table 4; Supplementary Table 12).

Analysis	PRS	Best Threshold	Nagelkerke's pseudo R^2	t	p
Depression [D+ vs D-]	**MDD**	**0.4**	**0.00848**	**28.1**	**$< 10^{-50}$**
	SCZ	**0.1**	**0.00191**	**13.4**	**1.32×10^{-40}**
	BIP	**0.2**	**0.000834**	**8.8**	**1.29×10^{-18}**
	BMI	**0.3**	**0.000354**	**5.7**	**9.70×10^{-9}**
	HB1Ac	0.001	0.0000217	1.4	0.155
Depression in trauma exposed [D+T+ vs D-T+]	**MDD**	**0.4**	**0.00810**	**14.0**	**1.16×10^{-44}**
	SCZ	**0.5**	**0.000717**	**4.2**	**3.24×10^{-5}**
	BIP	0.5	0.000358	2.9	0.00331
	BMI	**0.5**	**0.00113**	**5.2**	**1.71×10^{-7}**
	HB1Ac	0.001	0.000122	1.7	0.0860
Depression in trauma unexposed [D+T- vs D-T-]	**MDD**	**0.4**	**0.00396**	**14.0**	**2.42×10^{-44}**
	SCZ	**0.5**	**0.000908**	**6.7**	**2.37×10^{-11}**
	BIP	**0.2**	**0.000591**	**5.4**	**6.94×10^{-8}**
	BMI	0.3	0.0000578	1.7	0.0919
	HB1Ac	0.001	0.0000102	0.7	0.479

Table 4: Polygenic risk scores (PRS) associated with depression and reported trauma exposure overall and in stratified analyses. Bold = associations significant at $p < 0.000143$. Positive t indicates higher risk scores in the case group / those reporting exposure. Results are reported at the threshold with the lowest p-value - results across all thresholds are reported in Supplementary Table 12.

Sensitivity analyses focussed on reported trauma exposure

Results from parallel analyses focussed on reported trauma exposure (overall and stratified by depression diagnosis) were broadly similar to the results from analyses of depression (Supplementary Material). The pattern of phenotypic correlations observed between depression cases and controls was also observed for individuals reporting trauma exposure compared to those not reporting trauma exposure, both overall and when accounting for depression. The only exception to this was that individuals reporting trauma were more likely than unexposed individuals to have a degree (in contrast to depression cases, who were less likely than controls to have a degree; all $p < 0.001$).

Genome-wide association analyses identified six significant loci when comparing individuals reporting trauma exposure to those unexposed. These loci have previously been implicated in genetic studies of a variety of phenotypes, including attention deficit disorder, schizophrenia and educational attainment, but did not overlap with the locus identified in the depression analyses. The range of heritability estimates from analyses focussed on reported trauma exposure (16-24%) was similar to that from analyses of depression (10-24%), but the heritability of reported trauma exposure did not differ substantially when comparing depression cases to controls.

The pattern of external genetic correlations observed in depression, with a broader set of correlations with depression in individuals reporting trauma exposure, was mirrored in reported trauma exposure, where a broader set of correlations with reported trauma exposure was observed in cases compared to controls. Finally, the association between reported trauma exposure and polygenic risk scores was

largely similar to the results seen in the analyses with depression, with equivalent results from overall and stratified analyses with all risk scores.

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

Discussion

Our results show that depression is more heritable in individuals reporting trauma exposure than in those not reporting trauma exposure, and furthermore that the pattern of significant genetic correlations with depression is more diverse in this group than in unexposed individuals. This evokes past concepts of depression: the broader genetics of depression in individuals reporting trauma exposure mirrors reactive depression, while the narrower genetics of depression in unexposed individuals, with genetic correlations specifically with psychiatric phenotypes, is more akin to endogenous depression [47]. Our results also agree with the diathesis-stress model of psychiatric disorders, in that exposure to trauma appears to activate an underlying genetic liability as shown by the difference in SNP-heritability and polygenic risk scores for depression cases between the exposed and unexposed groups [48,49].

We examined individual genetic variants, variant by trauma interactions, genome-wide common variant heritability, genetic correlations, and polygenic risk scores. The estimated heritability of depression was significantly increased in individuals reporting trauma exposure [D+T+ vs D-T+] compared to unexposed individuals [D+T- vs D-T-]. There was a high genetic correlation between depression in individuals reporting trauma exposure [D+T+ vs D-T+] and individuals not reporting trauma exposure [D+T- vs D-T-], implying considerable genetic continuity.

This was supported by significant genetic correlations between depression and other psychiatric disorders when considering individuals reporting trauma exposure [D+T+ vs D-T+] and when considering unexposed individuals [D+T- vs D-T-]. However, depression in individuals reporting trauma exposure [D+T+ vs D-T+] showed significant genetic correlations with phenotypes known to be associated with depression, particularly anthropometric traits like body mass index [50–52].

Our results suggest that the genetic component of depression is composed of multiple overlapping components, including general and depression-specific psychiatric risk variants, but also including risk variants associated with anthropometric traits, such as BMI, and other traits thought to be independent risk factors for depression [1,53]. We find that these latter components are more prominent in individuals reporting trauma exposure. Another possible explanation of the increased heritability of depression in individuals reporting trauma exposure is that it gains additional, apparent heritability from the genetic variants associated with its risk factors, such as increased BMI [9,54].

A difference in the genetic correlation between BMI and depression was observed when stratifying by trauma, such that a significant positive genetic correlation was only seen in individuals reporting trauma exposure. Increased BMI is a considerable and common risk factor for increased, lifelong social adversity. The direction of association is unclear. Trauma in childhood has been reported to influence adult BMI [55]. However, in adults, Mendelian randomisation analyses suggest increased BMI may cause lifelong lower earnings on average, particularly in women [56]. Furthermore, in the recent PGC depression analyses [9], Mendelian randomisation showed a putative causal path from increased genetic risk of BMI to a higher risk of depression. As such, there is evidence for a bidirectional relationship

between BMI and social adversity (including trauma exposure), which is likely to influence the observed genetic correlation between BMI and depression in individuals reporting trauma exposure.

Sensitivity analyses focussed on trauma found that self-reported traumatic experience was significantly heritable, as has been previously observed [57]. We strongly emphasise that this does not imply that traumatic experiences themselves have a biological component - it is more likely that such experiences are associated with other significantly heritable traits, and that their biology is then reflected in the observed heritability of trauma exposure. This also does not mean that these individuals invite such experiences, but that their genetic predispositions may place them at a greater risk of experiencing or reporting traumatic experiences, as suggested by their observed genetic correlations with psychiatric traits and with BMI. A similar phenomenon of probable correlation with an unknown trait with a genetic component has been proposed to underlie observed genetic correlations with socioeconomic status [58].

The gene-environment correlation implied by the heritability of reported traumatic experience complicates the interpretation of the relationship between depression and trauma. This is notable when assessing genetic correlations with external phenotypes: the genetic correlation of reported trauma exposure with risk for various psychiatric disorders is also present in case-only [D+T+ vs D+T-] and control-only [D-T+ vs D-T-] analyses, and cannot be explained by the enrichment of reported traumatic experiences in depression cases alone. This suggests a number of potential explanations. There may be a true genetic correlation, but it may be more likely that many of the cases used in large studies of psychiatric illnesses are themselves trauma exposed, leading to a detectable correlation. This has parallels

with the positive genetic correlation seen between cannabis use and schizophrenia - one proposed explanation for this correlation is an excess of cannabis users in the cases (compared to the controls) used in the schizophrenia GWAS [59].

Our study has a number of limitations. Our trauma exposure measure relies on retrospective self-report, which is an imperfect measure of recorded trauma exposure, correlated with personality traits and mood at time of report [60]. This may explain the genetic correlations we observe with reported trauma exposure (including in controls, who do not report previous psychiatric illness). However, although retrospective self-report is not the ideal measure for this phenotype, it is the only option in the case of the UK Biobank, in which participants were recruited later in life. Also, the requirement for cohort sizes large enough to identify the small individual genetic effects typical of complex genetic traits such as depression makes self-report the most practicable method of data-collection.

The strength of the UK Biobank in enabling the integration of genetic and environmental data at scale in a reasonably homogeneous cohort is apparent. It is important to note that the UK Biobank cohort demonstrates some evidence of a "healthy volunteer bias", whereby the self-selected participants tend to be generally of better overall health and higher socioeconomic status than the equivalent overall population of this age [61]. Similarly, the White Western European ancestry group is predominant in the UK Biobank, and we focussed on these individuals; results (including genetic results) in one ancestry group may not generalise to others [62].

In addition, although genetic findings are robust to reverse causation, the trauma and depression data were reported at the same time, and so our results cannot strongly inform any temporal or causal hypotheses about the relationship between trauma and depression. 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 findings to inform approaches such as Mendelian randomisation.

In summary, we demonstrate the utility of biobank-level data for studying the complex relationship between genes and environments and their effects on complex traits. In particular, we have identified an intrinsic difference in the common genetic basis of depression when stratifying by reported exposure to trauma. Genetic correlates of depression may partially reflect specific depression risk factors that influence correlates of depression, such as BMI, and this is more apparent in individuals reporting exposure to trauma than in those unexposed. As our understanding of the genetics of depression increases, the nature and relevance of these findings may prove useful in informing clinical practise and our understanding of depression aetiology.

Acknowledgements

We thank the participants in the UK Biobank and the scientists involved in the construction of this resource. We also thank the members of the UK Biobank Mental Health Genetics Group for their valuable discussion and feedback on this work.

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. High

performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

References

1. Bromet, E. *et al.* Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med.* **9,** 90 (2011).
2. Surtees, P. G. *et al.* Depression and ischemic heart disease mortality: evidence from the EPIC-Norfolk United Kingdom prospective cohort study. *Am. J. Psychiatry* **165,** 515–523 (2008).
3. 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).
4. 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).
5. 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).
6. 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).
7. 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).
8. 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).

9. Major Depressive Disorder Working Group of the PGC. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv* 167577 (2017). doi:10.1101/167577
10. Dunn, E. C. *et al.* Genetic determinants of depression: recent findings and future directions. *Harv. Rev. Psychiatry* **23,** 1–18 (2015).
11. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK biobank data: come and get it. *Sci. Transl. Med.* **6,** 224ed4 (2014).
12. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
13. 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).
14. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33,** 2776–2778 (2017).
15. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).
16. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
17. 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).
18. 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).

19. 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).
20. 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).
21. Team, R. C. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. (2014).
22. Townsend, P., Phillimore, P. & Beattie, A. *Health and Deprivation: Inequality and the North*. (Croom Helm, 1988).
23. Joanes, D. N. & Gill, C. A. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47,** 183–189 (1998).
24. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. (2017).
25. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
26. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526,** 75–81 (2015).
27. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* **47,** 11.12.1–34 (2014).
28. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32,** 227–234 (2008).
29. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed model association for biobank-scale data sets. *bioRxiv* 194944 (2017). doi:10.1101/194944

30. Briere, J. & Elliott, D. M. Prevalence and psychological sequelae of self-reported childhood physical and sexual abuse in a general population sample of men and women. *Child Abuse Negl.* **27,** 1205–1222 (2003).
31. Molnar, B. E., Buka, S. L. & Kessler, R. C. Child sexual abuse and subsequent psychopathology: results from the National Comorbidity Survey. *Am. J. Public Health* **91,** 753–760 (2001).
32. Morgan, C. & Fisher, H. Environment and schizophrenia: environmental factors in schizophrenia: childhood trauma--a critical review. *Schizophr. Bull.* **33,** 3–10 (2007).
33. Stoltenborgh, M., van Ijzendoorn, M. H., Euser, E. M. & Bakermans-Kranenburg, M. J. A global perspective on child sexual abuse: meta-analysis of prevalence around the world. *Child Maltreat.* **16,** 79–101 (2011).
34. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
35. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33,** 272–279 (2017).
36. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
37. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).
38. Kemp, J. P. *et al.* Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat. Genet.* **49,** 1468–1475 (2017).
39. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).

40. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

41. 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).

42. 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).

43. 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).

44. 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).

45. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 118810 (2018).

46. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

47. Zimmerman, M., Coryell, W., Pfohl, B. & Stangl, D. The validity of four definitions of endogenous depression. II. Clinical, demographic, familial, and psychosocial correlates. *Arch. Gen. Psychiatry* **43,** 234–244 (1986).

48. Meehl, P. E. Schizotaxia, schizotypy, schizophrenia. *Am. Psychol.* **17,** 827–838 (1962).

49. McGuffin, P., Katz, R. & Bebbington, P. The Camberwell Collaborative Depression Study. III. Depression and adversity in the relatives of depressed probands. *Br. J. Psychiatry* **152,** 775–782 (1988).

50. de Wit, L. *et al.* Depression and obesity: a meta-analysis of community-based studies. *Psychiatry Res.* **178,** 230–235 (2010).
51. 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).
52. Afari, N. *et al.* Depression and obesity: do shared genes explain the relationship? *Depress. Anxiety* **27,** 799–806 (2010).
53. Rudisch, B. & Nemeroff, C. B. Epidemiology of comorbid coronary artery disease and depression. *Biol. Psychiatry* **54,** 227–240 (2003).
54. Milaneschi, Y. *et al.* Polygenic dissection of major depression clinical heterogeneity. *Mol. Psychiatry* **21,** 516–522 (2016).
55. Richardson, A. S., Dietz, W. H. & Gordon-Larsen, P. The association between childhood sexual and physical abuse with incident adult severe obesity across 13 years of the National Longitudinal Study of Adolescent Health. *Pediatr. Obes.* **9,** 351–361 (2014).
56. Tyrrell, J. *et al.* Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ* **352,** i582 (2016).
57. Power, R. A. *et al.* Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol. Med.* **43,** 1965–1971 (2013).
58. Hill, W. D. *et al.* Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Curr. Biol.* **26,** 3083–3089 (2016).
59. Power, R. A. *et al.* Genetic predisposition to schizophrenia associated with increased use of cannabis. *Mol. Psychiatry* **19,** 1201–1204 (2014).
60. 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).

61. 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).

62. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).