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

Jonathan R.I. Coleman[1,2], Kirstin L. Purves[1], Katrina A.S. Davis[2,3], Christopher Rayner[1], Shing Wan Choi[1], Christopher Hübel[1,2], Héléna A. Gaspar[1,2], Carol Kan[3], Sandra Van der Auwera[4], Mark James Adams[5], Donald M. Lyall[6], Wouter J. Peyrot[7], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[8], Erin C. Dunn[9,10,11], Evangelos Vassos[1,2], Andrea Danese[1,12,13], Hans J. Grabe[4], Cathryn M. Lewis[1,2], Paul F. O'Reilly[1], Andrew M. McIntosh[5], Daniel J. Smith[6], Naomi R. Wray[14,15], Matthew Hotopf[2,3], 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] Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience; King's College London; London; SE5 8AF; UK

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

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

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

[7] Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, Netherlands

[8] Consortium members listed in Supplementary Materials

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

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

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

[12] Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience; King's College London; London; SE5 8AF; UK

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

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

[15] 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.

Abstract

Depression is more frequently observed among individuals exposed to traumatic events. The relationship between trauma exposure and depression, including the role of genetic variation, is complex and poorly understood. The UK Biobank concurrently assessed depression and reported trauma exposure in 126,522 genotyped individuals of European ancestry. We compared the shared aetiology of depression and a range of phenotypes, contrasting individuals reporting trauma exposure with those who did not (final sample size range: 24,094- 92,957). Depression was heritable in participants reporting trauma exposure and in unexposed individuals, and the genetic correlation between the groups was substantial and not significantly different from 1. Genetic correlations between depression and psychiatric traits were strong regardless of reported trauma exposure, whereas genetic correlations between depression and body mass index (and related phenotypes) were observed only in trauma exposed individuals. The narrower range of genetic correlations in trauma unexposed depression and the lack of correlation with BMI echoes earlier ideas of endogenous depression.

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 depressive symptoms during their lifetime [2]. Both environmental and genetic factors influence depression. In particular, depression is more commonly observed among individuals reporting exposure to stressful life events and early-life traumas [3–6]. The relationship between depression and reported trauma exposure is complex, with studies showing both that reported trauma exposure is associated with subsequent depression, and depression is associated with subsequent reported trauma exposure [7,8]. However, the majority of people reporting exposure to traumatic experiences do not report depression [6,9,10].

The complexity in the relationship between traumatic experiences and depression is increased further by the often unavoidable use of retrospective self-reported measures of trauma exposure. Such reports are only mildly correlated with objective measures of traumatic experiences in the same individuals; and individuals currently experiencing low mood or depression are more likely to report previous trauma exposure [11,12]. However, retrospective self-report of trauma exposure is correlated with a range of adverse life outcomes including depression, demonstrating the value of using such data for examining this relationship [11].

Twin studies show that depression is moderately heritable, with 30-40% of the variance of depression attributable to genetic factors [13]. 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. Typically, such estimates are lower than can be obtained from twin approaches, due to the incomplete capture of genetic information in GWAS data

[14]. The most recent major depression GWAS from the Psychiatric Genomics Consortium combined ~35 cohorts, representing a range from severe to milder/more briefly assessed depression including the 23andMe discovery cohort [15]. This meta-analysis identified 44 loci significantly associated with depression, and estimated a SNP-heritability of 9-10% [16]. GWAS results strongly suggest depression is a complex polygenic disorder, with potentially thousands of variants with very small individual effects contributing to risk.

Genetic studies of reported trauma exposure are substantially less common than of depression. However, these studies have demonstrated that reported trauma exposure is heritable, with twin heritability estimates of 20-50% and SNP-heritability estimates of 30% [17–20]. The heritability of reported trauma exposure most likely reflects correlations between such reports and genetically-influenced traits, rather than a direct genetic influence on exposure itself. Combining measures of trauma exposure and GWAS on a large scale is challenging [21], and assessing trauma exposure, as with many environmental measures, requires careful phenotyping. Considerable heterogeneity often exists between the instruments used in individual studies, compromising meta-analysis, while individual study cohorts tend to be too small for GWAS analyses. However, there have been exceptions to this. A genome-wide SNP-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 depression identified three variants associated with depression specifically in individuals reporting exposed to traumatic events prior to depression onset, and indicated that depression in individuals reporting trauma exposure has a stronger genetic basis than

depression in unexposed individuals [23]. Further progress has been made by combining genetic associations with depression into polygenic risk scores (PRS) and performing PRS-by-trauma interaction analyses, investigating interactions as both departures from additivity (where the combined effect of PRS and trauma differs from the sum of the individual effects) and from multiplicativity (where the combined effect of PRS and trauma differs from the product of the individual effects). Results from these have been 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 recent release of mental health questionnaire data from the UK Biobank resource provides an opportunity to assess the relationship between genetic variation, risk for depression and reported trauma exposure. 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 depression, and a 16-item questionnaire component addressing traumatic events [Davis et al, In Press - included as a supplementary file for review purposes]. The availability of genetic, psychiatric, and reported trauma exposure data at this scale allows the simultaneous evaluation of their independent and joint effects. Specifically, we performed GWAS of depression with and without reported lifetime trauma exposure in UK Biobank. We used GWAS results to estimate the genetic correlation between depression in individuals with and without reported lifetime trauma exposure, and compared patterns of genetic correlation across a wide range of physical and psychiatric traits. We also performed polygenic risk scoring, using

depression-relevant external traits, and sought to extend previous analyses of PRS-by-trauma interactions on depression.

Methods

Genetic data

Genetic data for GWAS and genetic correlation analyses came from the full release of the UK Biobank data (N=487,410; [28]). 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 [28–30]. 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 were imputed from the HRC with high confidence (IMPUTE INFO metric > 0.4) [29].

Using the genotyped SNPs, individuals were removed if: they had withdrawn consent for analysis; 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 [31]); 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 minimises 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 [32]. Principal components analysis was also performed on the European-only subset of the data using the software

flashpca2 [33]. After quality control, individuals were excluded from analysis if they did not complete the online mental health questionnaire (N=126,522).

Polygenic risk score and SNP-heritability analyses used the genotyped variants [28]. 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.

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, In Press]. 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 [34] (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 [35–37]) and in adulthood

(developed by the UK Biobank Mental Health steering group to mirror the childhood items [Davis et al, In Press]), 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 with 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 the 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 [38].

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 [Davis et al, In Press]. Accordingly, participants were compared across a number of standard demographic variables and common correlates of depression - sex, age (at questionnaire), education (university degree vs. not), neighbourhood socioeconomic status (SES, as Townsend deprivation index [39]) 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; [40]) 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 [38,40,41]. An additional breakdown of individual trauma items by sex was performed. The tetrachoric correlation between depression and reported trauma was calculated using the psych package in R 3.4.1 [38,42].

Analyses

Genome Wide Association Studies (GWAS)

GWAS were performed using linear regressions on imputed genotype dosages in BGenie v1.2 [28], with residualised phenotypes as described above.

Phenotypes and genotypes were normalised. Genome-wide significance was defined at the conventional level $p < 5 \times 10^{-8}$ [43]. Betas from the GWAS were converted to odds ratios (OR) using LMOR (http://cnsgenomics.com/shiny/LMOR/) and observed sample prevalences [44]. Standard errors were calculated from the p-value and estimated OR [45]. 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 [38].

Results from each GWAS were clumped to define genetic loci in PLINK2 [46]. Loci were defined following established protocols [16]. 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 [47]) with a nearby (< 3Mb) variant with a lower p-value. Neighbouring (< 50kb) or overlapping clumps were merged using bedtools [48]. 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 NHGRI-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.

SNP-heritability

SNP-heritability was calculated on the observed scale using BOLT-LMM v2.3 [49]. Trauma exposure may be under-reported in the general population, which thus impairs accurate estimation of the incidence of trauma exposure in the general population [50–53]. Accordingly, the conversion of SNP-heritability estimates to the liability scale in the overall analysis of depression (and sensitivity analysis on reported trauma exposure) is reported for a range of population incidences from 5%-50%. It would be informative to compare the SNP-heritability of depression between individuals reporting trauma exposure and those not doing so; however, there are a number of limitations that prevent robust conclusions from such analyses. Further discussion of these limitations is included in the Supplementary Material.

Genetic correlations

Genetic correlations (r_g) between the stratified GWAS results were calculated in LD Score v1.0.0 [54] 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 [54–56]. Where multiple GWAS existed for a given trait, the most recent was retained. More recent GWAS of BMI [57], major depression [16] and heel bone mineral density [58] 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 [59]. 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 (PRS) were calculated using PRSice v2 (https://github.com/choishingwan/PRSice [46,60]). Specifically, PRS from analyses of major depression (MDD) [16], schizophrenia (SCZ) [61], bipolar disorder (BIP) [62], and body mass index (BMI) [57] were calculated and compared in all participants and stratifying by reported trauma exposure. Analyses used logistic regression, including all covariates used in the creating 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 depression. The major depression GWAS used to derive PRS contained participants from UK Biobank, so a restricted set of summary statistics without these individuals (but including individuals from 23andMe [15]) were used. PRS from an analysis of glycated haemoglobin (HbA1c), which were not expected to be associated with depression or reported trauma exposure, were used as a negative control [63]. In total, five external phenotypes were used to produce PRS 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 $p < 0.0004$.

In addition to stratified analysis, formal PRS-by-environment analyses were performed to assess the association of depression with multiplicative and additive interactions between reported trauma exposure and each PRS capturing the most variance in the main effect analyses. Analyses included the same covariates used in

the GWAS, and all PRS-by-covariate and reported trauma exposure-by-covariate interactions [64,65]. Multiplicative interaction was tested using logistic regression [25,26]. Additive interactions were initially tested using relative excess risk due to interaction (RERI; (Knol et al. 2007)); however, the presence of reported trauma exposure-by-covariate interactions in the model made the estimates of RERI unstable and inconclusive (Supplementary Material). Accordingly, additive interactions were assessed using linear regression. Correlations were calculated between the PRS and reported trauma measure to assess the potential for gene-environment correlation to confound gene-environment interaction [26].

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%). As such there is a significant enrichment of depression cases in individuals reporting trauma exposure (OR = 5.23; $p < 10^{-50}$, chi-square test) and a phenotypic tetrachoric correlation of 0.56. 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 university 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, N = 92,957);
b) depression in trauma exposed participants
([D+T+ vs D-T+], 13,393 cases, 10,701 controls, N = 24,094);
c) depression in trauma unexposed participants
([D+T- vs D-T-], 9,487 cases, 39,677 controls, N = 49,164)

Genome-wide association studies

GWAS were performed for depression 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). Liability scale estimates of depression SNP-heritability ranged 12-22% (Supplementary Table 7). One genome-wide significant locus ($p < 5x10^{-8}$) was identified in the analysis of depression [D+ vs D-] (Table 2), and remained significant when using logistic regression ($p = 4.69 \times 10^{-8}$, OR = 0.96, SE = 0.007). This locus has been repeatedly associated with depression [16,66] and with neuroticism [67–70], is intergenic, and is not annotated to any biological feature of interest (Supplementary Table 8).

Chr	Locus BP	Index SNP	BP	A0	A1	Freq A1	OR A1	SE A1	p	Genes +/- 100kb	Previous GWAS hits
9	11113891 - 11883299	rs11515172	11256041	T	C	0.216	0.920	0.0151	3.82×10^{-8}	-	Neuroticism, depression

Table 2: Genome-wide significant locus from depression GWAS [D+ vs D-].
FreqA1 = A1 frequency in non-Finnish samples from gnomAD [55]
BP = Base position of index SNP in hg19

Genetic correlations

The phenotypic correlation between depression and reported trauma exposure was mirrored genetically, with significant r_g in the full cohort (0.62). Given that trauma items were selected for association with depression, we also calculated the genetic correlation between depression in the full cohort and reported trauma exposure in the controls (0.31; Supplementary Table 10). This correlation persisted when using an independent source of depression genetics, as reported trauma exposure was correlated with the MDD PRS (Spearman's rho = 0.0675, $p < 10^{-50}$).

Genetic correlations were calculated between all internal phenotypes (those assessed in the UK Biobank in this analysis; Supplementary Table 10). 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 = 0.77 [95% CI: 0.48-1.05]; difference from 0: $p = 1.8 \times 10^{-7}$; difference from 1: p = 0.11). In contrast, patterns of genetic correlations with external phenotypes (from previously reported GWAS summary statistics) 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 [16], 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 (Supplementary Table 11). The genetic correlation between depression and other psychiatric traits was 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 1a). 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 1b, Supplementary Table 11).

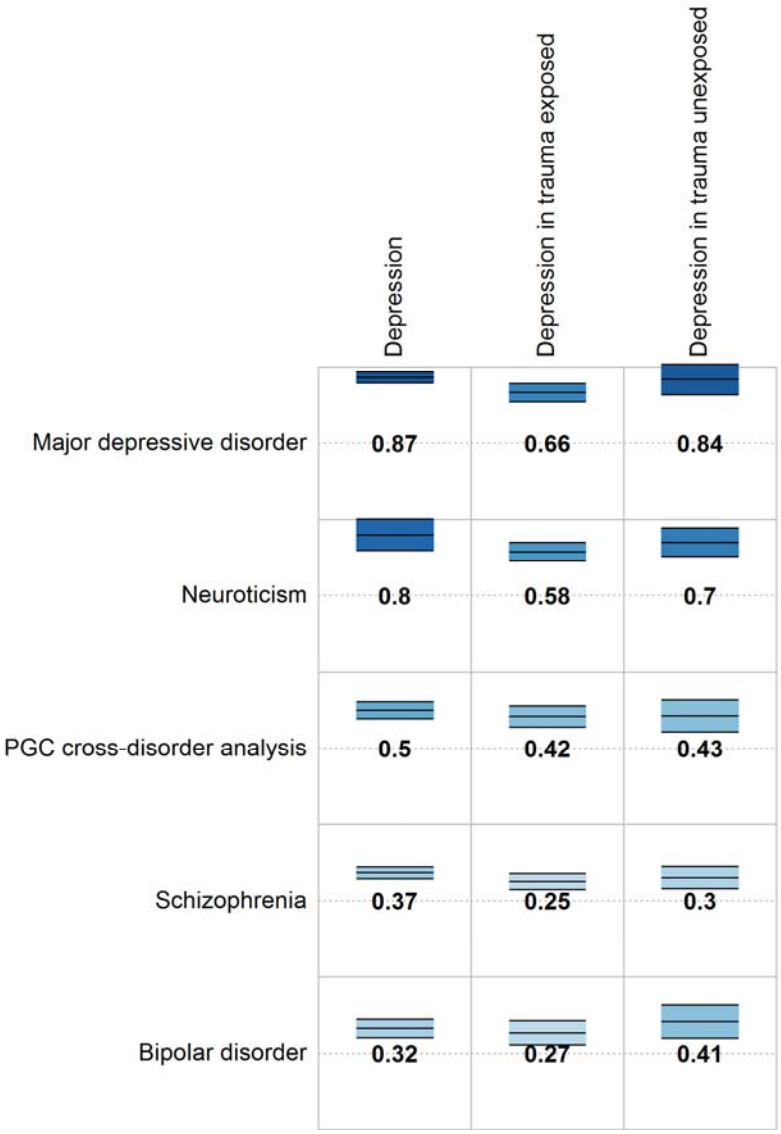


Figure 1a: Overlapping genetic correlations between depression and psychiatric disorders. Numbers = genetic correlations. Colour = direction of effect (blue = positive, red = negative). Colour intensity = size of correlation. Upper and lower bars are 95% confidence interval of genetic correlation.

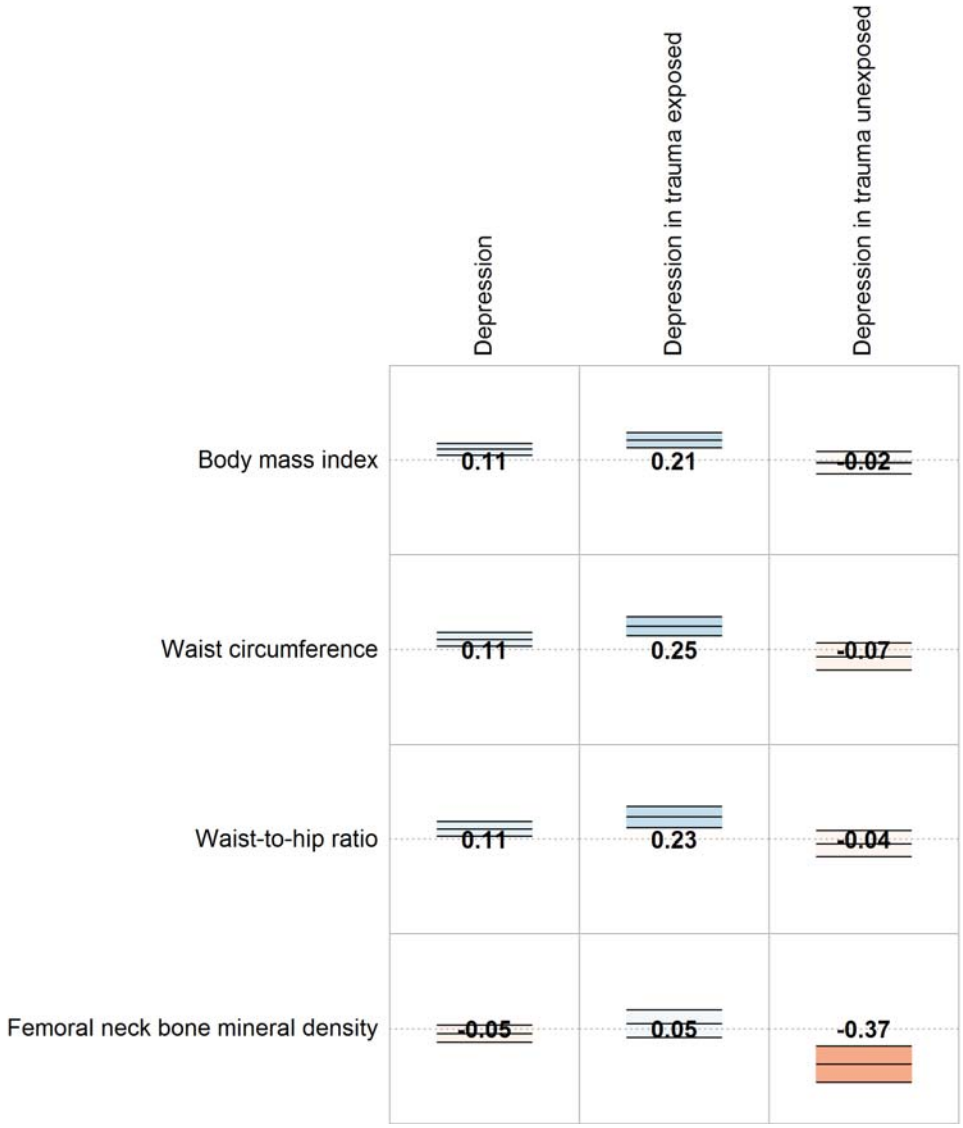


Figure 1b: Genetic correlations between body composition traits and depression, differing by reported trauma exposure (middle and right columns). Numbers = genetic correlations. Colour = direction of effect (blue = positive, red = negative). Colour intensity = size of correlation. Upper and lower bars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

Polygenic risk scores across strata

Genetic risk scores for MDD were higher in disorder cases overall and when stratifying by reported trauma exposure (Table 3; Supplementary Table 12). A similar pattern was observed with SCZ risk scores (Table 3; Supplementary Table 12). Greater BIP risk scores were observed in cases overall [D+ vs D-] and when

considering unexposed individuals [D+T- vs D-T-], but not when considering individuals reporting trauma exposure [D+T+ vs D-T+] (Table 3, Supplementary Table 12). Conversely, BMI risk scores were higher in depression cases than controls overall [D+ vs D-], and in individuals reporting trauma exposure [D+T+ vs D-T+], but not in unexposed individuals [D+T- vs D-T-] (Table 3; Supplementary Table 12). No significant differences were observed in the negative control analysis with HbA1c (Table 3; Supplementary Table 12).

Polygenic risk score by trauma exposure interactions

In the full sample, significant additive interaction terms were observed from linear regression when using the MDD 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 (Table 4). A significant multiplicative interaction term was also observed for the BMI PRS - the combined risk of depression from BMI PRS and reported trauma exposure was greater than expected from the product of the individual risks.

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
MDD	116,404 // 314,990	**0.5**	*Depression [D+ vs D-]*	**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
MDD	116,404 // 314,990	**0.4**	*Depression in trauma exposed [D+T+ vs D-T+]*	**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						
MDD	116,404 // 314,990	**0.4**	*Depression in trauma unexposed [D+T- vs D-T-]*	**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 3: Main effect and interaction effects for polygenic risk scores (PRS) associated with depression 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 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 university degree (in contrast to depression cases, who were less likely than controls to have a university degree; all $p < 0.001$).

Genome-wide association analyses identified six significant loci when comparing individuals reporting trauma exposure to those unexposed, which remained significant when assessed with logistic regression (Supplementary Figures 4-6; Supplementary Table 5). 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 liability-scale SNP-heritability estimate for reported trauma exposure (13-24%) was similar to that for depression (12-22%).

The pattern of external genetic correlations observed in analyses of depression (a broader set of correlations observed in individuals reporting trauma exposure) was mirrored in analyses of reported trauma exposure (a broader set of correlations observed in cases compared to controls; Supplementary Figures 11a and 11b). Finally, the associations between reported trauma exposure and PRS were largely similar to the results seen in the analyses with depression, with equivalent results from overall and stratified analyses with all PRS, and the same

interactions with MDD PRS (additive) and BMI PRS (additive and multiplicative) observed. Full results for sensitivity analyses are included in the Supplementary Material.

Discussion

We investigated the complex relationship between self-reported trauma exposure and depression in the largest data set available to date (N=92,957), examining individual genetic variants, genetic correlations, and polygenic risk scores. Our analyses provide evidence that the pattern of significant genetic correlations with depression is more diverse in individuals reporting trauma exposure compared to unexposed individuals. This evokes past concepts of depression, which described "reactive" depression, associated with a defined stressful event, and "endogeneous" depression, occurring in the absence of a stressful event [72]. 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 significantly greater PRS in exposed cases compared to unexposed cases (Supplementary Table 12) [73,74]. The genetic correlation between depression in individuals reporting trauma exposure and depression in individuals not reporting trauma exposure did not differ significantly from one, implying considerable genetic continuity. This was supported by significant genetic correlations between depression and other psychiatric disorders regardless of reported trauma exposure. These significant correlations were also observed in case-only and control-only analyses, and so do not result from more depression case individuals reporting trauma exposure. In individuals reporting trauma exposure, further significant genetic correlations were observed between depression and anthropometric traits like body mass index, which are known to be associated with depression [75–77]. This broader genetic basis of depression in

individuals reporting trauma exposure mirrors reactive depression whereas the narrower sets of correlations in unexposed individuals is more akin to endogenous depression [72].

Our results suggests that the common genetic component of depression is complex, including general and depression-specific psychiatric risk variants, but also risk variants associated with anthropometric traits, such as BMI [1,53]. Increased BMI is commonly associated with increased levels of social adversity but the direction of association is unclear [78,79]. In the recent PGC depression analyses [16], 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 reported trauma exposure), which is likely to influence the observed genetic correlation between BMI and depression in individuals reporting trauma exposure.

In PRS-by-reported trauma exposure interaction analyses, we identified a significant deviation from additivity for MDD PRS and reported trauma exposure on risk of depression, and a significant deviation from additivity and multiplicativity for the interaction between reported trauma exposure and BMI PRS. As such, the combined effect of psychiatric PRS and reported trauma exposure might fit a multiplicative model, consistent with the higher mean PRS observed in cases reporting trauma exposure. For BMI PRS however, the interaction with reported trauma exposure appears to be more complex, and does not fit well to simple additivity or multiplicativity. Analysis of interaction as departure from additivity used linear regression, which is not robust to sizable differences between the proportion of cases in the sample and in the population. However, there is only a minor difference between the proportion of cases in the sample (31.7%) and the population incidence

of depression (28%), so this limitation is unlikely to affect our additive interaction analysis. The gene-environment correlation between PRS and reported trauma exposure complicates the interpretation of these results further, as it may alter the observed interaction. Simulation study analyses have demonstrated the complex pattern of analysis results when the underlying relationship between a genotype and an environmental risk factor is known [26]. In real data, when the underlying relationship is unknown, caution is needed before drawing strong conclusions. However, this limitation is not unique for this study as it also impaired previous analyses of interaction with polygenic risk scores [24–26].

Sensitivity analyses focussed on trauma found that self-reported traumatic experience was significantly heritable, as has been previously observed [17]. 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 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 has been proposed to underlie observed genetic correlations with socioeconomic status [80].

Our finding of a broader genetic basis of depression in individuals reporting trauma exposure reinforces similar conclusions from a study of depression and adversity in Han Chinese women [23]. However, we did not replicate individual variant results from the same study. This is unsurprising, as there are a number of differences between the studies, including culture and ethnicity (although there are

not substantial differences in the allele frequencies of the relevant variants between East Asians and Europeans), as well as the methodology applied in the phenotyping and analysis of the cohorts [23]. Most notably, the previous study did not report a genetic correlation between depression and adversity, in contrast to that observed herein between depression and reported trauma exposure; this potentially results from the previous study focussing only on events occurring before the onset of depression.

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 [11]. 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 cohorts like 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. As noted, 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 [81]. 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 [82].

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 other complex traits. In particular, we identified no broad difference in the direction of common genetic effects on depression when stratifying by reported trauma exposure. However, subtle intrinsic differences may exist. In particular, the genetic correlates of depression may partially reflect specific 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). 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).

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. 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).

8. 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).
9. 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).
10. Collishaw, S. *et al.* Resilience to adult psychopathology following childhood maltreatment: evidence from a community sample. *Child Abuse Negl.* **31,** 211–229 (2007).
11. 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).
12. 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).
13. 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).
14. 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).
15. 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).
16. 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

17. Power, R. A. *et al.* Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol. Med.* **43,** 1965–1971 (2013).
18. 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).
19. 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).
20. Lyons, M. J. *et al.* Do genes influence exposure to trauma? A twin study of combat. *Am. J. Med. Genet.* **48,** 22–27 (1993).
21. Dunn, E. C. *et al.* Genetic determinants of depression: recent findings and future directions. *Harv. Rev. Psychiatry* **23,** 1–18 (2015).
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* appiajp201717060621 (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 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. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
29. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).
30. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
31. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26,** 2867–2873 (2010).
32. 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).
33. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33,** 2776–2778 (2017).
34. 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).
35. 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).

36. 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).
37. 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).
38. Team, R. C. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. (2014).
39. Townsend, P., Phillimore, P. & Beattie, A. *Health and Deprivation: Inequality and the North*. (Croom Helm, 1988).
40. 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).
41. 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).
42. Revelle, W. R. psych: Procedures for personality and psychological research. (2017).
43. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32,** 227–234 (2008).
44. 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
45. Marioni, R. *et al.* GWAS on family history of Alzheimer's disease. *bioRxiv* 246223 (2018). doi:10.1101/246223

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

47. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526,** 75–81 (2015).

48. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* **47,** 11.12.1–34 (2014).

49. 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

50. 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).

51. 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).

52. Morgan, C. & Fisher, H. Environment and schizophrenia: environmental factors in schizophrenia: childhood trauma--a critical review. *Schizophr. Bull.* **33,** 3–10 (2007).

53. 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).

54. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

55. 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).

56. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).

57. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

58. 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).

59. Daniel. *Biostatistics: A Foundation For Analysis In Health Sciences, 7Th Ed*. (Wiley India Pvt. Limited, 2006).

60. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).

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

62. 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).

63. 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).

64. 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).

65. Yzerbyt, V. Y., Muller, D. & Judd, C. M. Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *J. Exp. Soc. Psychol.* **40,** 424–431 (2004).

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

67. Nagel, M. *et al.* GWAS Meta-Analysis of Neuroticism (N=449,484) Identifies Novel Genetic Loci and Pathways. *bioRxiv* 184820 (2017). doi:10.1101/184820

68. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry* **21,** 749–757 (2016).

69. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50,** 6–11 (2018).

70. 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).

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

72. 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).

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

74. 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).

75. de Wit, L. *et al.* Depression and obesity: a meta-analysis of community-based studies. *Psychiatry Res.* **178,** 230–235 (2010).

76. 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).

77. Afari, N. *et al.* Depression and obesity: do shared genes explain the relationship? *Depress. Anxiety* **27,** 799–806 (2010).

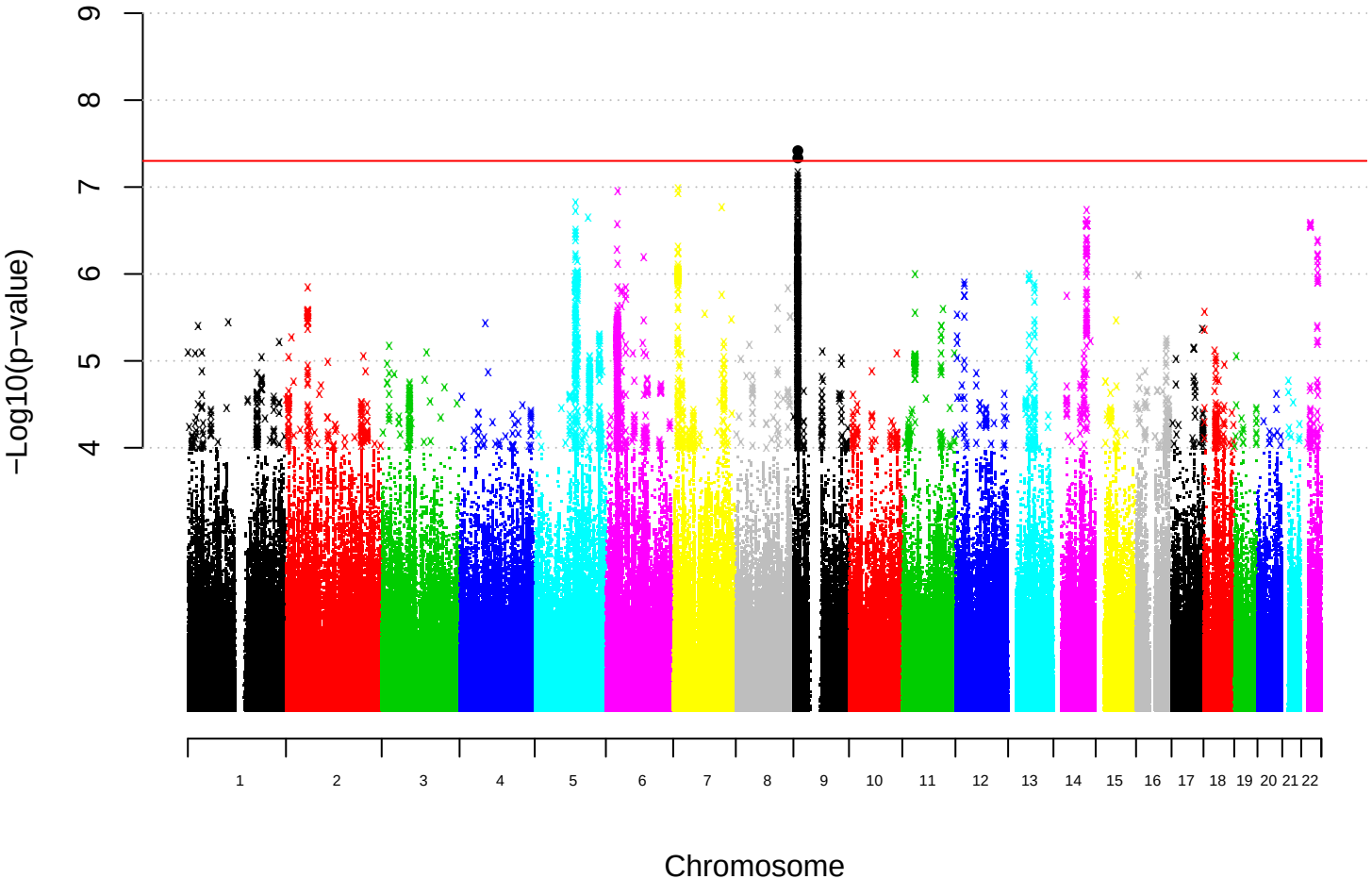
78. Richardson, A. S., Dietz, W. H. & Gordon-Larsen, P. The association between childhood sexual and physical abuse with incident adult severe obesity across 13 years of the National Longitudinal Study of Adolescent Health. *Pediatr. Obes.* **9,** 351–361 (2014).

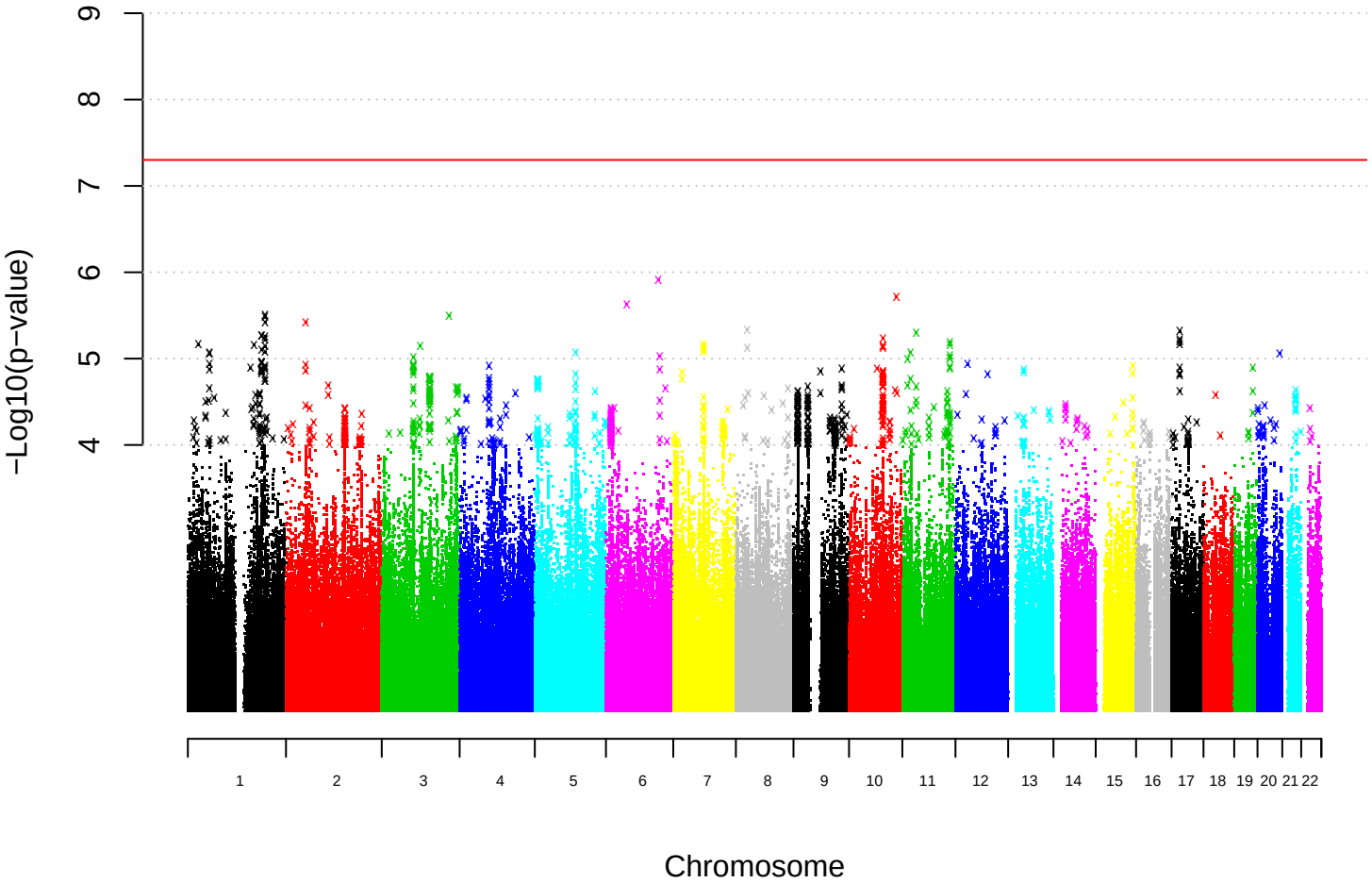
79. Tyrrell, J. *et al.* Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ* **352,** i582 (2016).

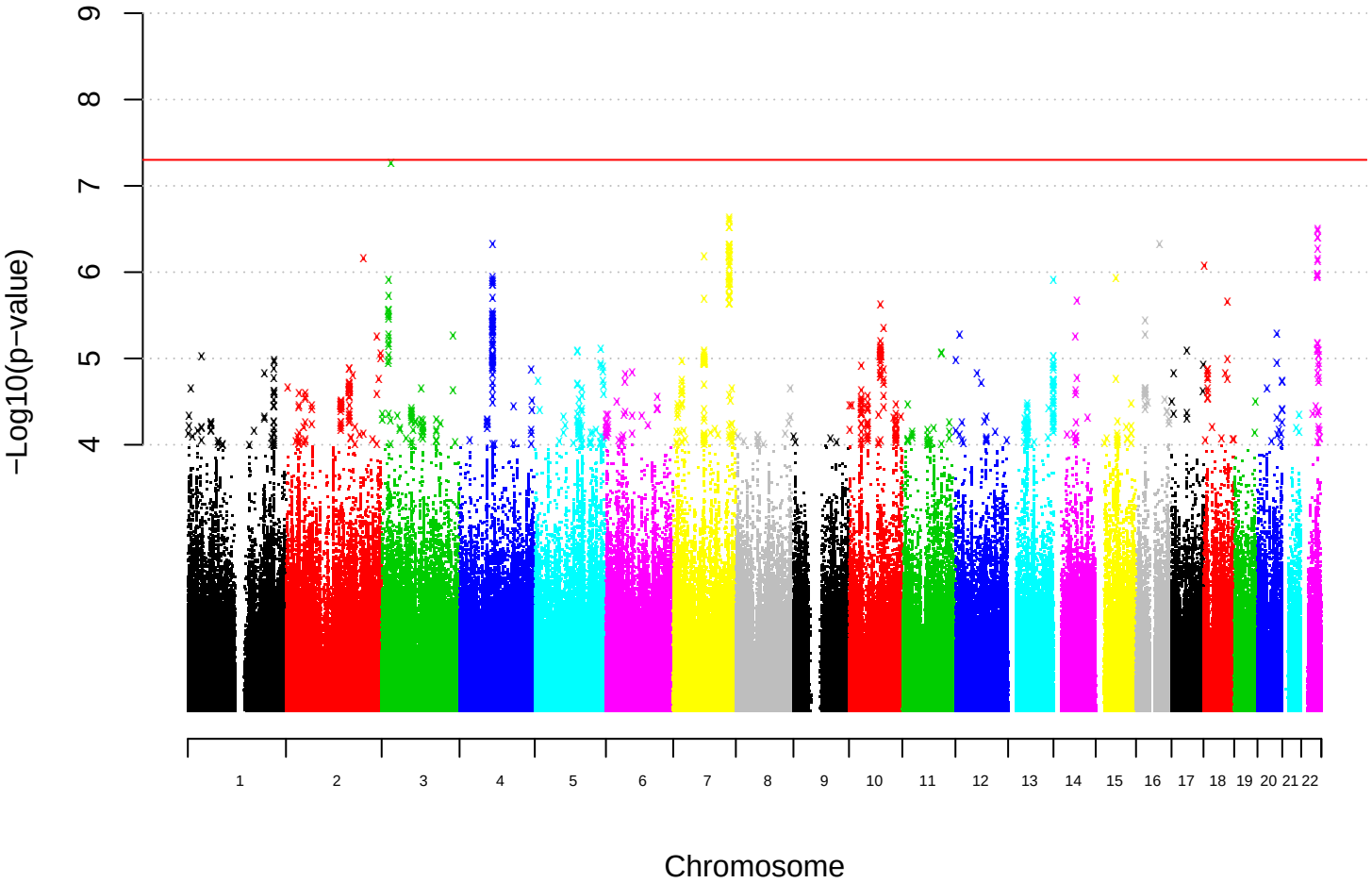
80. Hill, W. D. *et al.* Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Curr. Biol.* **26,** 3083–3089 (2016).

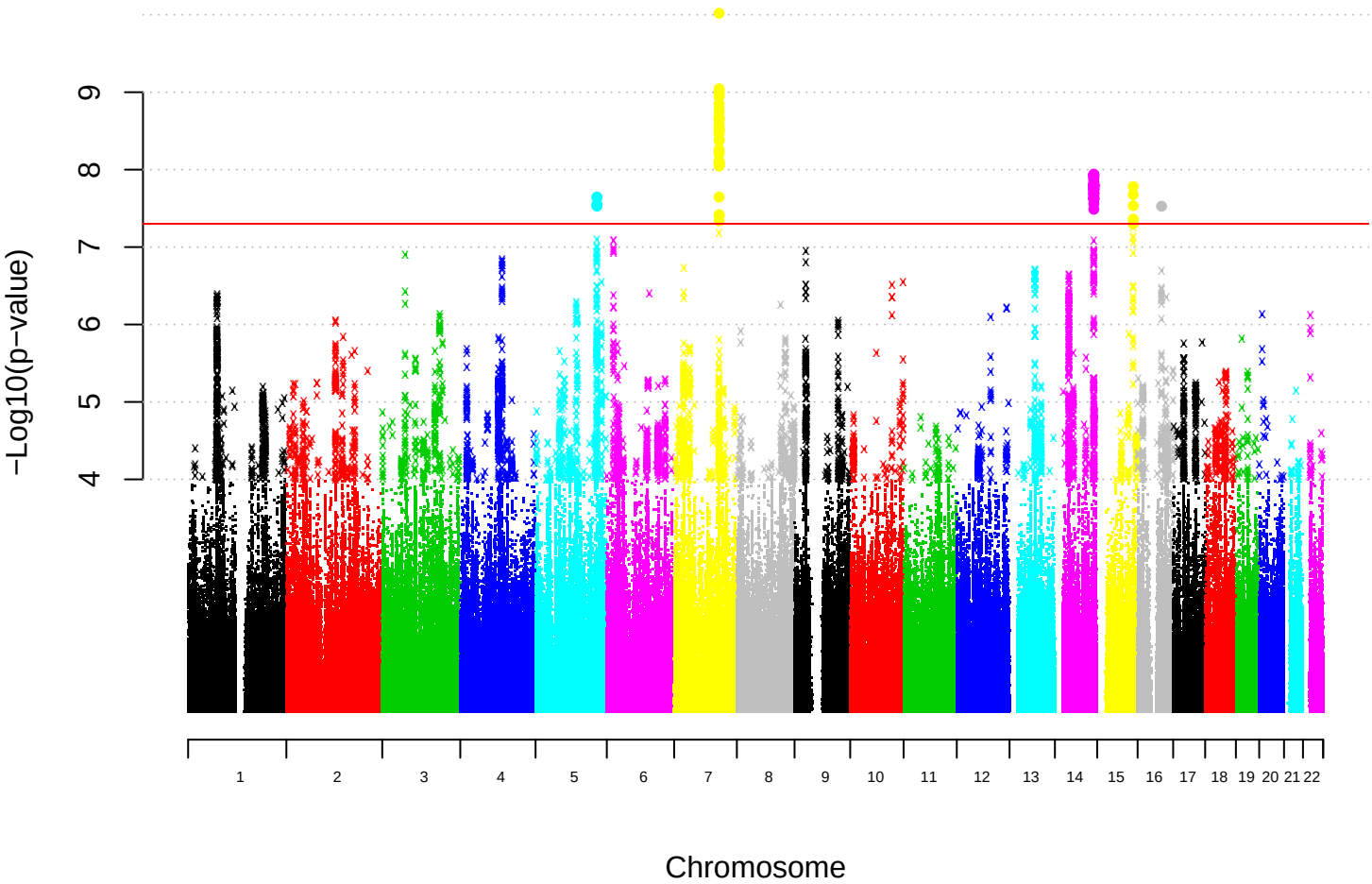
81. Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N. & Allen, N. E. OP41 The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. *J. Epidemiol. Community Health* **70,** A26–A26 (2016).

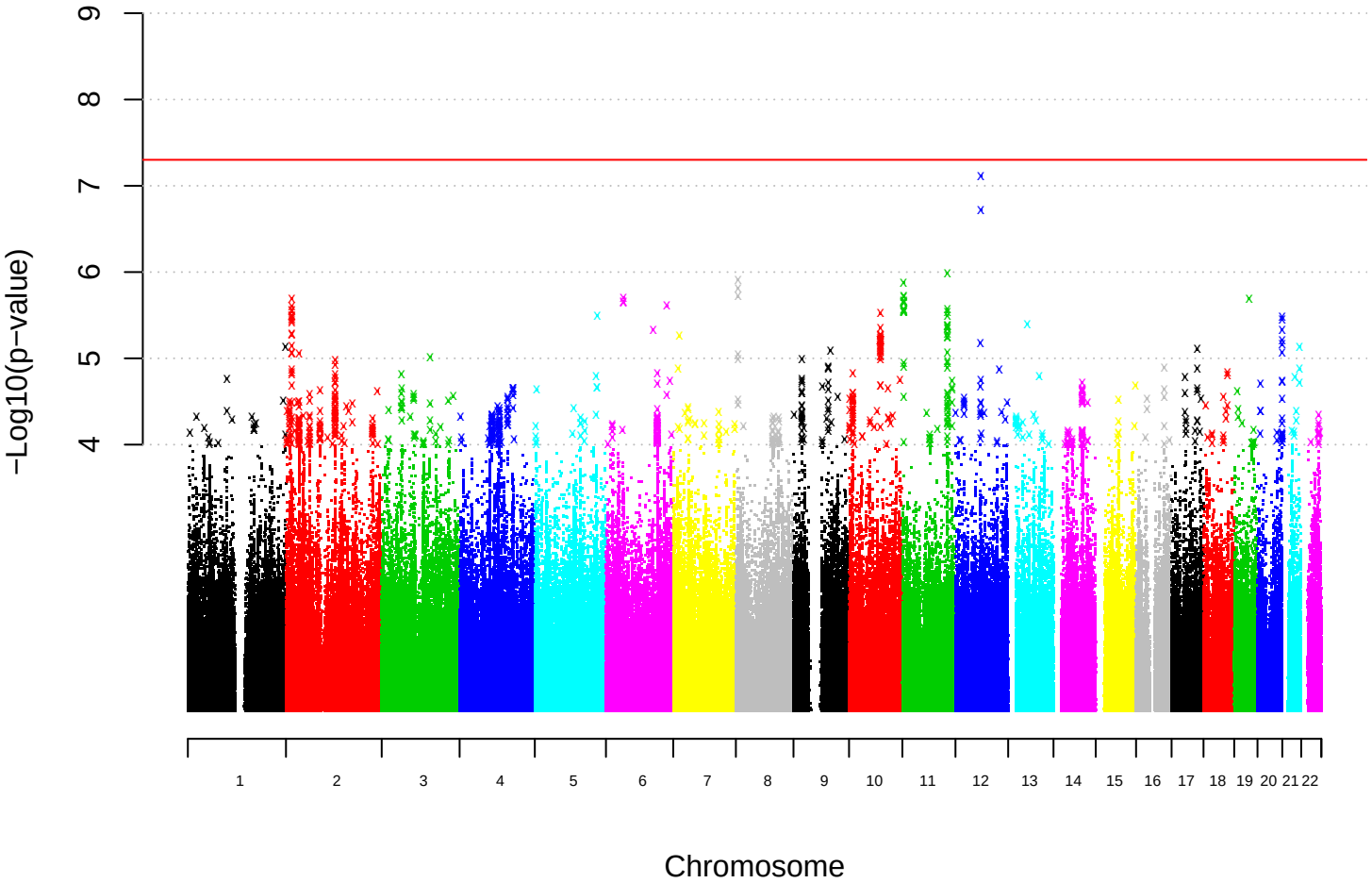
82. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).

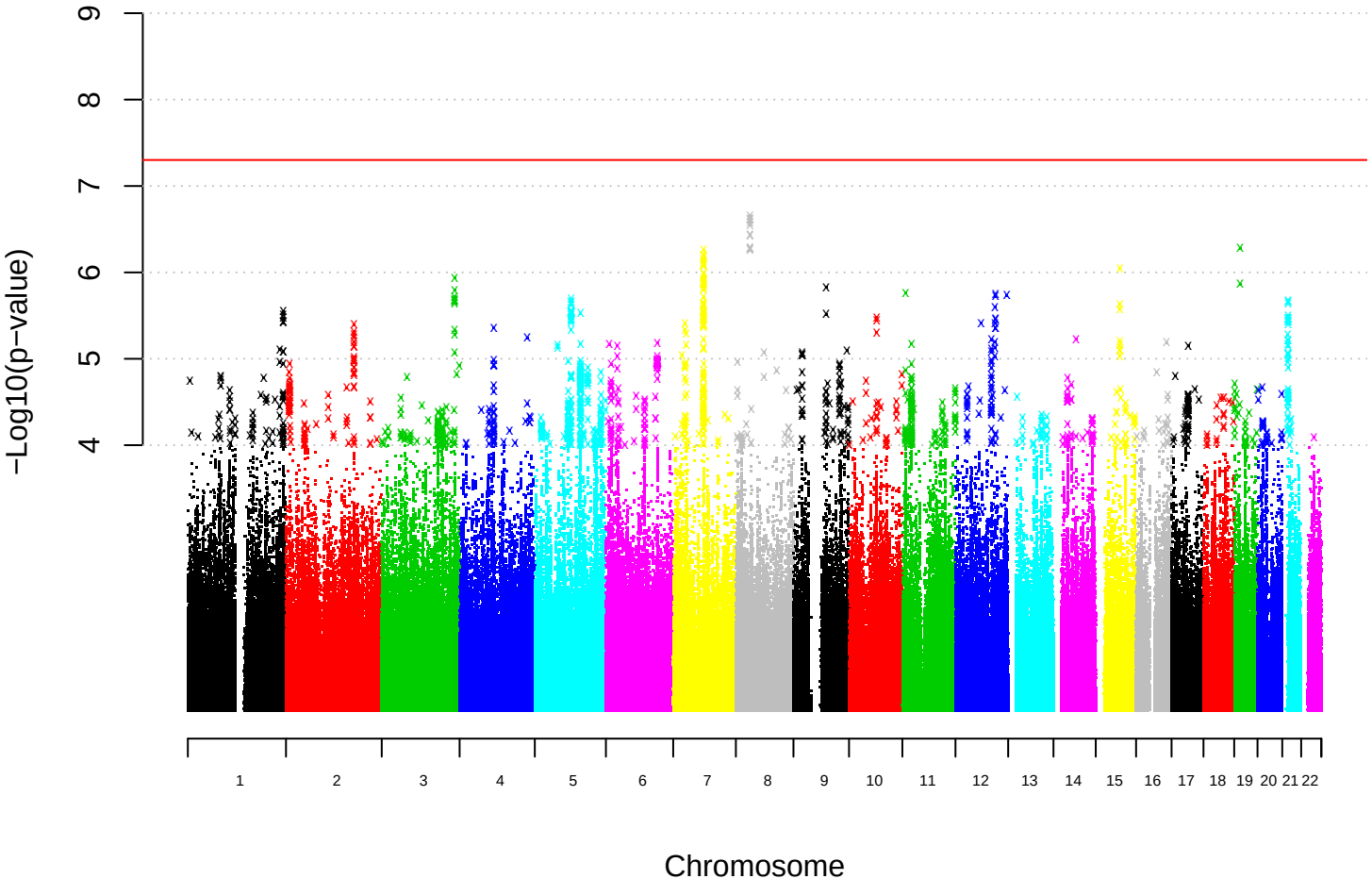

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

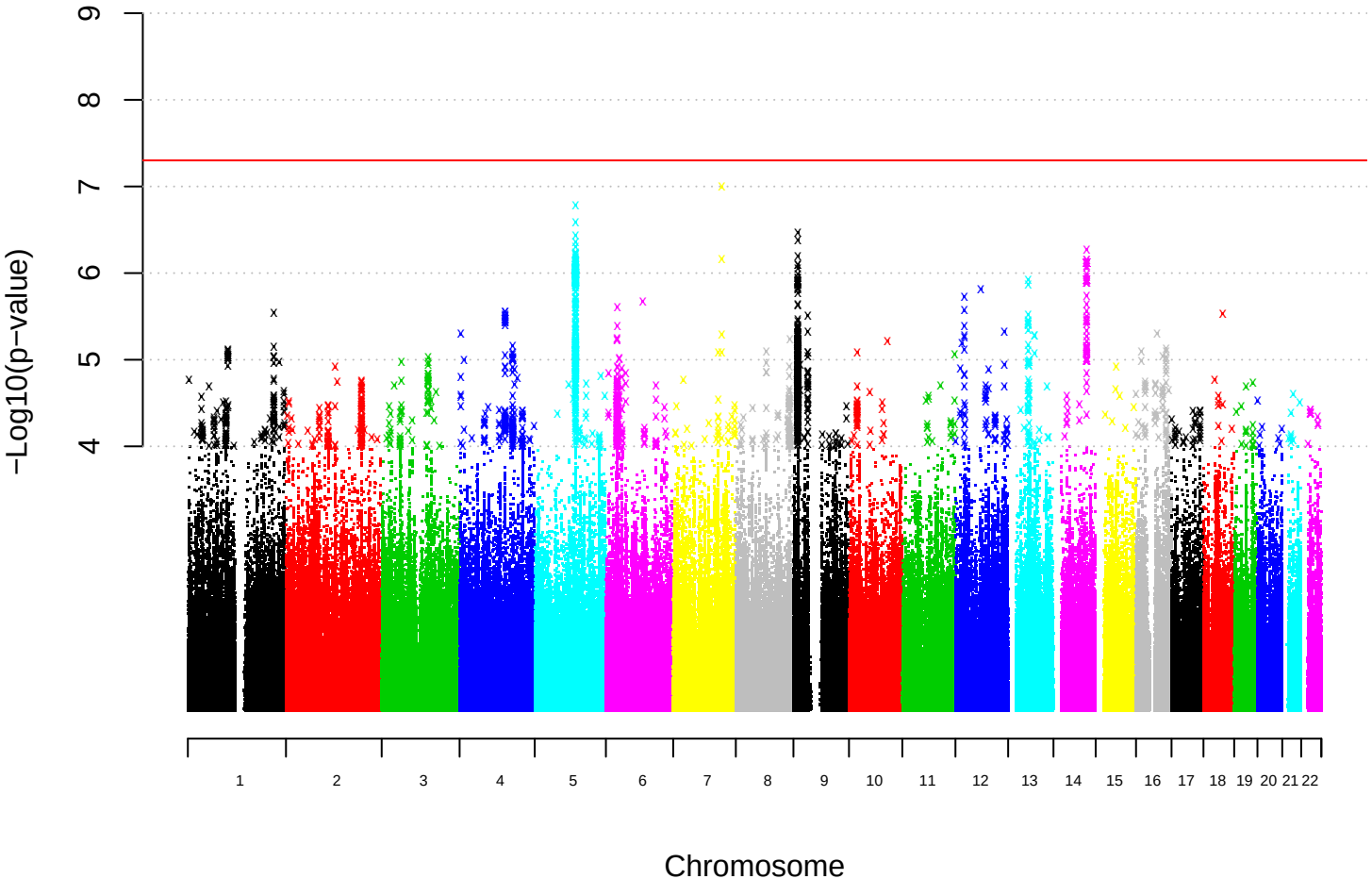

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

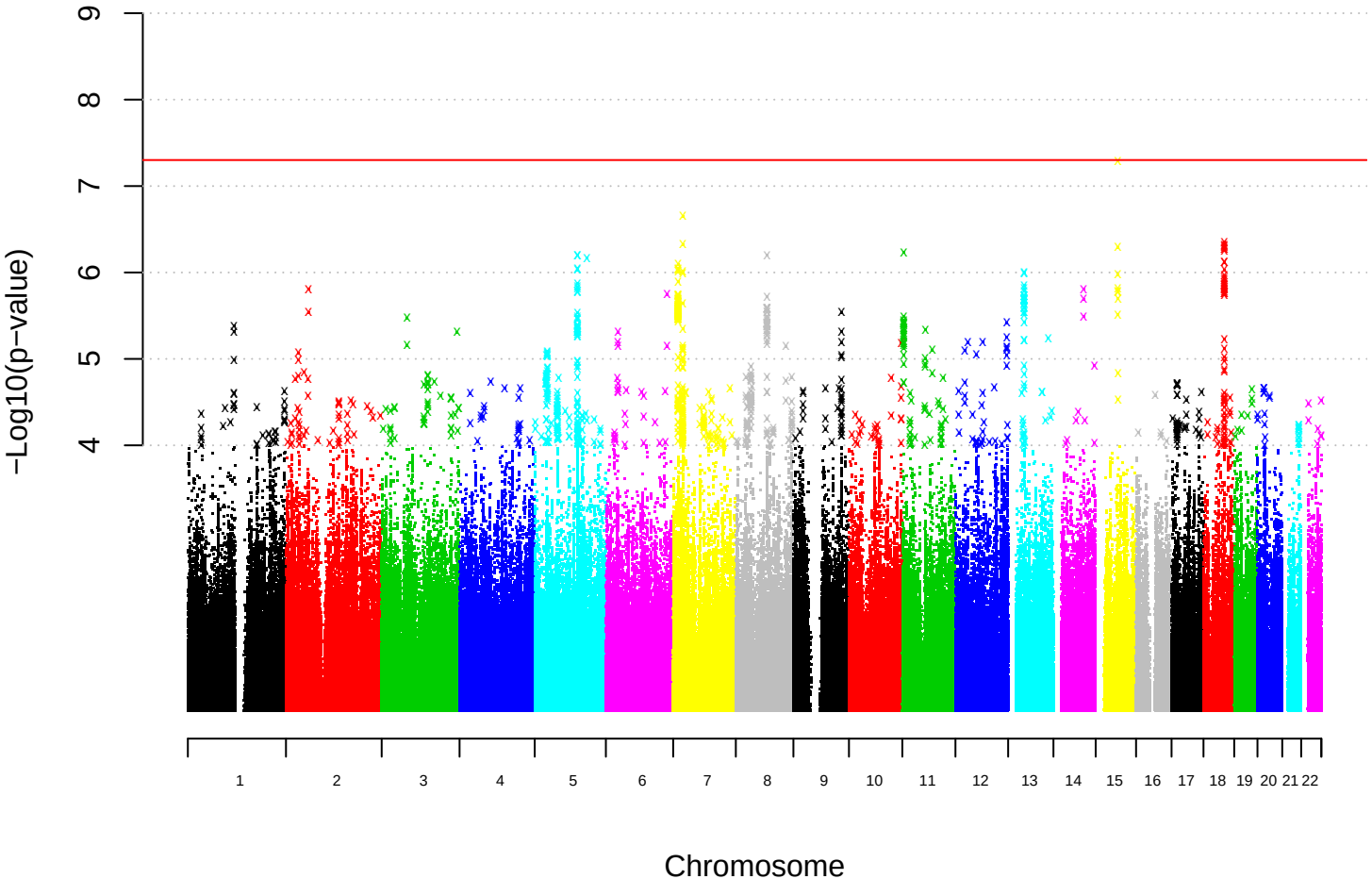

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

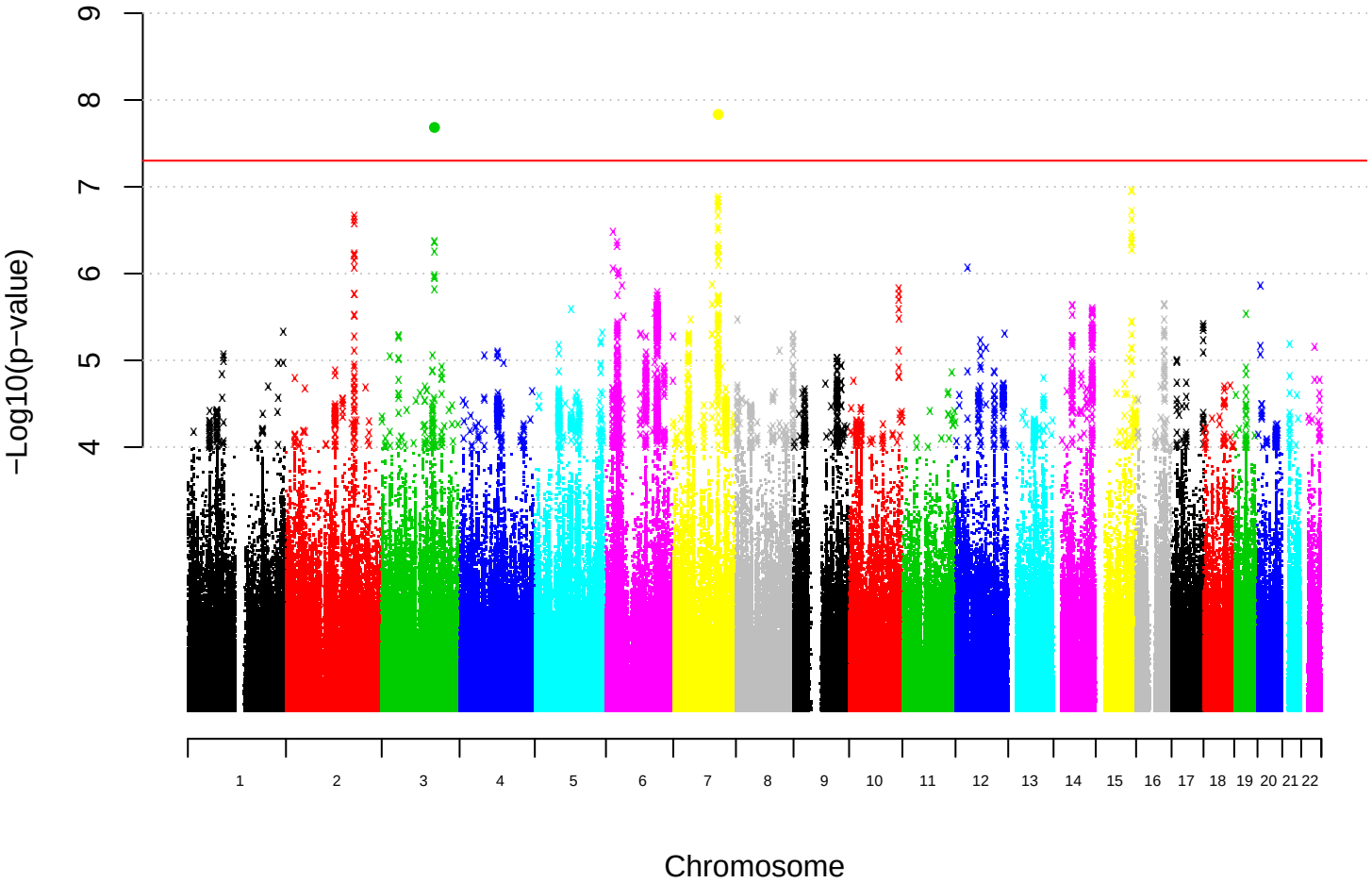

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

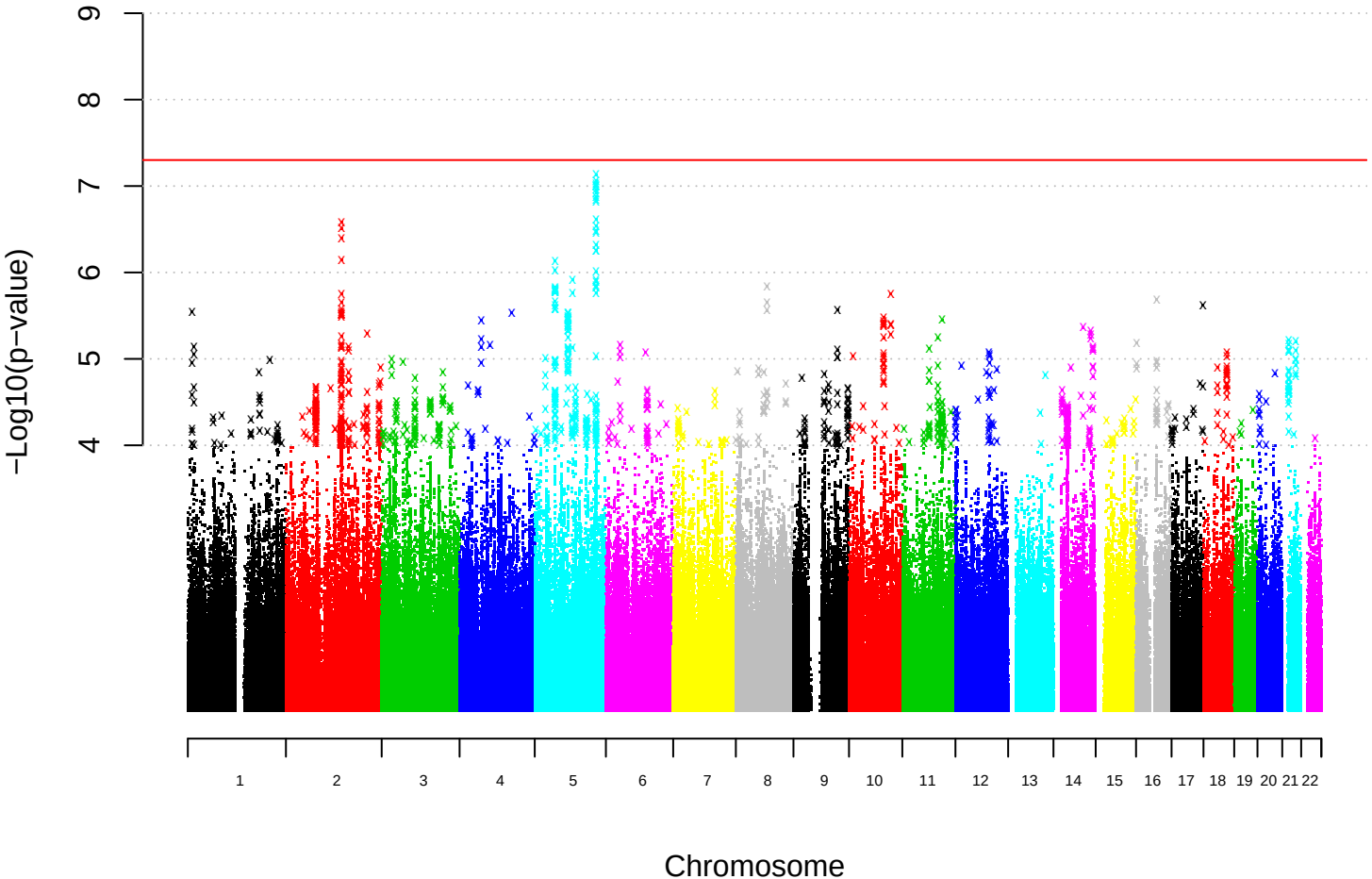

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

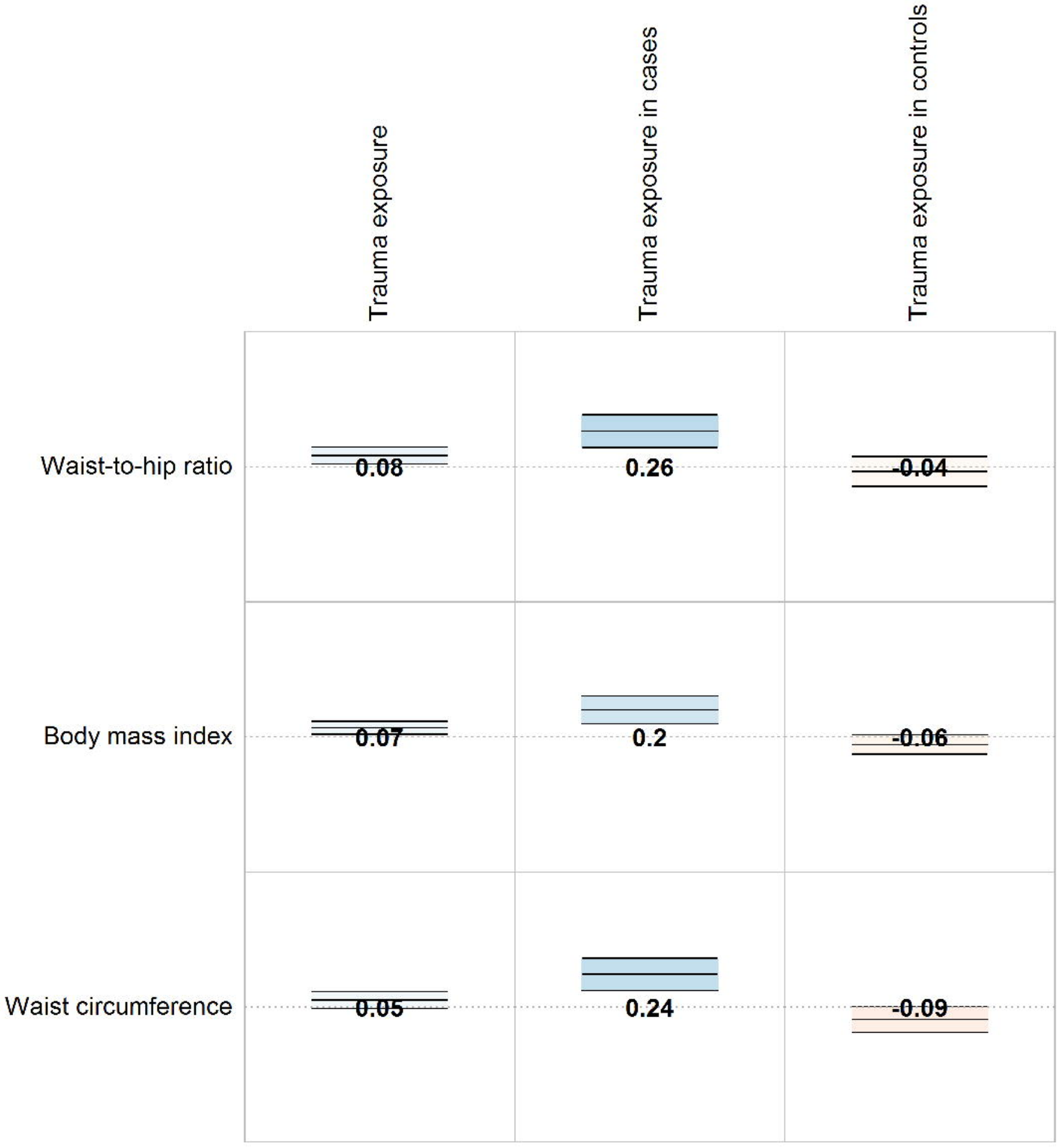

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

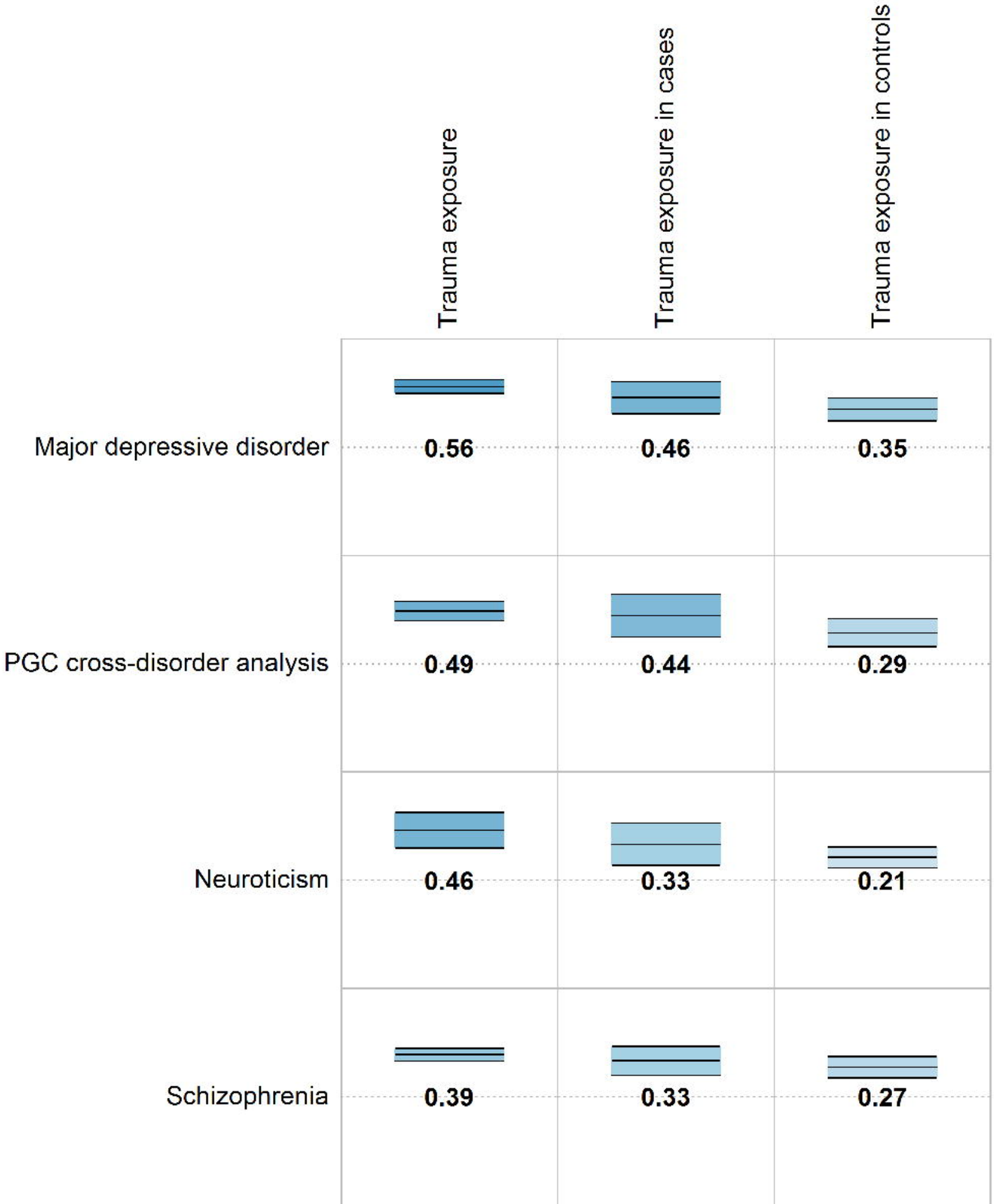

−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


−Log10(p−value)
9
8
7
6
5
4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


Trauma exposure
Trauma exposure in cases
Trauma exposure in controls
Waist-to-hip ratio
0.08
0.26
-0.04
Body mass index
0.07
0.2
-0.06
Waist circumference
0.05
0.24
-0.09


Trauma exposure
Trauma exposure in cases
Trauma exposure in controls
Major depressive disorder
0.56
0.46
0.35
PGC cross-disorder analysis
0.49
0.44
0.29
Neuroticism
0.46
0.33
0.21
Schizophrenia
0.39
0.33
0.27