Genomic Influences on Self-Reported Childhood Maltreatment

Shareefa Dalvie[1], Adam X. Maihofer[2,3,4], Jonathan R.I. Coleman[5,6], Bekh Bradley[7,8], Gerome Breen[5,6], Leslie A. Brick[9], Chia-Yen Chen[10,11,12], Karmel W. Choi[12,13,14], Laramie E. Duncan[15], Guia Guffanti[16,17], Magali Haas[18], Supriya Harnal[12], Israel Liberzon[19], Nicole R. Nugent[9,20], Allison C. Provost[18], Kerry J. Ressler[8,16,17], Katy Torres[2,3,4], Ananda B. Amstadter[21], S. Bryn Austin[22,23,24,25], Dewleen G. Baker[2,3,26], Elizabeth A. Bolger[17,16], Richard A. Bryant[27], Joseph R. Calabrese[28], Douglas L. Delahanty[29,30], Lindsay A. Farrer[31], Norah C. Feeny[32], Janine D. Flory[33], David Forbes[34], Sandro Galea[35], Aarti Gautam[36], Joel Gelernter[37,38,39], Rasha Hammamieh[40], Marti Jett[40], Angela G. Junglen[29], Milissa L. Kaufman[16,17], Ronald C. Kessler[41], Alaptagin Khan[17,41], Henry R. Kranzler[42,43], Lauren A. M. Lebois[16,17], Charles Marmar[44], Matig R. Mavissakalian[28], Alexander McFarlane[45], Meaghan O'Donnell[34], Holly K. Orcutt[46], Robert H. Pietrzak[47,48], Victoria B. Risbrough[2,3,4], Andrea L. Roberts[49], Alex O. Rothbaum[32], P. Roy-Byrne[50], Ken Ruggiero[51], Antonia V. Seligowski[17,16], Christina M. Sheerin[21], Derrick Silove[52], Jordan W. Smoller[10,12,14], Nadia Solovieff[11], Murray B. Stein[2,26,53], Martin H. Teicher[16,17], Robert J. Ursano[54], Miranda Van Hooff[45], Sherry Winternitz[16,17], Jonathan D. Wolff[17], Rachel Yehuda[33,55], Hongyu Zhao[56], Lori A. Zoellner[57], Dan J. Stein[1], Karestan C. Koenen[11,12,58], Caroline M. Nievergelt[2,3,4]

[1]University of Cape Town, SA MRC Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry, Cape Town, Western Cape, ZA. [2]University of California San Diego, Department of Psychiatry, La Jolla, CA, US. [3]Veterans Affairs San Diego Healthcare System, Center of Excellence for Stress and Mental Health, San Diego, CA, US. [4]Veterans Affairs San Diego Healthcare System, Research Service, San Diego, CA, US. [5]King's College London, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, London, GB. [6]King's College London, NIHR BRC at the Maudsley, London, GB. [7]Atlanta VA Health Care System, Mental Health Service Line, Decatur, GA, US. [8]Emory University, Department of Psychiatry and Behavioral Sciences, Atlanta, GA, US. [9]Alpert Medical School of Brown University, Providence, RI, US. [10]Massachusetts General Hospital, Analytic and Translational Genetics Unit, Boston, MA, US. [11]Massachusetts General Hospital, Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Boston, MA, US. [12]Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research, Cambridge, MA, US. [13]Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA, US. [14]Massachusetts General Hospital, Department of Psychiatry, Boston, MA, US. [15]Stanford University, Department of Psychiatry and Behavioral Sciences, Stanford, CA, US. [16]Harvard Medical School, Department of Psychiatry, Boston, MA, US. [17]McLean Hospital, Belmont, MA, US. [18]Cohen Veterans Bioscience, Cambridge, MA, US. [19]University of Michigan Medical School, Department of Psychiatry, Ann Arbor, MI, US.

[20]Bradley/Hasbro Children's Research Center of Rhode Island Hospital, Providence, RI, US. [21]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Richmond, VA, US. [22]Harvard Medical School, Department of Pediatrics, Boston, MA, US. [23]Harvard School of Public Health, Department of Social and Behavioral Sciences, Boston, MA, US. [24]Boston Children's Hospital, Division of Adolescent and Young Adult Medicine, Boston, MA, US. [25]Brigham and Women's Hospital, Channing Division of Network Medicine, Boston, MA, US. [26]Veterans Affairs San Diego Healthcare System, Psychiatry Service, San Diego, CA, US. [27]University of New South Wales, Department of Psychology, Sydney, NSW, AU. [28]University Hospitals, Department of Psychiatry, Cleveland, OH, US. [29]Kent State University, Department of Psychological Sciences, Kent, OH, US. [30]Kent State University, Research and Sponsored Programs, Kent, OH, US. [31]Boston University School of Medicine, Department of Medicine, Boston, MA, US. [32]Case Western Reserve University, Department of Psychological Sciences, Cleveland, OH, US. [33]Icahn School of Medicine at Mount Sinai, Department of Psychiatry, New York, NY, US. [34]University of Melbourne, Department of Psychiatry, Melbourne, VIC, AU. [35]Boston University, Department of Psychological and Brain Sciences, Boston, MA, US. [36]US Army Medical Research and Materiel Command, Fort Detrick, MD, US. [37]US Department of Veterans Affairs, Department of Psychiatry, West Haven, CT, US. [38]VA Connecticut Healthcare Center, West Haven, CT, US. [39]Yale University School of Medicine, Department of Genetics and Neuroscience, New Haven, CT, US. [40]US Army Medical Research and Materiel Command, USACEHR, Fort Detrick, MD, US. [41]Harvard Medical School, Department of Health Care Policy, Boston, MA, US. [42]University of Pennsylvania Perelman School of Medicine, Department of Psychiatry, Philadelphia, PA, US. [43]Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, US. [44]New York University School of Medicine, Department of Psychiatry, New York, NY, US. [45]University of Adelaide, Centre for Traumatic Stress Studies, Adelaide, South Australia, AU. [46]Northern Illinois University, Department of Psychology, DeKalb, IL, US. [47]U.S. Department of Veterans Affairs National Center for Posttraumatic Stress Disorder, West Haven, CT, US. [48]Yale University School of Medicine, Department of Psychiatry, New Haven, CT, US. [49]Harvard T.H. Chan School of Public Health, Department of Environmental Health, Boston, MA, US. [50]University of Washington, Department of Psychiatry, Seattle, WA, US. [51]Medical University of South Carolina, Department of Nursing and Department of Psychiatry, Charleston, SC, US. [52]University of New South Wales, Department of Psychiatry, Sydney, NSW, AU. [53]Veterans Affairs San Diego Healthcare System, Million Veteran Program, San Diego, CA, US. [54]Uniformed Services University, Department of Psychiatry, Bethesda, Maryland, US. [55]James J Peters VA Medical Center, Department of Mental Health, Bronx, NY, US. [56]Yale University, Department of Biostatistics, New Haven, CT, US. [57]University of Washington, Department of Psychiatry and Behavioral Sciences, Seattle, WA, US. [58]Harvard School of Public Health, Department of Epidemiology, Boston, MA, US.

Abstract

Childhood maltreatment is highly prevalent and serves as a risk factor for mental and physical disorders. Self-reported childhood maltreatment appears heritable, but the specific genetic influences on this phenotype are largely unknown. The aims of this study were to 1) identify genetic variation associated with reported childhood maltreatment, 2) calculate the relevant SNP-based heritability estimates, and 3) quantify the genetic overlap of reported childhood maltreatment with mental and physical health-related phenotypes. Genome-wide association analysis for childhood maltreatment was undertaken, using a discovery sample from the UK Biobank (UKBB) (n=124,000) and a replication sample from the Psychiatric Genomics Consortium–posttraumatic stress disorder working group (PGC-PTSD) (n=26,290). Heritability estimations for childhood maltreatment and genetic correlations with mental/physical health traits were calculated using linkage disequilibrium score regression (LDSR). Two genome-wide significant loci associated with childhood maltreatment, located on chromosomes 3p13 (rs142346759, beta=0.015, p=4.35x10^{-8}, *FOXP1*) and 7q31.1 (rs10262462, beta=-0.016, p=3.24x10^{-8}, *FOXP2*), were identified in the discovery dataset but were not replicated in the PGC-PTSD sample. SNP-based heritability for childhood maltreatment was estimated to be ~6%. Childhood maltreatment was most significantly genetically correlated with depressive symptoms (r$_g$=0.70, p=4.65x10^{-40}). This is the first large-scale genetic study to identify specific variants associated with self-reported childhood maltreatment. *FOXP* genes could influence traits such as depression and thereby be relevant to childhood maltreatment. Alternatively, these variants may be associated with a greater likelihood of reporting maltreatment. A clearer understanding of the genetic relationships of childhood maltreatment, including particular abuse subtypes, with various psychiatric disorders, may ultimately be useful in in developing targeted treatment and prevention strategies.

Introduction

The lifetime prevalence of childhood physical, sexual and emotional abuse ranges from 8-36%[1]. In addition to being highly prevalent, childhood maltreatment is associated with the development of mental disorders, including depression[2, 3], and physical ill health, including non-communicable diseases[4, 5]. Although these associations are now well established, estimates of effect size vary considerably across epidemiological studies, likely reflecting methodological challenges, including uncertainty about how best to assess childhood maltreatment [6].

A twin-based study found that retrospective reports of childhood maltreatment has a heritability of 6% [7]. Although the idea that childhood maltreatment is heritable may be counter-intuitive, the field of behavior genetics has long documented the heritability of many exposures seen as environmental. This is referred to as gene-environment correlation (rGE). Three potential rGE mechanisms to explain the heritability of childhood maltreatment may be posited. First, a "passive" rGE: parental genes affecting parental behavior may influence the childhood environment (e.g. aggressive parents may be more likely to physically punish their children who have also inherited their parents' genetic variants influencing aggressive behavior [8]). Second, an "active" rGE: individuals with genetic variants associated with certain behavioral phenotypes may be more at risk of selecting or creating adverse situations (e.g. risk-taking personality is heritable and children who are high in risk taking may be exposed to more trauma)[9, 10]. Third, an "evocative" rGE: genetic variation may influence child behavior, which in turn is associated with responses to the child (e.g. genetic factors may influence infant "difficultness", which in turn is associated with maternal hostile-reactive behavior that is correlated with child abuse[11, 12]). The latter two rGEs are sometimes collectively referred to as non-passive correlations[7].

While a number of key risk factors for childhood maltreatment, including child behavioral characteristics and parental mental health, have been investigated[6], studies have seldom focused on associated genetic variation. The few genetic association studies of childhood maltreatment have only considered variants in candidate genes[13] and have had insufficient power to detect the small polygenic effect sizes typically associated with behavioral phenotypes[14]. Also, there are no studies of the genetic overlap of childhood maltreatment with mental and physical health-related traits, using genome-wide single nucleotide

polymorphism (SNP) data. Knowledge of specific genetic variation for childhood maltreatment, the heritability of this phenotype, and the genetic overlap with other traits may be useful in informing our understanding of the identified risk factors, the etiology and the outcomes of childhood maltreatment. This, in turn, may have important implications for the design of prevention and treatment programs for adverse health outcomes. For example, an observed association between an adverse health outcome and genetic variants may be mediated by an environmental exposure, if that exposure is shown to be causal of the health outcome. Thus, preventative strategies would focus on decreasing the risk conferred by the environmental exposure without needing to specifically consider the genetic influences on the health outcome[9].

The PGC-PTSD consortium has collaborated to obtain access to well-powered genetic studies of trauma and PTSD that have allowed a number of key genetic questions in this field to be investigated[15-17], providing a unique opportunity to address knowledge gaps in the area of childhood maltreatment. This study aims to: 1) identify genetic variants associated with childhood maltreatment using a genome-wide association study (GWAS) design, 2) quantify the heritability of childhood maltreatment using SNP-based methods, and 3) assess the degree of genetic overlap of childhood maltreatment with mental and physical health-related phenotypes.

Methods

Participating studies

Nineteen studies, comprising subjects of European ancestry only, were used in this analysis. The discovery dataset consisted of 124,711 individuals from the UK Biobank (UKBB)[18] and the replication sample comprised 26,290 individuals, a subset of the PGC-PTSD Freeze 1.5 dataset (PGC1.5)[17]. The details of these studies, including the demographics and instruments used to assess maltreatment can be found in **Supplementary Table 1**.

Phenotype Harmonization

For the childhood maltreatment phenotype, Childhood Trauma Questionnaire (CTQ) scores on physical, sexual, and emotional abuse [19] were obtained from the participating studies. From this, an overall childhood maltreatment count score of 0-3 was constructed, based on

a count of the three abuse categories listed above. An individual was considered to have endorsed a childhood abuse category if they scored in the moderate to extreme range for that particular category, per established cut-offs [20] (**Supplementary Table 2**). If CTQ data were not available, the event assessment during childhood (occurring before 18 years of age) that was most validated for that particular study was obtained, providing a count of the total number of different categories of reported childhood events (e.g. physical, sexual or severe emotional abuse) along with the range of possible scores for the measure. The reported maltreatment exposure from the UKBB dataset comprised a score of three items where participants were asked whether they were i) "physically abused by family as a child", ii) "sexually molested as a child", and whether they iii) "felt hated by family member as a child". The childhood maltreatment count score, whether it was generated from the CTQ or another instrument, was used as the main outcome measure in the association analysis. The range of this maltreatment count score for each study can be seen in **Supplementary Table 1**.

Global Ancestry Determination, Genotyping quality control and Imputation

Study participants from the PGC-PTSD were genotyped with a number of different arrays (**Supplementary Table 1**). Genotype data were quality controlled and processed using the standard PGC pipeline, Ricopili-MANC (https://sites.google.com/a/broadinstitute.org/ricopili/ and https://github.com/orgs/Nealelab/teams/ricopili) as part of the PGC-PTSD Freeze 2 data analysis[17, 21]. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. A detailed outline of these methods can be found in[17]. Briefly, ancestry was determined with pre-QC genotypes using a SNPweights panel of 10,000 ancestry informative markers from a reference panel comprising 2911 subjects from 71 diverse populations and six continental groups (https://github.com/nievergeltlab/global_ancestry). Samples were excluded if they had call rates <98%, deviated from the expected inbreeding coefficient (fhet < -0.2 or > 0.2), or had a sex discrepancy between reported and genotypic sex (based on inbreeding coefficients calculated from SNPs on the X chromosome). Markers were excluded if they had call rates <98%, >2% difference in missing genotypes between PTSD cases and controls, or were monomorphic. Markers with a Hardy-Weinberg equilibrium (HWE) $p<1\times10^{-6}$ in controls were excluded from all subjects. Principal components (PCs) were calculated using the smartPCA algorithm in EIGENSTRAT[22]. Pre-phasing and phasing was performed using

SHAPEIT2 v2.r837[23]. Imputation was performed with IMPUTE2 v2.2.2[24] using the 1000 Genomes (1000G) phase 3 data[25] as the reference.

Details regarding the QC, imputation and ancestry determination of the UKBB dataset can be found in[26]. Briefly, study participants were genotyped with two custom genotyping arrays (with ~800,000 markers). A two-stage imputation was performed using the Haplotype Reference Consortium (HRC)[27] and the UK10K[28] as the reference panels. Variants were filtered to include only those with a minor allele frequency (MAF) of > 1% and an INFO score of > 0.4. Related individuals (third degree and closer) and those with a genotyping call rate < 98% were excluded. Ancestry was determined by 4-means clustering on the first two PCs provided by the UKBB[29]. Additional principal component analysis was conducted on the European-only data subset using flashpca2[30].

Main GWAS

GWAS analysis was conducted separately for each study. Best-guess genotypes were tested for association to reported childhood maltreatment using an ordinal logistic regression model with age, sex, and the first five PCs included as covariates. Variants with a MAF < 0.5% and a genotyping rate < 98% were excluded. These analyses were implemented in PLINK 1.9[31] using the plug-in *Rserve*. To ensure computational efficiency, linear regression models were run for four of the larger contributing studies (NSS1; NSS2; PPDS; and UKBB, N=143,392 subjects)[17]. For the NSS1, NSS2, and PPDS studies, age, sex, and 5 PCs were included as covariates in the regression model. For the UKBB dataset, the regression analysis was implemented in BGenie v1.2[32] with age, sex, 6 PCs, batch, and site included as covariates.

Meta-analysis

As both linear and ordinal logistic models were implemented in the GWASs, which resulted in different effect statistics, fixed effects meta-analysis was conducted across studies using p-values and direction of effect, weighted according to the effective sample size as the analysis scheme, in METAL (v. March 25 2011)[33]. Heterogeneity across datasets was tested using the Cochran's Q-test for heterogeneity, also implemented in METAL. Only variants with an INFO score of greater than 0.8 and a MAF of greater than 5% were included in the meta-analysis, except where otherwise indicated in the results. Forest plots were generated for genome-wide significant hits using the R package *meta*[34].

Functional mapping and annotation

Genome-wide significant hits identified from the GWAS meta-analysis were annotated using the web-based tool FUnctional Mapping and Annotation (FUMA) v1.3.4c (http://fuma.ctglab.nl/)[35]. Default settings were used and annotations were based on the human genome assembly GRCh37 (hg19). The *SNP2GENE* module was used to identify genomic risk loci and these were mapped to protein coding genes within a 10kb window. An r^2 of ≥ 0.6 was used to identify variants in LD with lead SNPs. The 1000G European Phase 3 was used as the reference dataset. Variants were functionally annotated using ANNOVAR, Combined Dependent Depletion (CADD), RegulomeDB (RDB) and chromatin states (only tissues/cells from brains were included). The NHGRI-EBI GWAS catalog was used to determine any previous associations with the identified risk variants. The GTEx v7 brain tissue, RNAseq data from the CommonMind Consortium and the BRAINEAC database were used to perform eQTL mapping for significant SNP-gene pairs (FDR q < 0.05).

A gene-based analysis was performed within FUMA using MAGMA whereby SNPs were mapped to 18,989 protein coding genes. Genome-wide significance was set at a Bonferroni-corrected threshold $p < 2.63 \times 10^{-6}$. In addition, gene-based test statistics were used to determine whether specific biological pathways are associated with childhood maltreatment. This was performed for 18,989 curated gene sets and GO terms obtained from MsigDB, using MAGMA. The significance threshold was set at a Bonferroni-corrected threshold of $p = 2.63 \times 10^{-6}$ (0.05/18,989).

Heritability and Genetic Correlation Estimation

Linkage disequilibrium score regression (LDSR) is a technique for quantifying polygenicity and confounding, such as population stratification, in GWAS summary statistics[36]. This is accomplished by evaluating the relationship between linkage disequilibrium scores (the average squared correlation of a SNP with all neighboring SNPs) and SNP test statistics. Using this approach, the LDSR intercept was used to estimate the proportion of inflation in test statistics due to polygenic signal (rather than inflation due to population stratification and cryptic relatedness), with the equation *1 - (LDSR intercept -1)/(mean observed chi-square - 1)*[17]. One of the applications of LDSR is the estimation of SNP-based heritability based on GWAS summary statistics. Another application of LDSR is the measurement of genetic correlation i.e. the degree and direction of shared genetic effects between different

traits[36, 37]. Heritability and cross-cohort genetic correlation (r_g) was calculated using LDSR. The web-based interface for LDSR, LD Hub (http://ldsc.broadinstitute.org/ldhub/), was used to further calculate pairwise genetic correlations between childhood maltreatment and 247 non-UKBB traits of interest including psychiatric, anthropomorphic, smoking behavior, reproductive, aging, education, autoimmune, and cardio-metabolic categories.

Results

GWAS and meta-analysis

We report GWAS results from our discovery dataset (UKBB) (n=124,711) and meta-analysis (n=151,001). In our UKBB discovery dataset, we identified two genome-wide significant loci ($p<5 \times 10^{-8}$) associated with childhood maltreatment (**Table 1**, **Figure 1**), rs142346759 (chr3, beta=0.015, $p=4.35 \times 10^{-8}$) and rs10262462 (chr7, beta=-0.016, $p=3.24 \times 10^{-8}$). These variants remained significant in the meta-analysis (**Table 1**, **Supplementary Figures 1 and 2**). Additional variants on chromosome 7 (rs1859100, beta=0.015, $p=3.91 \times 10^{-8}$) and chromosome 12 (rs917577, beta=0.017, $p=2.64 \times 10^{-8}$) (**Figure 2**), also achieved genome-wide significance in the meta-analysis. None of these hits were replicated in PGC1.5 (**Table 1, Supplementary Figure 3**).

Quantile-quantile (qq) plots indicate minimal inflation of p-values across studies (**Supplementary Figures 4-6**). Using the LDSR intercept method, polygenic effects account for 93% and 94% of the observed inflation in test statistics for the UKBB dataset (intercept=1.0096, SE=0.0064) and meta-analysis (intercept=1.0095, SE=0.0077), respectively (**Supplementary Figures 4 and 6**), consistent with minimal population stratification and cryptic relatedness. Heterogeneity estimates in the meta-analyses were not significant (**Table 1**).

Integration with Functional Genomic Data

Using the web-based tool FUMA, the two UKBB GWAS hits were each annotated to two genes, *FOXP1* and *FOXP2* (**Table 2**). Gene-based analysis of the UKBB GWAS summary statistics further identified three gene-wide significant genes, *KIF26B* ($p=1.67 \times 10^{-7}$), *CNTNAP5* ($p=8.89 \times 10^{-7}$), and *EXOC2* ($p=2.04 \times 10^{-6}$) from a total of 18,989 protein coding genes. Gene-set analysis did not reveal any significant pathways associated with childhood maltreatment. Limited functionality of the two risk variants (rs142346759 and rs10262462) was observed (**Table 2**). One of the SNPs in LD for the risk variant on chromosome 3,

rs142346759, obtained a CADD score of greater than 12.37, indicating that this SNP may be deleterious. Six of the SNPs in LD with the risk variant on chromosome 7, rs10262462, had a CADD score of greater than 12.37. No significant eQTLs were identified for either risk locus.

The chromosome 7 variant identified in the meta-analysis, rs1859100, also mapped to the gene *FOXP2* and is located in the same genomic risk locus (chr7:114015707-114287116) as rs10262462. The other hit observed in the meta-analysis, rs917577, was mapped to an intergenic region on chromosome 12. This variant obtained an RDB categorical score of 2B, indicating that it is likely to affect transcription factor binding. No eQTLs exist in the selected tissue types for this region (**Table 2**).

Heritability of Reported Childhood Maltreatment

GWAS summary statistics were used to estimate the SNP-based heritability (h^2_{snp}) of childhood maltreatment with the tool LDSR (**Table 3**). The h^2_{snp} was estimated at 0.057 (p=1.60x10^{-32}) for the UKBB discovery dataset and 0.123 (p=0.002) for PGC1.5. The h^2_{snp} for the meta-analysis was 0.057 (p= 4.48x10^{-46}).

Genetic Correlations of Reported Childhood Maltreatment with Other Traits and Disorders

The r_g for childhood maltreatment between the UKBB and PGC1.5 datasets was 0.63 (p=3.28x10^{-6}). To determine whether there is significant genetic overlap between childhood maltreatment and other traits and disorders, pairwise genetic correlations were calculated using the web-based tool LD Hub. A total of 27 significant correlations (Bonferroni-corrected p-value threshold= 0.05/247=0.0002) were found between childhood maltreatment in the meta-analysis and 247 non-UKBB traits. The top 10 highest genetic correlations are plotted in **Figure 3** with depressive symptoms (r_g=0.70, p=4.65x10^{-40}) having the most significant correlation with childhood maltreatment. There were also positive genetic correlations with "Major depressive disorder" (r_g=0.71, p=4.13x10^{-11}), "PGC cross-disorder analysis" (r_g=0.47, p=1.62x10^{-14}) and "neuroticism" (r_g=0.44, p=1.14x10^{-17}). Significant negative genetic correlations between childhood maltreatment and "age of first birth" (r_g=-0.47, p=2.61x10^{-27}), "subjective well-being" (r_g=-0.46, p=1.00x10^{-18}) and "mother's age at death" (r_g=-0.36, p=7.42x10^{-6}) were also observed.

Discussion

The main findings of this study were that 1) variants located in the genes *FOXP1* and *FOXP2* and on chromosome 12 are significantly associated with childhood maltreatment, 2) the SNP-based estimate of childhood maltreatment is approximately 6%, 3) childhood maltreatment is significantly genetically correlated with "depressive symptoms" and "major depressive disorder", "neuroticism", "age of first birth", and "subjective well-being".

Two genome-wide loci for childhood maltreatment identified in our discovery dataset were also significant in the meta-analysis: rs142346759 (chr3p13), an intronic variant in *FOXP1* and rs10262462 (chr7q31.1) an intronic variant located in *FOXP2*. Both genes form part of the forkhead box superfamily of transcription factors which are widely expressed, and which play important roles during development and adulthood. *FOXP1* and *FOXP2* fall under the FOXP sub-family (also comprising *FOXP3* and *FOXP4*) which has functions in oncogenic and tumour suppressive pathways[38]. *FOXP2* contains highly conserved genomic sites, including an intronic region within this gene, located about 107kb downstream from our risk variant[39]. FOXP1 and FOXP2 have approximately 60% homology at the amino acid level (https://www.ncbi.nlm.nih.gov/books/NBK7023/) and both proteins have been implicated in cognitive disorders, including expressive language impairment[40]. In the meta-analysis, we observed an additional genome-wide variant, located in an intergenic region on chromosome 12, but as this variant does not map to a particular gene, its possible biological mechanism is unclear.

Notably, variation within *FOXP1* has been found to have associations with language impairment, internalizing symptoms, and externalizing symptoms[41]. *FOXP2* has mainly been investigated in regards to speech and language development[42], but has also been found to be associated with depression[43] and attention deficit hyperactivity disorder (ADHD)[44]. Further, an intronic variant in the *FOXP2* gene, rs727644, has been associated with risk taking behavior[45, 46]. While most work on childhood maltreatment has emphasized subsequent risk for mental and physical disorders, it is possible that externalizing behaviors increase risk for childhood trauma[47], consistent with a non-passive rGE mechanism. Alternatively, phenotypes such as depression or neuroticism may increase the likelihood of individuals recalling childhood maltreatment[48, 49].

In this study we estimated SNP-based heritability for childhood maltreatment to be ~6%. A first possibility, in line with a link between *FOXP* variants and externalizing symptoms, is that genetic factors influence environmental factors indirectly through personality and behavior [9]. A second possibility, consistent with a link of *FOXP* variants with internalizing symptoms and depression, is that genetic factors influence the recall of childhood maltreatment. In particular, retrospective assessment of childhood maltreatment may be limited by recall bias and the respondent's subjective assessment of the event[50, 51]. Indeed, a recent systematic review found very low concordance between prospective and retrospective measures of childhood maltreatment[52] and those who retrospectively report childhood adversity are at a greater risk for having psychopathology than those who prospectively reported childhood maltreatment[53].

A twin-based study estimated the heritability of reported childhood maltreatment (comprising physical, and sexual maltreatment and neglect) to be 6%[7], the same as our SNP-based estimate. As twin-based studies capture latent heritability across the entire genome, these heritability estimates are generally higher than SNP-based heritability estimates, which are limited to common variation and by the number of markers present and tagged on the genotyping array used[15]. However, in this twin study, when considering each maltreatment category separately, the heritability of childhood physical maltreatment, sexual maltreatment, and neglect was 28%, 0%, and 24%, respectively. This suggests that only physical abuse and neglect are heritable and that sexual abuse is not genetically influenced. It is notable that these twin data, then, do not support an rGE for some abuse types.

Our finding of positive genetic correlations between childhood maltreatment, depressive symptoms, and major depressive disorder further suggests that the genetic factors predisposing to reporting early life maltreatment overlap with the genetic variation underlying depression. Genetic correlations between depression, stressful life events and lifetime trauma have led to the hypothesis that genes increasing risk for the development of depression predispose individuals to entering into adverse environments[54, 55]. Depressed individuals with and without trauma exposure differ in associated genetic variation, with trauma-exposed individuals having greater SNP-based heritability, supporting this hypothesis[26, 56]. On the other hand, polygenic scores for major depressive disorder (MDD) were associated with greater reporting of stressful life events in individuals with MDD[57].

Indeed, current mood can influence the recall of childhood experiences, and individuals with current depression are at an increased likelihood of reporting early life adversity[58].

In addition to depression, we found significant positive genetic correlations between childhood maltreatment and "neuroticism" and "PGC cross-disorder analysis" (comprised of GWAS summary statistics of five psychiatric disorders: autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, MDD, and schizophrenia). We observed negative genetic correlations of childhood maltreatment with "age of first birth" and "subjective well-being". Associations between early life maltreatment and each of these above traits have previously been observed[56, 59-67]. Further investigation is required to delineate the mechanisms that play a role in the relationship between childhood maltreatment and these outcomes.

Our study had a number of limitations. First, the genetic correlation between the UKBB and PGC1.5 datasets was only 0.63, indicating differences between the datasets, which possibly explains the non-replication of our top hit and of greater SNP heritability in PGC1.5. The UKBB dataset comprises healthy volunteers who are typically of a higher socioeconomic status and in better overall health than the general population of comparable age[68]. Thus, the findings reported here may not be generalizable to the general population. However, it is also worth noting that the top hits were significant in the meta-analysis, where additional hits for childhood maltreatment were detected in an intergenic region on chromosome 12.

Second, although many of the study sites included in the final meta-analysis utilized the well-validated CTQ, childhood maltreatment was measured in a diversity of ways for many of the other studies. Thus, our main phenotype was not homogenous and may reflect different aspects of childhood maltreatment across contributing studies. However, we performed GWAS for each study separately, and meta-analysis was conducted with *p-value* and *direction of effect* to minimize heterogeneity. In addition, based on Cochran's Q-test, none of our significant hits in the meta-analysis showed significant heterogeneity across studies.

This is the first large-scale genetic study to identify specific variants associated with self-reported childhood maltreatment. Variation in *FOXP* genes and other variants associated

with childhood maltreatment may put individuals at greater risk for maltreatment. Alternatively, however, these variants may be associated with a greater likelihood of reporting maltreatment, given their association with depression and neuroticism. With the data available, we are unable to indicate definitively which of these explanations is a better one, and it is possible that different mechanisms have more robust explanatory power in accounting for different abuse subtypes as well as different associated psychopathologies. A clearer understanding of the genetic relationships of childhood maltreatment, including particular abuse subtypes, with a range of different psychiatric disorders, may ultimately be useful in developing targeted treatment and prevention strategies.

Acknowledgements

Conflict of interest

H.R.K. is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative (ACTIVE), which in the last three years was supported by AbbVie, Alkermes, Amygdala Neurosciences, Arbor, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, and Pfizer. **H.R.K** and **J.G.** are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018.

In the past 3 years, **D.J.S.** has received research grants and/or consultancy honoraria from Lundbeck and Sun.

In the past 3 years, **R.C.K**. received support for his epidemiological studies from Sanofi Aventis; was a consultant for Johnson & Johnson Wellness and Prevention, Sage

Pharmaceuticals, Shire, Takeda; and served on an advisory board for the Johnson & Johnson. Services Inc. Lake Nona Life Project. Kessler is a co-owner of DataStat, Inc., a market research firm that carries out healthcare research.

M.B.S. has in the past three years been a consultant for Aptinyx, Bionomics, Dart Neuroscience, Janssen, Jazz Pharmaceuticals, Neurocrine Biosciences, Oxeia Biopharmaceuticals, and Pfizer.

R.Y. is a co-inventor of the following patent application: "Genes associated with posttraumatic-stress disorder. European Patent# EP 2334816 B1

References

1.   Child maltreatment.Global estimates. http://apps.who.int/gho/data/view.main.VIOLENCECHILDMALTREATMENTv, 2016, Accessed Date Accessed 2016 Accessed.

2.   McLaughlin KA, Koenen KC, Bromet EJ, Karam EG, Liu H, Petukhova M *et al.* Childhood adversities and post-traumatic stress disorder: evidence for stress sensitisation in the World Mental Health Surveys. *The British journal of psychiatry : the journal of mental science* 2017; **211**(5)**:** 280-288.

3.   Copeland WE, Shanahan L, Hinesley J, Chan RF, Aberg KA, Fairbank JA *et al.* Association of childhood trauma exposure with adult psychiatric disorders and functional outcomes. *JAMA network open* 2018; **1**(7)**:** e184493-e184493.

4.   Basu A, McLaughlin KA, Misra S, Koenen KC. Childhood Maltreatment and Health Impact: The Examples of Cardiovascular Disease and Type 2 Diabetes Mellitus in Adults. *Clinical psychology : a publication of the Division of Clinical Psychology of the American Psychological Association* 2017; **24**(2)**:** 125-139.

5.   Suglia SF, Koenen KC, Boynton-Jarrett R, Chan PS, Clark CJ, Danese A *et al.* Childhood and Adolescent Adversity and Cardiometabolic Outcomes: A Scientific Statement From the American Heart Association. *Circulation* 2018; **137**(5)**:** e15-e28.

6.   Koenen KC, Roberts AL, Stone DM, Dunn EC. The epidemiology of early childhood trauma. *The impact of early life trauma on health and disease: The hidden epidemic* 2010; **1**.

7.   Schulz-Heik RJ, Rhee SH, Silvern L, Lessem JM, Haberstick BC, Hopfer C *et al.* Investigation of genetically mediated child effects on maltreatment. *Behav Genet* 2009; **39**(3)**:** 265-276.

8.   Knafo A, Jaffee SR. Gene-environment correlation in developmental psychopathology. *Development and psychopathology* 2013; **25**(1)**:** 1-6.

9.   Jaffee SR, Price TS. Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular psychiatry* 2007; **12**(5)**:** 432.

10. Rutter M. Gene–environment interdependence. *Developmental science* 2007; **10**(1)**:** 12-18.

11. Boivin M, Perusse D, Dionne G, Saysset V, Zoccolillo M, Tarabulsy GM *et al.* The genetic-environmental etiology of parents' perceptions and self-assessed behaviours toward their 5-month-old infants in a large twin and singleton sample. *Journal of child psychology and psychiatry, and allied disciplines* 2005; **46**(6)**:** 612-630.

12. Berlin LJ, Dodge KA, Reznick JS. Examining pregnant women's hostile attributions about infants as a predictor of offspring maltreatment. *JAMA pediatrics* 2013; **167**(6)**:** 549-553.

13. Rehan W, Antfolk J, Johansson A, Aminoff M, Sandnabba NK, Westberg L *et al.* Gene-Environment Correlation Between the Dopamine Transporter Gene (DAT1) Polymorphism and Childhood Experiences of Abuse. *Journal of interpersonal violence* 2018; **33**(13)**:** 2059-2072.

14. Munafo MR, Flint J. Dissecting the genetic architecture of human personality. *Trends in cognitive sciences* 2011; **15**(9)**:** 395-400.

15. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry* 2017.

16. van der Merwe C, Jahanshad N, Cheung JW, Mufford M, Groenewold NA, Koen N *et al.* Concordance of genetic variation that increases risk for anxiety disorders and posttraumatic stress disorders and that influences their underlying neurocircuitry. *Journal of affective disorders* 2019; **245:** 885-896.

17. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen C-Y, Choi KW *et al.* Largest genome-wide association study for PTSD identifies genetic risk loci in European and African ancestries and implicates novel biological pathways. *bioRxiv* 2018**:** 458562.

18. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: come and get it. American Association for the Advancement of Science2014.

19. Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T *et al.* Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child abuse & neglect* 2003; **27**(2)**:** 169-190.

20. Bernstein DP, Fink L. *Childhood trauma questionnaire: A retrospective self-report: Manual*. Harcourt Brace & Company1998.

21. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *bioRxiv* 2019**:** 587196.

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* 2006; **38**(8)**:** 904.

23. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods* 2013; **10**(1)**:** 5.

24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* 2009; **5**(6)**:** e1000529.

25. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526**(7571)**:** 68.

26. Coleman J, Purves K, Davis K, Rayner C, Choi S, Hübel C. Genome-wide gene-environment analyses of depression and reported lifetime traumatic experiences in UK Biobank. bioRxiv. November2018.

27. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**(10)**:** 1279-1283.

28. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S *et al.* The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**(7571)**:** 82-90.

29. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* 2017; **49**(3)**:** 403-415.

30. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* 2017; **33**(17)**:** 2776-2778.

31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**(1)**:** 7.

32. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726)**:** 203.

33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17)**:** 2190-2191.

34. Schwarzer G. meta: An R package for meta-analysis. *R news* 2007; **7**(3)**:** 40-45.

35. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* 2017; **8**(1)**:** 1826.

36. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* 2015; **47**(3)**:** 291.

37. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* 2015; **47**(11)**:** 1236.

38. Lam EW-F, Brosens JJ, Gomes AR, Koo C-Y. Forkhead box proteins: tuning forks for transcriptional harmony. *Nature reviews Cancer* 2013; **13**(7)**:** 482.

39. Atkinson EG, Audesse AJ, Palacios JA, Bobo DM, Webb AE, Ramachandran S *et al.* No evidence for recent selection at FOXP2 among diverse human populations. *Cell* 2018; **174**(6)**:** 1424-1435. e1415.

40. Bacon C, Rappold GA. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Human genetics* 2012; **131**(11)**:** 1687-1698.

41. Siper PM, De Rubeis S, del Pilar Trelles M, Durkin A, Di Marino D, Muratet F *et al.* Prospective investigation of FOXP1 syndrome. *Molecular autism* 2017; **8**(1)**:** 57.

42. Fisher SE, Scharff C. FOXP2 as a molecular window into speech and language. *Trends in Genetics* 2009; **25**(4)**:** 166-177.

43. Arnau-Soler A, Macdonald-Dunlop E, Adams MJ, Clarke T-K, MacIntyre DJ, Milburn K *et al.* Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational psychiatry* 2019; **9**(1)**:** 14.

44. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* 2019; **51**(1)**:** 63.

45. Clifton EA, Perry JR, Imamura F, Lotta LA, Brage S, Forouhi NG *et al.* Genome–wide association study for risk taking propensity indicates shared pathways with body mass index. *Communications biology* 2018; **1**(1)**:** 36.

46. Strawbridge RJ, Ward J, Lyall LM, Tunbridge EM, Cullen B, Graham N *et al.* Genetics of self-reported risk-taking behaviour, trans-ethnic consistency and relevance to brain gene expression. *Translational psychiatry* 2018; **8**(1)**:** 178.

47. Cui N, Liu J. Cognitive and behavioral risk factors for child physical abuse among Chinese children: a multiple-informant study. *Child and adolescent psychiatry and mental health* 2016; **10**(1)**:** 36.

48. Bernet CZ, Stein MB. Relationship of childhood maltreatment to the onset and course of major depression in adulthood. *Depression and anxiety* 1999; **9**(4)**:** 169-174.

49. Stein MB, Jang KL, Taylor S, Vernon PA, Livesley WJ. Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: a twin study. *The American journal of psychiatry* 2002; **159**(10)**:** 1675-1681.

50. Widom CS, Morris S. Accuracy of adult recollections of childhood victimization, part 2: childhood sexual abuse. *Psychological assessment* 1997; **9**(1)**:** 34.

51. Hardt J, Rutter M. Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *Journal of child psychology and psychiatry, and allied disciplines* 2004; **45**(2)**:** 260-273.

52. Baldwin JR, Reuben A, Newbury JB, Danese A. Agreement between prospective and retrospective measures of childhood maltreatment: a systematic review and meta-analysis. *JAMA psychiatry* 2019.

53. Newbury JB, Arseneault L, Moffitt TE, Caspi A, Danese A, Baldwin JR *et al.* Measuring childhood maltreatment to predict early-adult psychopathology: comparison of prospective informant-reports and retrospective self-reports. *Journal of psychiatric research* 2018; **96:** 57-64.

54. Kendler KS, Kessler RC, Neale MC, Heath AC, Eaves LJ. The prediction of major depression in women: toward an integrated etiologic model. *American Journal of Psychiatry* 1993; **150:** 1139-1139.

55. Kendler KS, Karkowski-Shuman L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychological medicine* 1997; **27**(3)**:** 539-547.

56. Peterson RE, Cai N, Dahl AW, Bigdeli TB, Edwards AC, Webb BT *et al.* Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. *American Journal of Psychiatry* 2018; **175**(6)**:** 545-554.

57. Mullins N, Power R, Fisher H, Hanscombe K, Euesden J, Iniesta R *et al.* Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological medicine* 2016; **46**(4)**:** 759-770.

58. Colman I, Kingsbury M, Garad Y, Zeng Y, Naicker K, Patten S *et al.* Consistency in adult reporting of adverse childhood experiences. *Psychol Med* 2016; **46**(3)**:** 543-549.

59. Rosenman S, Rodgers B. Childhood adversity and adult personality. *The Australian and New Zealand journal of psychiatry* 2006; **40**(5)**:** 482-490.

60. Pesonen AK, Raikkonen K, Heinonen K, Kajantie E, Forsen T, Eriksson JG. Reproductive traits following a parent-child separation trauma during childhood: a natural experiment during World War II. *American journal of human biology : the official journal of the Human Biology Council* 2008; **20**(3)**:** 345-351.

61. Matheson SL, Shepherd AM, Pinchbeck RM, Laurens KR, Carr VJ. Childhood adversity in schizophrenia: a systematic meta-analysis. *Psychol Med* 2013; **43**(2)**:** 225-238.

62. Palmier-Claus JE, Berry K, Bucci S, Mansell W, Varese F. Relationship between childhood adversity and bipolar affective disorder: systematic review and meta-

analysis. *The British journal of psychiatry : the journal of mental science* 2016; **209**(6)**:** 454-459.

63. Kerns CM, Newschaffer CJ, Berkowitz S, Lee BK. Brief Report: Examining the Association of Autism and Adverse Childhood Experiences in the National Survey of Children's Health: The Important Role of Income and Co-occurring Mental Health Conditions. *Journal of autism and developmental disorders* 2017; **47**(7)**:** 2275-2281.

64. Bjorkenstam E, Bjorkenstam C, Jablonska B, Kosidou K. Cumulative exposure to childhood adversity, and treated attention deficit/hyperactivity disorder: a cohort study of 543 650 adolescents and young adults in Sweden. *Psychol Med* 2018; **48**(3)**:** 498-507.

65. Corcoran M, McNulty M. Examining the role of attachment in the relationship between childhood adversity, psychological distress and subjective well-being. *Child Abuse Negl* 2018; **76:** 297-309.

66. Mandelli L, Petrelli C, Serretti A. The role of specific early trauma in adult depression: a meta-analysis of published literature. Childhood trauma and adult depression. *European psychiatry* 2015; **30**(6)**:** 665-680.

67. Li M, D'arcy C, Meng X. Maltreatment in childhood substantially increases the risk of adult depression and anxiety in prospective cohort studies: systematic review, meta-analysis, and proportional attributable fractions. *Psychological medicine* 2016; **46**(4)**:** 717-730.

68. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American journal of epidemiology* 2017; **186**(9)**:** 1026-1034.
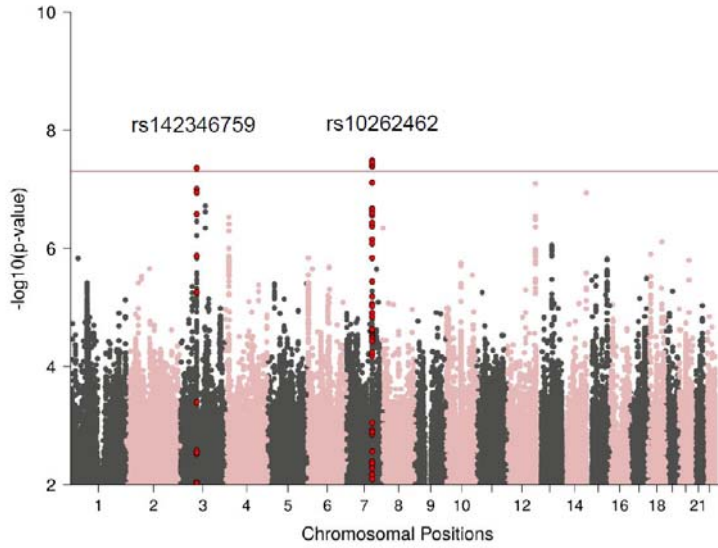
Figure 1: Manhattan plot of UKBB GWAS for childhood maltreatment, showing the top variants. The horizontal line represents genome-wide significance at $p<5\times10^{-8}$.
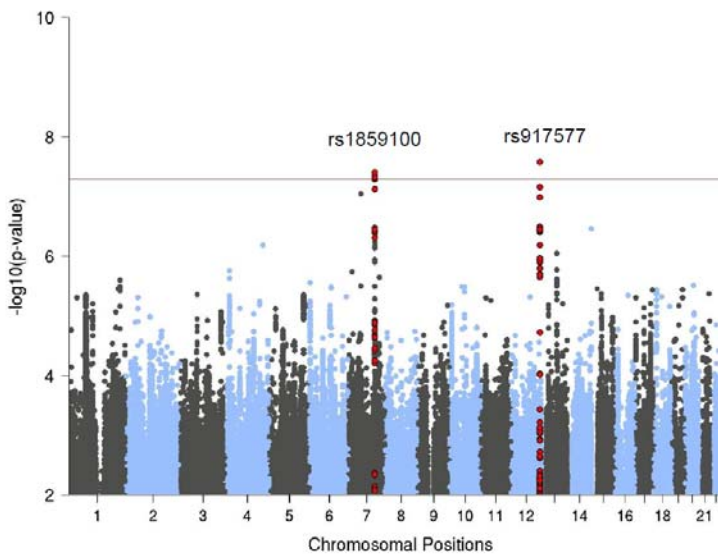

Figure 2: Manhattan plot of the meta-analysis results for childhood maltreatment, showing the top variants. The horizontal line represents genome-wide significance at $p<5\times10^{-8}$.
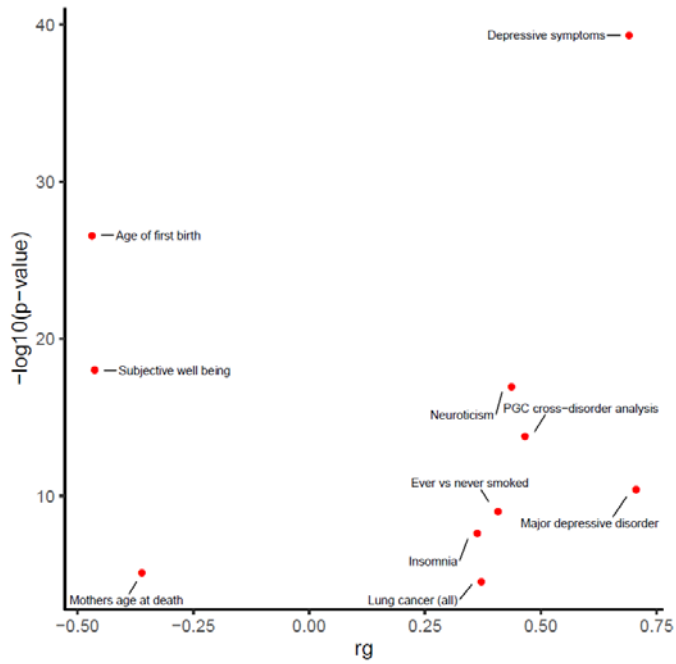
Figure 3: Top ten genetic correlations between several groups of traits (from psychiatric, anthropomorphic, smoking behavior, reproductive, aging, education, autoimmune, and cardio-metabolic categories) and childhood maltreatment (meta-analysis).