This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: https://orca.cardiff.ac.uk/id/eprint/144840/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:


Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Enhancing Discovery of Genetic Variants for Posttraumatic Stress Disorder Through Integration of Quantitative Phenotypes and Trauma Exposure Information

Posttraumatic stress disorder (PTSD) may develop after exposure to traumatic life events. PTSD can severely impact the mental and physical health of affected individuals and impair their interpersonal relationships (1). While the estimated community prevalence of PTSD in the United States is 5% to 10% (2), the rate of PTSD differs based on the nature of trauma exposure (3) and other environmental (4) and genetic (5–7) factors. Identifying the biological mechanisms associated with the etiology of PTSD will facilitate the discovery of biomarkers for screening and diagnostic purposes (7) and the development of new treatments.

Genome-wide association studies (GWASs) facilitate biological understanding of PTSD (8,9), but are well known to be limited by statistical power to identify risk variation (10). Quantitative measures of PTSD enhance discovery power over binary trait definitions (9,11). Appropriately accounting for trauma exposure hypothetically enhances power, as individuals will not develop the disorder unless they are exposed to trauma, regardless of high genetic vulnerability for PTSD (12,13). Moreover, the notion that genetic variants can predispose to trauma exposure is only starting to be explored (14).

As trauma exposure is a prerequisite for the development and manifestation of PTSD, investigating the genetics of trauma exposure will hypothetically lead to a clearer picture of PTSD genetics.

The Psychiatric Genomics Consortium (PGC)–PTSD is a global collaborative effort to study the genetic basis of PTSD through meta-analysis of diverse cohorts (13). Subsequent to a case-control GWAS (8), our collaborators have provided quantitative measures of PTSD and lifetime trauma exposure (LTE). To obtain genomic insights from the quantitative PTSD phenotyping, we performed a GWAS of PTSD symptoms in 182,199 participants from the PGC-PTSD Freeze 2 dataset. To determine if accounting for LTE would provide the hypothesized increase in discovery power, we performed a GWAS of PTSD with covariate adjustment for LTE, showing that it lowers PTSD signal. We investigated the possibility that multicollinearity arising from high genetic correlation (r_g) of PTSD and LTE was responsible for this result. To perform this investigation, we performed a GWAS of LTE in the most powered and unbiased (15) subsample of the data, 132,988 participants from the UK Biobank (UKBB) (16), then evaluated the r_g of PTSD and LTE. To explore the r_g further, we contrasted the r_g that PTSD and LTE have with other traits. We showed that the high r_g of PTSD and LTE can be leveraged to enhance the power of PTSD GWASs using multivariate methods. We replicated PTSD GWAS findings in the Million Veteran Program (MVP) GWAS of total PTSD symptoms (MVP_TOT). We contextualized genomic findings through functional annotation, tissue expression analyses, and phenome-wide association study (PheWAS).

METHODS AND MATERIALS

Study Population and Phenotyping

Participants were drawn from a collection of 51 cohorts within the PGC-PTSD Freeze 2 dataset, as previously described in Nievergelt et al. (8). All participants included in the present study were of genetically estimated European ancestry. PTSD symptoms and LTE were measured within each cohort using structured clinical interviews, self-reported inventories, or clinical evaluation. A summary of the assessment and scoring methods for the various studies is presented in Table S1 in Supplement 2, and a complete description is available in Nievergelt et al. (8). All participants provided written informed consent, and studies were approved by the relevant institutional review boards.

Four of these 9 PTSD loci were independently replicated in the Million Veteran Program GWAS of PTSD total symptoms. To determine if accounting for LTE would provide the hypothesized increase in discovery power, we performed a GWAS of PTSD with covariate adjustment for LTE, showing that it lowers PTSD signal. We investigated the possibility that multicollinearity arising from high genetic correlation (r_g) of PTSD and LTE was responsible for this result. To perform this investigation, we performed a GWAS of LTE in the most powered and unbiased (15) subsample of the data, 132,988 participants from the UK Biobank (UKBB) (16), then evaluated the r_g of PTSD and LTE. To explore the r_g further, we contrasted the r_g that PTSD and LTE have with other traits. We showed that the high r_g of PTSD and LTE can be leveraged to enhance the power of PTSD GWASs using multivariate methods. We replicated PTSD GWAS findings in the Million Veteran Program (MVP) GWAS of total PTSD symptoms (MVP_TOT). We contextualized genomic findings through functional annotation, tissue expression analyses, and phenome-wide association study (PheWAS).
institutional review boards and the University of California San Diego Human Research Protection Program.

**GWAS Quality Control**

Genotyping, quality control (QC), and imputation methods for the included studies have been described in detail (8). In brief, participating cohorts provided phenotype and genotype data or GWAS summary statistics to the PGC-PTSD for quality control and analysis. For studies in which the PGC-PTSD analyst had direct access to genotype data, RICOPILI (17) was used to perform QC and imputation. QC included standard filters for single nucleotide polymorphism (SNP) call rates (exclusion of SNPs with call rate <98% or a missing difference >0.02 between cases and controls), call rate for participant genotypes (samples with <98% call rate excluded), Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$ in controls), and heterozygosity (within £ 0.2). Datasets were phased using SHAPEIT (18) and imputed using IMPUTE2 (19) with the 1000 Genomes Phase 3 reference panel data (20). For the UKBB, QC and imputation were carried out centrally by UKBB investigators as previous described (18) and GWAS was carried out by the PGC-PTSD analyst. For cohorts with data-sharing restrictions, analyses were performed using similar protocols by the study team that had individual-level data access, and GWAS summary statistics were provided to the PGC-PTSD.

**Genome-wide Association Study**

Only unrelated ($r < 0.2$) participants were retained for analysis. Principal components (PCs) were calculated within each cohort using EIGENSOFT v6.3.4 (21). The PTSD GWAS was performed within cohorts using PLINK 2.0 alpha with the –glm option, with the exception of UKBB and VETSA (Vietnam Era Twin Study of Aging) data, which were analyzed using BOLT-LMM v2.3.4 (22). Where available, PTSD symptom scores were analyzed using linear regression ($n = 36$ cohorts); PTSD case-control status was used if symptom scores were not available, using logistic regression ($n = 15$ cohorts). In both cases, 5 PCs were included as covariates to account for population stratification and genotyping artifacts. The UKBB PTSD GWAS included an additional PC as well as batch and assessment center covariates. Studies providing summary data used similar analytic strategies, as previously described (8). For each GWAS, SNPs with minor allele frequency <1% or imputation information score $<0.6$ were excluded. To perform a GWAS of PTSD conditioned on LTE, the GWAS was performed with LTE included as an additional covariate as either a count of LTEs or a binary variable, depending on data availability. The GWAS of the LTE count phenotype in the UKBB sample was performed in BOLT-LMM using 6 PCs, batch, and assessment center as covariates.

**PTSD Meta-analysis**

Sample size–weighted fixed-effects meta-analysis was performed using METAL (23). To account for different analytic methods and measure scales, effect estimates were converted into z scores by dividing effect sizes by standard errors (24). Case-control and quantitative GWAS subsets were evaluated for $r_g$ to determine if they could be meta-analyzed. To account for differences in ascertainment, heritability, and power between case-control and quantitative subsets, modified sample size weights were derived as previously described (25), assuming 10% population prevalence of PTSD, the estimates of SNP-based heritability ($h^2_{SNP}$), $r_g$, and sample PTSD prevalence. Meta-analysis was conducted on the reweighted $z$ scores. Only SNPs available in >90% of all samples ($N \geq 163,979$) were included in analyses. Regional annotation plots of genome-wide significant loci were produced using Locus-Zoom (26).

**Heritability and Genetic Correlation Estimation With Linkage Disequilibrium Score Regression**

Trait $h^2_{SNP}$ and $r_g$ were estimated from GWAS summary statistics using linkage disequilibrium score regression (27). The linkage disequilibrium score intercept was used to test for inflation of test statistics owing to residual population stratification or other artifacts, and the attenuation factor $\left[\frac{\text{Intercept} - 1}{\text{mean}(\nu^2) - 1}\right]$ was used to determine the proportion of inflation of test statistics owing to residual population stratification (Table S2 in Supplement 2). Heritabilities were contrasted using a z test where standard errors were estimated using the block-jackknife approach. To estimate $r_g$ with other disorders, the LD Hub web interface was used (28). To identify genetic differences between PTSD and LTE, the $r_g$s observed for PTSD and LTE were contrasted using $z$ tests, where significance level was determined using Bonferroni correction for the 772 traits tested ($\alpha < 6.47 \times 10^{-5}$).

**FUMA**

FUMA v1.3.6a (29) was used with the default settings (Supplement 1) to visualize and annotate GWAS results. The FUMA pipeline integrates the MAGMA (30) tool to perform gene-based, gene-pathway, and tissue-enrichment analyses, with significance based on Bonferroni correction. 1000 Genomes Europeans were used as reference genotypes. Tissue-enrichment analysis included Genotype-Tissue Expression (GTEx) v8 expression data (31).

**Cis-Quantitative Trait Locus Mapping**

The effects of GWAS loci on transcriptionic regulation of surrounding genes (locus within $\pm 1$ Mb of the gene transcription starting site) were tested for 49 tissues in GTEx v8 with genome-wide false discovery rate correction applied. Using the same criteria, GTEx v8 data were also used to investigate the effects of GWAS loci on the regulation of alternative splicing isoforms. A detailed description regarding GTEx v8 quantitative trait locus (QTL) mapping data by the GTEx Consortium is available (32). Briefly, cis-expression QTL (eQTL) and cis-splicing QTL mapping was performed using FastQTL (33) including the top 5 genotyping PCs, probabilistic estimation of expression residuals factors (34), sequencing platform, sequencing protocol, and sex as covariates.

**Replication Analysis**

Summary data from MVP_TOT (dbGaP study accession phs001672/v4.p1) was used to replicate GWAS results. MVP_TOT included 186,689 European ancestry participants who completed the PTSD Checklist–Civilian Version and passed QC. Details of MVP_TOT have been published (35). SNPs were
deemed replicated in MVP\(_{\text{TOT}}\) if they had matching effect direction and were nominally significant after Bonferroni correction for the 9 SNPs tested (\(p < .006\)).

**Multi-Trait Analysis of GWAS**

Multi-Trait Analysis of GWAS (MTAG) (36) performs multivariate analysis of genetically correlated traits to increase discovery power for each input trait, providing trait-specific effect estimates and \(p\) values. MTAG was used to perform multivariate analysis with PTSD and LTE GWASs. The maxFDR statistic was used to test for MTAG model assumptions (Supplement 1).

**Phenome-wide Association Study**

To understand further how functional changes of significant loci are associated with human traits and diseases, we conducted a PhenWAS of leading SNPs from PTSD and LTE using data from the GWAS Atlas (available at https://atlas.ctgibl.nl/) (37). Bonferroni correction was applied to account for the 4756 phenotypes available that were tested (\(p < 1.05 \times 10^{-5}\)).

**RESULTS**

The PTSD GWAS meta-analysis included 182,199 participants of European ancestry from 51 cohorts (Table S1 in Supplement 2). The largest cohort was the UKBB (\(N = 134,586\) participants). Across the cohorts, PTSD was assessed using a variety of different methods (\(n = 19\) methods); the most common methods were versions of the Clinician-Administered PTSD Scale (\(n = 18\) studies) and PTSD Checklist (\(n = 14\) studies). The majority of participants (91%, \(n = 165,825\), 36 studies) were analyzed based on PTSD symptom scores; the remaining participants (9%, \(n = 16,374\), 15 studies) did not have symptom scores available and were analyzed based on PTSD case-control status.

**PGC-PTSD GWAS Meta-analysis**

The \(h^2_{\text{SNP}}\) of meta-analysis of cohorts analyzed by symptom scores was 0.0547 (SE = 0.0042, \(p = 8.9 \times 10^{-39}\)) (Table S2 in Supplement 2). The \(h^2_{\text{SNP}}\) was similar, albeit not significant, in the smaller meta-analysis of case-control cohorts (observed scale \(h^2_{\text{SNP}} = 0.0580\), SE = 0.0259, \(p = 17\)). The \(r_g\) between the symptom score and case-control analyses was very high (\(r_g = 0.9646\), SE = 0.36, \(p = 0.0074\)). Thus, symptom score and case-control GWASs were meta-analyzed. We identified 5 genome-wide significant loci (Table 1, Figure 1A). Leading variants in significant loci mapped to an intergenic locus on chromosome 1, the intronic region of the GABBR1 gene on chromosome 6, the intronic regions of MMP6 and DFNAS on chromosome 7, an intron of FOXP2 on chromosome 7, and the intronic region of FAM120A on chromosome 9. Gene-based analysis identified 6 significant genes (DCAF5, EXD2, FAM120A, FOXP2, GALNT16, and PHF2) (Table S3 in Supplement 2).

**PGC-PTSD GWAS Covariate Adjustment for LTE**

We repeated the GWAS of PTSD with covariate adjustment for LTE. \(h^2_{\text{SNP}}\) was 0.0389 (SE = 0.00340, \(p = 2.6 \times 10^{-30}\)), 31% lower than the PTSD GWAS without LTE covariate adjustment (\(p = 8.6 \times 10^{-29}\)). There was a genome-wide significant locus in an uncharacterized region, CTC-340A15.2, on chromosome 5 that was not identified in the PTSD GWAS (Table S4 in Supplement 2). Effects changed slightly for the loci previously identified in the unadjusted PTSD GWAS (Table S4 in Supplement 2). Gene-based analysis identified no significant genes.

**UKBB LTE GWAS**

We performed GWAS of LTE count in the UKBB subset of the PGC-PTSD GWAS data (132,988 UKBB participants). Of participants, 30.9% reported 1 LTE, 14.8% reported 2 LTEs, 6.3% reported 3 LTEs, and 3.3% reported 4 or more LTEs (Table S5 in Supplement 2). The \(h^2_{\text{SNP}}\) of LTE count was 0.0734 (SE = 0.005, \(p = 8.7 \times 10^{-48}\)). Six loci showed genome-wide significance (Figure 1B, Table 2). Leading variants in significant loci mapped to an intron of PRUNE on chromosome 1, the intron of noncoding RNA AC068490.2 on chromosome 2, the intronic region of the FOXP2 gene on chromosome 7, and the intronic regions of SGCd on chromosome 5, an intron of FOXP2 on chromosome 7 (also identified in the PGC-PTSD GWAS), an intergenic region in chromosome 14 near MDGA, and

### Table 1. Genome-wide Significant Loci From PTSD GWASs and MTAGs With Replication in MVP\(_{\text{TOT}}\), GWAS

<table>
<thead>
<tr>
<th>Analysis</th>
<th>rsID</th>
<th>Chr</th>
<th>Position(^a)</th>
<th>A1</th>
<th>A2</th>
<th>A1 Freq</th>
<th>P Value</th>
<th>Score</th>
<th>z Score</th>
<th>p Value</th>
<th>A1 freq</th>
<th>z Score</th>
<th>p Value(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified in GWAS</td>
<td>rs72657988</td>
<td>1</td>
<td>35668541</td>
<td>T</td>
<td>G</td>
<td>0.08</td>
<td>6.44</td>
<td>12.10</td>
<td>(10^{-10})</td>
<td>5.34</td>
<td>9.4 (\times) 10^{-8}</td>
<td>0.07</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>rs146918648</td>
<td>6</td>
<td>28548674</td>
<td>A</td>
<td>G</td>
<td>0.04</td>
<td>6.04</td>
<td>15.90</td>
<td>(10^{-9})</td>
<td>6.50</td>
<td>8.0 (\times) 10^{-11}</td>
<td>0.04</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>rs27218161</td>
<td>7</td>
<td>24699329</td>
<td>A</td>
<td>G</td>
<td>0.82</td>
<td>-5.27</td>
<td>1.4 (\times) 10^{-7}</td>
<td>-5.80</td>
<td>6.5 (\times) 10^{-9}</td>
<td>0.82</td>
<td>-1.45</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>rs10266297</td>
<td>7</td>
<td>114143407</td>
<td>C</td>
<td>T</td>
<td>0.59</td>
<td>5.38</td>
<td>7.4 (\times) 10^{-6}</td>
<td>6.72</td>
<td>1.8 (\times) 10^{-11}</td>
<td>0.59</td>
<td>4.97</td>
<td>6.7 (\times) 10^{-7}</td>
</tr>
<tr>
<td></td>
<td>rs10821140</td>
<td>11</td>
<td>96253169</td>
<td>A</td>
<td>C</td>
<td>0.35</td>
<td>-5.71</td>
<td>1.2 (\times) 10^{-6}</td>
<td>-6.02</td>
<td>1.8 (\times) 10^{-9}</td>
<td>0.34</td>
<td>-3.89</td>
<td>1.0 (\times) 10^{-4}</td>
</tr>
<tr>
<td>Identified in MTAG</td>
<td>rs4557006</td>
<td>2</td>
<td>22443840</td>
<td>A</td>
<td>G</td>
<td>0.45</td>
<td>4.26</td>
<td>2.0 (\times) 10^{-5}</td>
<td>5.83</td>
<td>5.7 (\times) 10^{-9}</td>
<td>0.45</td>
<td>5.53</td>
<td>3.2 (\times) 10^{-8}</td>
</tr>
<tr>
<td></td>
<td>rs1504930</td>
<td>5</td>
<td>155852066</td>
<td>A</td>
<td>C</td>
<td>0.62</td>
<td>-4.26</td>
<td>2.0 (\times) 10^{-5}</td>
<td>-5.58</td>
<td>2.5 (\times) 10^{-8}</td>
<td>0.62</td>
<td>-2.40</td>
<td>2.7 (\times) 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs8059002</td>
<td>16</td>
<td>25417390</td>
<td>T</td>
<td>G</td>
<td>0.86</td>
<td>-4.43</td>
<td>9.3 (\times) 10^{-6}</td>
<td>-5.46</td>
<td>4.8 (\times) 10^{-8}</td>
<td>0.85</td>
<td>-1.50</td>
<td>(0.13)</td>
</tr>
<tr>
<td></td>
<td>rs7264419</td>
<td>20</td>
<td>47701309</td>
<td>A</td>
<td>G</td>
<td>0.75</td>
<td>-5.06</td>
<td>4.1 (\times) 10^{-7}</td>
<td>-5.85</td>
<td>5.0 (\times) 10^{-9}</td>
<td>0.76</td>
<td>0.55</td>
<td>(0.58)</td>
</tr>
</tbody>
</table>

\(^a\)Base pair position on chromosome (hg19/GR37 Human Genome Build).

\(^b\)Significant in MVP if \(p < .006\) (Bonferroni-corrected for 9 loci).

\(^c\)Linkage disequilibrium proxy for rs2721817, the leading single nucleotide polymorphism in this locus.
upstream of CCDC8 on chromosome 19. Gene-based analysis identified SGCD (chromosome 5: 155,297,354–156,194,799 base pairs, 2965 SNPs, 99 parameters, \( z = 5.53, p = 1.5 \times 10^{-8} \)) and C20orf112 (chromosome 20:31,030,862–31,172,876 base pairs, 296 SNPs, 21 parameters, \( z = 4.73, p = 1.13 \times 10^{-6} \)). GWAS of LTE count weighted by trauma-specific PTSD.

**Table 2. Genome-wide Significant Loci From GWASs of LTE**

<table>
<thead>
<tr>
<th>rsID</th>
<th>Chr</th>
<th>Position</th>
<th>A1</th>
<th>A2</th>
<th>A1 Frequency</th>
<th>z Score</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6661135</td>
<td>1</td>
<td>150999414</td>
<td>C</td>
<td>T</td>
<td>0.93</td>
<td>−5.52</td>
<td>3.3 \times 10^{-8}</td>
</tr>
<tr>
<td>rs4665501</td>
<td>2</td>
<td>22546151</td>
<td>G</td>
<td>T</td>
<td>0.44</td>
<td>−5.77</td>
<td>7.7 \times 10^{-9}</td>
</tr>
<tr>
<td>rs4704792</td>
<td>5</td>
<td>155757946</td>
<td>A</td>
<td>T</td>
<td>0.26</td>
<td>5.75</td>
<td>9.2 \times 10^{-9}</td>
</tr>
<tr>
<td>rs1476535</td>
<td>7</td>
<td>114071035</td>
<td>C</td>
<td>T</td>
<td>0.44</td>
<td>−5.77</td>
<td>8.0 \times 10^{-9}</td>
</tr>
<tr>
<td>rs2933196</td>
<td>14</td>
<td>47285917</td>
<td>G</td>
<td>A</td>
<td>0.59</td>
<td>−5.51</td>
<td>3.6 \times 10^{-8}</td>
</tr>
<tr>
<td>rs770444611</td>
<td>19</td>
<td>46917381</td>
<td>INS</td>
<td>T</td>
<td>0.59</td>
<td>5.66</td>
<td>1.5 \times 10^{-8}</td>
</tr>
</tbody>
</table>

A1, allele 1 (coded); A2, allele 2; Chr, chromosome; GWAS, genome-wide association study; LTE, lifetime trauma exposure; rsID, reference SNP ID number.

- Base pair position on chromosome (hg19/GR37 Human Genome Build).
- Insertion of TGAGGCCAGGAGTTC.

**Figure 1.** Manhattan plots of genome-wide association study (GWAS) associations. The x-axis is the position on the genome, ordered by chromosome and base-pair position. The y-axis is the \(-\log_{10} p\) value of association. Each dot represents the association between a given single nucleotide polymorphism and the trait. Colors alternate between chromosomes, with odd chromosomes colored blue and even chromosomes colored teal. (A) Results of posttraumatic stress disorder GWASs. (B) Results of lifetime trauma exposure GWASs. (C) Posttraumatic stress disorder–specific results of MTAG (Multi-Trait Analysis of GWAS) analysis of posttraumatic stress disorder and lifetime trauma exposure.
Compared to PTSD and LTE loci included height and body mass phenotypes, educational attainment, social interaction, sexual activity, risk tolerance, and sleep phenotypes (Supplement 1). Several PTSD loci showed widespread pleiotropy across multiple psychiatric traits: rs10266297 (35 significant

Extended Phenotyping Identifies PTSD Genetic Variants

**Figure 2.** Comparison of the genetic correlations of posttraumatic stress disorder (PTSD) and lifetime trauma exposure (LTE) with other traits. The x-axis is the genetic correlation between PTSD and a given trait from the LD Hub. The y-axis is the genetic correlation between PTSD and LTE after Benferroni adjustment. Noncolored (gray) dots indicate traits where genetic correlation with PTSD than with LTE (e.g., neuroticism score \( r_g = 1.09 \)). Twelve traits showed significantly higher correlation with PTSD than with LTE (e.g., neuroticism score \( r_g = 1.09 \)). Conversely, risk taking showed a significantly higher correlation with LTE than PTSD (e.g., neuroticism score \( r_g = 1.09 \)). There was no significant enrichment detected in any sets from the list of curated gene sets and Gene Ontology terms (Table S11 in Supplement 2).

**Genetic Overlap Between LTE and PTSD**

The \( r_g \) between PTSD and LTE was high \( r_g = 0.7239, p < 1 \times 10^{-10} \). To explore this genetic overlap, we contrasted patterns of \( r_g \) of PTSD and LTE to other traits. Testing 772 human traits and diseases, we observed 269 and 217 \( r_g \)s that survived Bonferroni multiple testing correction (\( p < 6.47 \times 10^{-4} \)) for PTSD and LTE, respectively (Table S6 in Supplement 2). There was complete directional concordance between PTSD and LTE among the 187 \( r_g \)s that were significant in both analyses. For several traits, while the effect direction was concordant, the magnitude of correlation with PTSD was significantly different from the correlation with LTE (\( p < 6.47 \times 10^{-5} \)) (Figure 2). Fifteen traits showed significantly higher genetic correlation with PTSD than with LTE (e.g., neuroticism score \( p = 2.74 \times 10^{-24}; \) fed-up feelings \( p = 1.83 \times 10^{-15}; \) mood swings \( p = 9.92 \times 10^{-15}; \) loneliness \( p = 8.07 \times 10^{-8}; \) depressive symptoms \( p = 1.94 \times 10^{-7}; \) irritability \( p = 2.27 \times 10^{-1}; \) Tense/Highly strung \( p = 0.22 \times 10^{-5}; \) Risk Taking \( p = 2.71 \times 10^{-56}; \) PTSD \( r_g = 0.33, p = 3.9 \times 10^{-33}; \) PTSD \( r_g = 0.33, p = 8.09 \times 10^{-5} \)).

**Multivariate Analysis of PTSD and Trauma Exposure**

MTAG analysis that combined PTSD GWAS meta-analysis and UKBB LTE GWAS reported an effective sample size increase of PTSD GWAS from 182,199 to 217,491. There were 8 genome-wide significant loci for the MTAG PTSD analysis, including 4 loci not identified in the PTSD GWAS meta-analysis (Table 1, Figure 1C). Leading variants from additional loci mapped to an intergenic region in chromosome 2, the intron of SGCG on chromosome 5, an intergenic region on chromosome 16 near ZKSCAN2 and AQP8, and the intron on STAU1 on chromosome 20. In gene-based analysis, there were 8 significant genes, including 5 genes not identified from the original GWAS gene-based analysis (CSE1L, DFNA5, FOXP1, SGCG, TRIM26) (Table S3 in Supplement 2).

**Cross-Replication in MVP**

Of the 9 loci identified across the PTSD GWASs (5 from the PGC GWAS and 4 from the MTAG), 4 replicated significantly in MVP (\( p < .006 \)) (Table 1, Figures S2–S10 in Supplement 1). Of the 11 genes identified in gene-based analyses (6 in the GWAS and 5 in the MTAG), 7 replicated at least at a nominally significant level in MVP (\( p < .05 \)) (Table S3 in Supplement 2). Additionally, of 15 loci identified in MVP GWASs, 9 nominally replicated in PGC-PTSD (Table S7 in Supplement 2). Overall, \( r_g \) between PGC-PTSD and MVP was high \( r_g = 0.8359, SE = 0.0376, p = 2.5 \times 10^{-10} \).
associations, 40% psychiatric domain, top psychiatric result: risk taking $p = 1.27 \times 10^{-11}$, rs10821140 (37 significant associations, 38% psychiatric domain, top psychiatric result: loneliness $p = 1.11 \times 10^{-11}$, rs146918648 (44 significant associations, 48% psychiatric domain, top psychiatric result: tensionness/restlessness $p = 2.13 \times 10^{-5}$).

**DISCUSSION**

Our GWASs aimed to advance understanding of PTSD genetics by integrating quantitative PTSD phenotypes and LTE exposure information in 182,199 participants of European ancestry from 51 cohorts. Overall, quantitative PTSD phenotyping captured similar genetic signal to our prior case-control analysis ($r_g = 0.92–1.14$) (8), but with substantially higher power. However, by using LTE as a covariate, which hypothetically accounts for unexpressed genetic vulnerability among unexposed participants (12), we found a significant reduction in heritability and gene discovery. As high $r_g$ between PTSD and LTE would be one hypothetical explanation for this result (i.e., multicollinearity), we performed a GWAS of LTE and contrasted it to GWAS results for PTSD. We found that LTE has $h^2_{\text{SNP}}$ comparable to PTSD and high $r_g$ compared with PTSD. We leveraged the $r_g$ to significantly enhance PTSD discovery power using a multivariate approach (36).

One explanation for $h^2_{\text{SNP}}$ of PTSD adjusted for LTE being lower than the unadjusted estimate is that it may have removed genetic effects on PTSD mediated by trauma exposure (12,13). Given that trauma is a prerequisite for PTSD, genetic effects on trauma exposure can have mediated (i.e., indirect) effects on PTSD. Indeed, this seems plausible, as our LTE GWAS suggested a substantial amount of $h^2_{\text{SNP}}$ related to trauma exposure. Therefore, the estimated $h^2_{\text{SNP}}$ of PTSD conditional on LTE would theoretically reflect only nonmediated (i.e., direct) effects and thus would be smaller.

We used $r_g$ to quantify the genetic overlap between LTE and PTSD, finding similar magnitude to findings from twin studies (5,6). At the same time, incomplete $r_g$ between these two phenotypes also suggested meaningful genetic differences. To investigate this, we contrasted the magnitudes of $r_g$ that PTSD and LTE shared with other traits. For most traits, $r_g$ with PTSD was quite similar in magnitude to $r_g$ with LTE. However, we also found that negative affect traits, such as neuroticism and irritability, were more strongly correlated with PTSD than LTE, whereas risk-taking behavior showed higher correlation with LTE than PTSD. This suggests that some variants influence PTSD and LTE through somewhat distinct psychological and behavioral mechanisms (5).

The high $r_g$ between PTSD and LTE facilitates the application of multivariate approaches to PTSD GWASs. Whereas the $r_g$ between PTSD and LTE induces loss of power in the PTSD analysis when conditioned on LTE, a multivariate approach can benefit from it. Our multivariate (36) analysis resulted in a 19% increase in the effective sample size by adding LTE count data from the UKBB and identified replicable loci and patterns of tissue expression not identified in a standard PTSD GWAS.

The biological mechanisms associated with several of the protein products of identified genes have been linked to PTSD pathophysiology in animal and cell models: amygdala-mediated fear extinction [FAM120A (38)], neuronal transcriptional regulation [FOXP2 (39)], brain excitatory/inhibitory balance [ARHGGEF2, GABBR1, STAU1 (40)], intracellular vesicular trafficking and other synaptic activities [ARHGGEF2 (41), MMP6 (42), SEMA6C (43), SGCD (44)], and inflammation [HIATL1, TRIM26 (45), TRIM27 (46), ZMYM4, ZNF165 (47)]. Blood and brain transcription-wide association and differential gene expression studies of PTSD have also implicated some of these genes, including a blood-based prediction of downregulation of ARHGGEF2 in the dorsolateral prefrontal cortex (48) and a postmortem study of human PTSD cortex indicating downregulation of CTSS expression in the dorsal anterior cingulate cortex and downregulation of OSBPL3 expression in the dorsolateral prefrontal cortex (49).

Interestingly, PTSD loci show widespread pleiotropic associations in PheWAS (50–52). Some loci point to factors associated with existing clinical presentations of PTSD (e.g., sleep), while others point to potential risk/protective factors for PTSD, such as educational attainment and cognitive functioning. Loci may affect PTSD through their direct influence on these risk/protective factors. Alternatively, the high degree of pleiotropy shown by these loci suggests that they could influence PTSD risk through a more general alteration of biological function (37), such as general predisposition to psychiatric illness (53). In particular, metabolic phenotypes such as height and body mass also appeared to be enriched in our PheWAS. This could be the influence of these loci on previously implicated inflammatory mechanisms for PTSD (8) or simply an artifact of their overrepresentation in the GWAS Atlas. Nevertheless, the broad variety of behavioral and clinical domains associated with these loci suggest complex etiologic heterogeneity of PTSD that could relate to subtypes (54).

Further characterization of significant loci via eQTL analyses identified expression across a variety of tissue types. Given the high degree of shared eQTL architecture between tissues, the presence of some of these tissues might not be directly related to PTSD pathogenesis. Indeed, on the genome-wide level, our tissue enrichment analysis suggests that only brain tissues are relevant. The brain regions implicated are consistent with functional magnetic resonance imaging and structural magnetic resonance imaging findings of PTSD. BA 24 (as part of the ventral anterior cingulate cortex) is implicated in PTSD response to trauma-, fear-, and threat-related stimuli (55,56). BA 9 (as part of the dorsomedial prefrontal cortex) reflects response to self-referential thought, theory of mind, empathy, and moral judgments and shows greater engagement in people with PTSD and trauma-exposed individuals (55,57,58). Nucleus accumbens expression is consistent with the neuroimaging evidence of its role in the reward system, which is prominently affected with emotional numbing symptoms of PTSD (59–62).

**Limitations**

Stress-related disorders are phenotypically complex and heterogeneous (63), which limits discovery power and complicates translation to clinical application. The strategies proposed for understanding and addressing heterogeneity in major depressive disorder, such as harmonization of measures, additional phenotypic measures, and investigations of subtypes, could be applied to PTSD as additional avenues to enhance discovery power (64). Sex differences may also contribute a significant
Extended Phenotyping Identifies PTSD Genetic Variants

source of heterogeneity (8,65–68). Our analyses were restricted to participants of European ancestry given power limitations for other ancestry groups. However, urgent scientific and ethical reasons call for extending analyses to individuals of non-European ancestry (69). The PGC-PTSD group has actively been gathering data to increase representation from diverse ancestry and developing methods to optimize analyses in admixed populations (70). As sample sizes increase, future investigations will be powered to investigate ancestry and sex-specific genetic influences on PTSD and trauma exposure. In performing a GWAS of cumulative LTE, we identified several significant loci, including loci previously identified in GWASs of childhood trauma exposure (14). A full investigation of the genetic basis of LTE is clearly warranted. Future work could also examine the relationship between PTSD and specific types or numbers of trauma exposure, as they plausibly have different relationships with PTSD (6) and may therefore be more informative than our cumulative measure for LTE. Finally, trauma was assessed via participant self-report, which may vary with mood and PTSD symptoms at the time of reporting (71) and could inflate genetic associations with PTSD.

Conclusions Novel replicable risk loci for PTSD identified by incorporating quantitative symptom data and trauma exposure information into GWASs offer us new insights into the genetic architecture of PTSD. Beyond the nature of LTE as an environmental exposure, there is a heritable component to LTE that overlaps highly with PTSD to impart an enhanced understanding of PTSD genetics. In future investigations, the genetic architectures of PTSD and LTE could be further delineated using causal mediation analysis (72), which can provide estimates of LTE-related mediation and gene-by-environment interaction. Our results reinforce the notion that in addition to larger samples, more detailed phenotyping and sophisticated modeling are needed to account for the role of environmental exposure in developing PTSD, as these influence GWAS discovery power. Widespread pleiotropy of significant loci suggests that cross-disorder analysis with PTSD (73,74) will enhance our understanding of how these loci modify risk for PTSD and related disorders.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Institute of Mental Health/ U.S. Army Medical Research and Development Command (Grant No. R01MH106595 [to CMN, IL, MBS, KJRe, and KCK]), National Institutes of Health (Grant No. SU01MH109539 to the Psychiatric Genomics Consortium), and Brain & Behavior Research Foundation (Young Investigators Grant [to KWC]). Genotyping of samples was provided in part through the Stanley Center for Psychiatric Genetics at the Broad Institute supported by Cohen Veterans Bioscience. Statistical analyses were carried out on the LISA/Genetic Cluster Computer (https://userinfo.surfsara.nl/systems/lisa) hosted by SURFsara. This research has been conducted using the UK Biobank resource (Application No. 41209). This work would have not been possible without the financial support provided by Cohen Veterans Bioscience, the Stanley Center for Psychiatric Genetics at the Broad Institute, and One Mind.

This material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting true views of the U.S. Department of the Army or the Department of Defense.

We thank the investigators who comprise the PGC-PTSD working group and especially the more than 206,000 research participants worldwide who shared their life experiences and biological samples with PGC-PTSD investigators. We thank Mark Zervas for his critical input. Full acknowledgments are in Supplement 1.

MBs has in the past 3 years received consulting income from Actelion, Acadia Pharmaceuticals, Atrypnix, Bionomics, BioXcel Therapeutics, Clexio, EmpowerPharm, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, and Roche/Genentech and has stock options in Oxeia Biopharmaceuticals and Empirico. In the past 3 years, NPd has held a part-time paid position at Cohen Veterans Bioscience, has been a consultant for Sunovion Pharmaceuticals, and is on the scientific advisory board for Sento Solutions for unrelated work. In the past 3 years, KjRe has been a consultant for Datastat, Inc., RallyPoint Networks, Inc., Sage Pharmaceuticals, and Takeda. JLM-K has received funding and a speaking fee from COMPASS Pathways. Mu has been a consultant for System Analytic. HRK is a member of the Dicema scientific advisory board and a member of the American Society of Clinical Psychopharmacology Alcohol Clinical Trials Initiative, which during the past 3 years was supported by Alkermes, Amyrgydata Neurosciences, Arbor Pharmaceuticals, Dicema, Elymphm, Indivior, Lundbeck, Mitsubishi, and Otsuka. HRK and JG are named as inventors on Patent Cooperative Treaty patent application number 15/878,640, entitled “Genotype-guided dosing of opioid agonists,” filed January 24, 2018. RP and JG are paid for their editorial work on the journal Complex Psychiatry. OAA is a consultant to HealthLytx. All other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Departments of Psychiatry (AXM, EK, KT, DGB, CEF, WSK, SBN, MSP, VBR, JS, MBS, CMN), Family Medicine and Public Health (AXM, RMS), Radiology (AMD), Neurosciences (AMD), and Cellular and Molecular Medicine (JS), University of California San Diego; Herbert Wertheim School of Public Health and Human Longevity Science (WKT, MBS), University of California San Diego; Moores Cancer Center (TWu), University of California San Diego, La Jolla; Center of Excellence for Stress and Mental Health (AXM, EK, KT, DGB, WSK, VBR, MBS, CMN, SBN), Research Service (EK, KT, VBR, CMN), and Psychiatry Service (DGB), Veterans Affairs Healthcare System; Department of Epidemiology and Health Sciences (RAS), Naval Health Research Center and Division of Epidemiology and Biostatistics (TWu), San Diego State University School of Public Health, San Diego; Department of Psychiatry and Behavioral Sciences (LED), Stanford University, Stanford, California; Departments of Epidemiology (KWC, CAD, ARA, CCZ, BG, MAW, KCK), Social and Behavioral Sciences (SBAu), and Environmental Health (ALR), Harvard T.H. Chan School of Public Health; Psychiatric and Neurodevelopmental Genetics Unit (SR, MJD, KCK), Department of Psychiatry (KWC, JWS), and Analytic and Translational Genetics Unit (JWS), Massachusetts General Hospital; Division of Adolescent and Young Adult Medicine (SBAu), Boston Children’s Hospital; Channing Division of Network Medicine (SBAu), Brigham and Women’s Hospital; Departments of Pediatrics (SBAu), Psychiatry (NPd, EAB, GG, MLK, AK, LAML, AVS, MHT, SW, KJRe), and Health Care Policy (RKc), Harvard Medical School; Biomedical Genetics Section (LAF, MWL, MMW, EJW) and Departments of Neurology (LAF), Ophthalmoology (LAF), and Epidemiology (LAF), Boston University School of Medicine; Division of Biostatistics (LAF, MWL), Boston University School of Public Health; Department of Psychological and Brain Sciences (SG) and Dean of Students’ Office (MJL), Boston University; National Center for PTSD (MWL, MMW, EJW), Translational Research Center for TBI and Stress Disorders (REM, WM), and Geriatric Research, Education, and Clinical Center (REM, WM), Veterans Affairs Boston Healthcare System, Boston; Department of Psychology (KAM), Harvard University; Translational Biology (C-YC), Biogen; Stanley Center for Psychiatric Research (NPd, CAD, CCZ, JWS, SR, KCK), Broad Institute of MIT and Harvard; Cohen Veterans Bioscience (MH, HL, ACP), Cambridge; Center of Excellence in Depression and Anxiety Disorders (NPd) and Developmental Biopsychiatry Research Program (MHT), McLean Hospital (EAB, GG, MLK, AK, LAML, AVS, SW, KJRe), Belmont, Massachusetts; Duke Molecular Physiology Institute (RAM, AEA-K, MEG, X-JQ) and Department of Psychiatry and Behavioral Sciences (JCB, MFD, MAH, DEW,
Extended Phenotyping Identifies PTSD Genetic Variants
Extended Phenotyping Identifies PTSD Genetic Variants

University of Freiburg, Freiburg: Department of Psychiatry and Psychotherapy (SR), Charité – Universitätsmedizin, Berlin, Germany; Department of Psychiatry (AGo), University Clinical Center of Kosovo, Pristina, Kosovo; Department of Psychiatry (MJa), University Hospital Centre Zagreb, Zagreb, Croatia; and Department of Medicine (SES), Universidad Peruana de Ciencias Aplicadas Facultad de Ciencias de la Salud, Lima, Peru.

Received Apr 28, 2021; revised Aug 25, 2021; accepted Sep 21, 2021. Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsycho.2021.09.020.

REFERENCES