This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: https://orca.cardiff.ac.uk/id/eprint/151967/

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

Citation for final published version:

Maihofer, Adam X., Engchuan, Worrawat, Huguet, Guillaume, Klein, Marieke, MacDonald, Jeffrey R., Shanta, Omar, Thiruvahindrapuram, Bhooma, Martineau, Jean-louis, Saci, Zohra, Jacquemont, Sebastien, Scherer, Stephen W., Ketema, Elizabeth, Aiello, Allison E., Amstadter, Ananda B., Avdibegovic, Esmina, Babic, Dragan, Baker, Dewleen G., Bisson, Jonathan I., Boks, Marco P., Bolger, Elizabeth A., Bryant, Richard A., Bustamante, Angela C., Caldas-de-Almeida, Jose Miguel, Cardoso, Graca, Deckert, Jurgen, Delahanty, Douglas L., Domschke, Katharina, Dunlop, Boadie W., Dzubur-Kulenovic, Alma, Evans, Alexandra, Feeny, Norah C., Frantz, Carol E., Gautam, Aarti, Geuze, Elbert, Goci Uka, Aferdita, Hammamieh, Rasha, Jakovljevic, Miro, Jett, Marti, Jones, Ian, Kaufman, Milissa L., Kessler, Ronald C., King, Anthony P., Kremen, William S., Lawford, Bruce R., Lebois, Lauren A. M., Lewis, Catrin, Liberton, Isreal, Linnstaed, Sarah D., Lugonja, Bozo, Luykx, Jurjen J., Lyons, Michael J., Mavissakalian, Matig R., McLauiglin, Katie A., McLean, Samuel A., Mehta, Divya, Mellor, Rebecca, Morris, Charles Philip, Muhie, Seid, Orcutt, Holly K., Peverill, Matthew, Ratnatharathorn, Andrew, Risbrough, Victoria, Rizzo, Albert, Roberts, Andrea L., Rothbaum, Alex O., Rothbaum, Barbara O., Roy-Byrne, Peter, Ruggiero, Kenneth J., Rutten, Bart P. F., Schijven, Dick, Seng, Julia S., Sheerin, Christina M., Sorenson, Michael A., Teicher, Martin H., Uddin, Monica, Ursano, Robert J., Vinkers, Christiaan H., Voiye, Joanne, Weber, Heike, Winternitz, Sherry, Xavier, Miguel, Yang, Ruoting, Young, Ros McD., Zoellner, Lori A., Salem, Rany M., Shaffer, Richard, Wu, Tianying, Ressler, Kerry J., Stein, Murray B., Koenen, Karestan C., Sebat, Jonathan and Nievergelt, Caroline M. 2022. Rare copy number variation in posttraumatic stress disorder. Molecular Psychiatry 27, pp. 5062-5069. 10.1038/s41380-022-01776-4 file

Publishers page: https://doi.org/10.1038/s41380-022-01776-4

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.
Posttraumatic stress disorder (PTSD) is a heritable (h² = 24–71%) psychiatric illness. Copy number variation (CNV) is a form of rare genetic variation that has been implicated in the etiology of psychiatric disorders, but no large-scale investigation of CNVs in PTSD has been performed. We present an association study of CNV burden and PTSD symptoms in a sample of 114,383 participants (13,036 cases and 101,347 controls) of European ancestry. CNVs were called using two calling algorithms and intersected to a structural forms of genetic variation are hypothesized to be these forms of variation have been examined more thoroughly in disorders [8]. Recent large investigations of PTSD genetics have uncovered gene-sets, and CNVs overlapping individual genes and implicated neurodevelopmental regions. The total distance covered by deletions crossing over known neurodevelopmental CNV regions was significant (beta = 0.0056, P = 6.3 × 10⁻⁹). The genome-wide neurodevelopmental CNV burden identified explains 0.034% of the variation in PTSD symptoms. The 15q11.2 BP1-BP2 microdeletion region was significantly associated with PTSD (beta = 0.0206, SE = 0.0056, P = 0.0002). No individual significant genes identified were interrupted. 22 gene pathways related to the function of the nervous system and brain were significant (FDR q < 0.05), but these associations were not significant once NDD regions were removed. A larger sample size, better detection methods, and annotated resources of CNV are needed to explore this relationship further.

Molecular Psychiatry; https://doi.org/10.1038/s41380-022-01776-4

INTRODUCTION

Posttraumatic stress disorder (PTSD) has a substantial genetic component [1]. Recent large investigations of PTSD genetics have focused on common genetic variation [2, 3], but rare and structural forms of genetic variation are hypothesized to be important contributors to the development of psychiatric disorders [4]. Rare and structural variation have not received substantial empirical study in the context of PTSD [5]. However, these forms of variation have been examined more thoroughly in association with other psychiatric disorders, where many investigations have specifically focused on copy number variants (CNVs) [6]. CNV associations have been identified for attention-deficit/ hyperactivity disorder (ADHD) [7], autism spectrum disorder (ASD) [8], depression [9, 10], obsessive-compulsive disorder [11], and schizophrenia [12]. Many of the identified psychiatric associations involved specific CNVs that have been implicated in neurodevelopmental disorders (NDD) [9, 10, 13], but also the cumulative burden of CNVs across the genome and enrichment over specific

© The Author(s) 2022
pathways related to the brain and development of the nervous system [12]. Largely owing to lack of available data, there has been no major reported investigation of CNVs and PTSD. However, the recent availability of large sample size PTSD genetic data [2] and available techniques to leverage this data to identify CNVs [14], means that it is now possible to investigate the association between PTSD and CNV burden with an unprecedented level of discovery power.

We present an association study between CNVs and PTSD, conducted in a sample of 114,383 (13,036 cases and 101,347 controls) European ancestry participants from the Psychiatric Genomics Consortium—PTSD [2, 15]. We detected rare (≤1% population frequency) CNVs using algorithms [16–18] applied to the SNP genotyping array intensity data. Following this, we examined the impact of CNV on PTSD on multiple scales: genome-wide CNV burden, enrichment over 46 neuropsychiatric gene-sets [15], CNV burden on individual genes, and CNV carrier status over 53 previously implicated NDD CNV regions [9]. We conclude by comparing the risk contribution from CNVs to the contribution of common variant polygenic risk scores (PRFs).

**METHODS**

**Participants and phenotyping**

The study sample consisted of 114,383 (13,036 cases and 101,347 controls) participants across 20 cohorts from the Psychiatric Genomics Consortium—PTSD freeze 2 data collection. The Psychiatric Genomics Consortium for PTSD is a large scale international effort to investigate genomic contributions to PTSD via meta-analysis of diverse cohorts [2]. For a given PGC-PTSD freeze 2 cohort to be included in this investigation, genotype intensity data had to be available, so that CNV calling could be performed. To reduce the potential for population stratification, we only included subjects of genetically determined [2] European ancestry, the largest homogeneous subset of the data. Within each cohort, participants were assessed for PTSD using either clinical assessment, clinician administered inventory, or self-reported inventory (Supplementary Table 1). Methods of PTSD assessment varied across cohort, and included the BSSS, CAPS, DEQ, IES, NSA, NWS, PCL, PSS, SCID, TSC, and WMH-CIDI. All cohorts provided a PTSD case/control status variable as determined from standard criteria. PTSD assessment varied across cohort, and included the BSSS, CAPS, DEQ, IES, NSA, NWS, PCL, PSS, SCID, TSC, and WMH-CIDI. All cohorts provided a PTSD case/control status variable as determined from standard criteria.

**CNV detection**

DNA was extracted from blood samples. All details regarding DNA extraction and genotyping processes have been published [2]. Participants were genotyped using Illumina arrays (Supplementary Table 1), with the exception that the UK Biobank (UKBB) cohort, which used AxioN genotyping arrays (ThermoFisher). Illumina genotype platform data was self-clustered in GenomeStudio 2.0 and exported as intensity data inputs for CNV callers (SNP name, chromosome, position, allele 1, allele 2, allele frequency, log R ratio, X, and Y). Affymetrix platform genotype data clustering methods have been described previously [9], and log R ratio and allele frequency data were downloaded directly from the UKBB database. For Illumina data, CNVs were called using PennCNV [17] and iPAT [16]. For Affymetrix data, CNVs were called using PennCNV and QuantiSNP [18]. For PennCNV calling, the population frequency of B allele files were generated using the data itself. Waviness correction was applied using a GC content model file generated from UCSC gc_model data ([https://genome.ucsc.edu/cgi-bin/hgTables](https://genome.ucsc.edu/cgi-bin/hgTables)). For the Hidden Markov Model input of PennCNV, the pre-supplied files were used: hhall.hm for Illumina data and afgyw6.6mm for the UKBB data ([https://pennsvv.openbioinformatics.org/en/latest/user-guide/input/hhm- file](https://pennsvv.openbioinformatics.org/en/latest/user-guide/input/hhm-file)). iPAT calls were made using the default program settings, in batches of up to 196 samples. Batches were selected such that samples within a batch were genotyped on the same plate or genotyped at approximately the same time. QuantiSNP calls were made with 10 iterations of the EM algorithm, where the characteristic length used to calculate transition probabilities was set to 2,000,000. GC based correction was performed using UCSC gc_model files.

**CNV quality control**

CNV quality controlled according to the PGC CNV calling pipeline [12]. To ensure that the analysis included a reliable set of calls, CNV calls produced by the different calling algorithms were intersected to produce a consensus set. CNVs called as gain by one method and loss by the other were also excluded from further analyses. Fragmented large CNVs in a log R ratio profile were removed if the gap length between them was less than 30% of the overall length of the annotated CNV. CNV quality metrics calculated by PennCNV were used to perform sample QC. Subjects were removed if their values for SD of log R ratio, B allele frequency, or waviness were > = Q3 + 3IQR, if >20% of any chromosome was copy number variant (aneuploidy), or if they had excessive CNV count (>Q3 + 3IQR CNVs) or KB burden (>≥3 kb). Participants who failed standard genotype QC were removed: sample missingness rates > 2%, excess heterozygosity, mismatch between self-reported sex and genetically determined sex, π relatedness coefficient > 0.2. We removed CNVs for any of the following reasons: 50% overlap with centromeres, telomere, immunoglobulin or T-cell receptor loci, >50% overlap with known segmental duplications, CNV frequency >1% (measured within the data) in cases and controls and >10 kb in CNV length or intersecting <10 probes.

**CNV burden calculation**

CNV burden was measured and evaluated for association with PTSD in multiple ways: The cumulative burden of CNVs was calculated as the genome-wide total distance (in megabases) by spans by CNVs. For each of the 53 NDD CNV regions, NDD CNV carrier status was determined as having at least 50% of the NDD CNV region overlapped by CNV. As a sensitivity analysis, two different overlap criteria (>0% or 100% overlap) were also evaluated. For gene-level CNV burden, first gene positions (GRCh37 human genome build) were downloaded from the UCSC table browser ([https://genome.ucsc.edu/cgi-bin/hgTables](https://genome.ucsc.edu/cgi-bin/hgTables)). Genes were filtered to protein coding genes, based on having an “NM_” accession prefix in the National Center for Biotechnology Information reference sequence database [19]. For genes with multiple isoforms, the minimum start and the maximum end positions were used. For each CNV, the CNV was mapped to all genes it overlapped by at least one base pair. The CNV burden variable was then calculated for each gene, coded 1 if the subject carried a CNV that mapped onto the gene, and 0 otherwise. For gene-set analysis, a gene-set burden variable was calculated for each set tested, coded as the number of genes within the set overlapped by the CNVs. The gene-set analysis included 53 gene-sets, consisting of 23 gene-sets related to neurofunction or nervous system, 6 brain expression from BrainSpan consortium and 7 mouse phenotype negative control gene-sets from previous neurological disorders studies [12, 20], a set of loss-of-function intolerant genes as defined by gnomAD v2.0 [21], and 16 brain-expressed gene-sets from human neocortex scRNA data [22]. A list of genes belonging to each set is included in Supplementary Table 2.

**Statistical analyses**

A two stage meta-analytic approach was conducted, where analyses were performed within each cohort separately then results were combined via meta-analysis. As all subjects belonging to a given cohort were genotyped using the same platform, this analysis was effectively performed stratified by platform, thus accounting for potential confounding due to CNV calling across platforms. Within each cohort, the association between PTSD and CNVs was tested using a regression model of PTSD as predicted by the CNV variable, five principal components calculated from genotype call data using Eigenstrat 6.0.2 [23], and the log R ratio standard deviation sample quality metric from PennCNV. For the gene-set analyses, in order to follow the enrichment test model outlined by Raychaudhuri et al. [24] analyses also contained predictors for genome-wide total CNV count and genome-wide average length of CNVs. Cohorts with continuous PTSD symptom measures were analyzed using linear regression and cohorts with only case/control phenotypes were analyzed using logistic regression. Results across cohorts were meta-analyzed using fixed effects inverse variance weighted meta-analysis in the metafor [25] R package. For the meta-analysis, to account for the different PTSD measure scales used across cohorts, PTSD measures were scaled from 0 to 1 according to the theoretical range of scores of the assessment method (i.e. = no PTSD symptoms, = theoretical maximum possible PTSD symptoms), and case/ control estimates were interpreted as being the observed, censored variable for a latent symptom measure variable. Statistical significance was declared based on Benjamini-Hochberg false discovery rate (FDR).
q value < 0.05 calculated within a family of tests. To enhance interpretability of results, we also provide odds ratio effect estimates, via analyzing cohorts with continuous data using an ordinal logistic regression. For this analysis, odds ratio estimates were directly meta-analyzed across studies (i.e., not rescaled) using inverse-variance weighted meta-analysis. The statistical inferences made in this manuscript are however based only on the linear regression based results. To examine if outliers strongly contributed to the results of analyses of the 16p11.2 deletion and 2q13 deletion CNVs, linear regression was also performed using heteroskedasticity consistent (HC3) standard errors [26].

We estimated PRS for PTSD in all participants. SNP weights were obtained from the Million Veteran Program PTSD GWAS [3] of European ancestry participants, with weights adjusted using PRS-CS [27] under default parameters, with 1000 Genomes Phase 3 European data [28] used to model linkage disequilibrium. SNPs were filtered to common (minor allele frequency > 1%), strand unambiguous variants. PRS were computed as the weighted sum of risk alleles at each markers using the -score option in PLINK [29]. PRS were standardized to mean zero and unit standard deviation, such that the effects reported refer to PTSD risk relative to every unit standard deviation increase in PRS. The proportion of variance in PTSD explained by PRS and CNV was estimated as the difference in model r-squared values between a baseline model that included all relevant covariates and the model with additional terms for PRS and CNV. Standard errors for the proportion of variance explained were calculated using the formulae from Cohen et al. [30].

**RESULTS**

The PTSD CNV meta-analysis included 114,383 participants (13,036 cases and 101,347 controls) of European ancestry across 20 cohorts (Supplementary Table 1, Table 1). The method of PTSD assessment varied across cohorts (11 distinct methods), with most participants being assessed via a version of the PCL (N = 106,353). The majority of subjects (N = 113,320, 99%) were analyzed using PTSD symptom scores, the remaining subjects were analyzed using case/control status. 15 cohorts were genotyped using the Psych array (N = 6,813 samples), 1 with the Psych Chip (N = 756 samples), 3 with the OmniExpressExome8+Custom content (N = 9,432 samples), and 1 with the Affymetrix UK Biobank Axiom array (N = 97,382). CNVs were produced as the consensus call of iPattern and PennCNV (Illumina arrays, N = 19 studies) or PennCNV and QuantiSNP (Affymetrix array, N = 1). The final dataset included 103,036 CNVs (41,473 gains and 61,563 losses). The median length of CNVs was 122,756 BP (range = 10,000 to 9,911,819 BP) (Supplementary Fig. 1). 60.1% of participants were carriers of at least one CNV (Table 1). Among CNV carriers, the average total span of CNV carried was 0.32 megabases (SD = 0.35), and the average of within subject average CNV lengths was 0.23 megabases (SD = 0.26).

**Genome-wide CNV burden analysis**

Genome-wide cumulative CNV burden was significantly associated with PTSD (beta = 0.0028, SE = 0.0008, p = 0.0003, q = 0.001; OR = 1.025, 95%CI[1.002,1.049]) (Fig. 1). We examined CNV burden stratified by type (duplication or deletion), finding that the total distance covered by deletions was significant (beta = 0.0046, SE = 0.0013, p = 0.0004, q = 0.001; OR = 1.042, 95%CI = [1.007,1.080]) but the total distance covered by duplications was not (beta = 0.0018, SE = 0.0010, p = 0.065, q = 0.11; OR = 1.054, 95%CI = [0.985–1.043]). Next, we examined CNV burden stratified by overlap with any of 53 previously implicated NDD CNV regions. The cumulative burden of CNV deletions that overlapped NDD regions was significantly associated with PTSD (beta = 0.0290, SE = 0.0054, p = 6.3 x 10^{-6}, q = 1 x 10^{-6}; OR = 1.576, 95% CI = [1.314,1.889]), while the duplication burden was not (beta = 0.0053, SE = 0.0023, p = 0.024, q = 0.06; OR = 1.055, 95%CI = [0.972,1.146]). The genome-wide burden of non-NDD CNV deletions was not significant (beta = 0.0031, SE = 0.0013, p = 0.023, q = 0.054; OR = 1.008,95%CI = [0.978–1.040]) (Supplementary Table 3).

**Table 1. Cohorts analyzed.**
Three NDD CNV were significantly associated with increased PTSD symptoms, the 2q13 deletion (chr2:111,394,040–112,012,649, N = 529 carriers, beta = 0.1455, SE = 0.0367, p = 0.0001, q = 0.0027; OR = 2.508, 95% CI = [0.956, 6.583]), the 15q11.2 BP1-BP2 microdeletion (chr15:22,805,313–23,094,530, N = 529 carriers, beta = 0.0206, SE = 0.0056, p = 0.0002, q = 0.0027; OR = 1.275, 95% CI = [1.093, 1.488]), and the 16p11.2 deletion (N = 16 carriers, beta = 0.0702, SE = 0.025, p = 0.0041, q = 0.0369; OR = 2.619, 95% CI = [1.019, 6.728]) (Fig. 1). Given the limited number of carriers for 2q13 deletion and 16p11.2 deletion, we tested their association again under models with robust standard errors, finding that the neither the 2q13 deletion nor the 16p11.2 deletion were significantly associated (p = 0.11 and p = 0.25, respectively). The overall results were similar under a stricter definition of carrier status (100% overlap of NDD CNV region) (Supplementary Table 4), whereas under a loose definition of carrier status (>0% overlap of NDD CNV region), four regions were FDR significant: the 8p23.1 del (beta = 0.0233, SE = 0.0078, p = 0.003, q = 0.04; OR = 1.271, 95% CI = [1.021, 1.582]), 15q11.2 BP1-BP2 del (beta = 0.0201, SE = 0.0056, p = 0.0003, q = 0.007; OR = 1.27, 95% CI = [1.090, 1.480]), 15q11.2-q12 Prader-Willi/Angelman syndrome del (beta = 0.0186, SE = 0.0053, p = 0.0004, q = 0.007; OR = 1.25, 95% CI = [1.080, 1.447]), and 22q11.2 dup (beta = 0.0216, SE = 0.0055, p = 8.3 x 10^{-3}, q = 0.0041; OR = 1.277, 95% CI = [1.128, 1.444]). We note that in this less restrictive analysis, the association of the 15q11.2-q12 Prader-Willi/Angelman syndrome was driven by the smaller 15q11.2 BP1-BP2 deletion and that no subjects in this study carried a deletion with a >50% overlap with the Prader-Willi/Angelman syndrome critical region.

Specific NDD CNV regions confer risk for PTSD

We investigated the association between PTSD and NDD CNV carrier status. 33 out of 53 NDD CNVs had at least 1 carrier (Supplementary Table 4). The most common NDD CNV was the 15q11.2 BP1-BP2 deletion (N = 529 carriers, frequency = 0.046%). Three NDD CNV were significantly associated with increased PTSD symptoms, the 2q13 deletion (chr2:111,394,040–112,012,649, N = 15 carriers, beta = 0.1455, SE = 0.0367, p = 0.0001, q = 0.0027; OR = 2.508, 95% CI = [0.956, 6.583]), the 15q11.2 BP1-BP2 microdeletion (chr15:22,805,313–23,094,530, N = 529 carriers, beta = 0.0206, SE = 0.0056, p = 0.0002, q = 0.0027; OR = 1.275, 95% CI = [1.093, 1.488]), and the 16p11.2 deletion (N = 16 carriers, beta = 0.0702, SE = 0.025, p = 0.0041, q = 0.0369; OR = 2.619, 95% CI = [1.019, 6.728]) (Fig. 1). Given the limited number of carriers for 2q13 deletion and 16p11.2 deletion, we tested their association again under models with robust standard errors, finding that the neither the 2q13 deletion nor the 16p11.2 deletion were significantly associated (p = 0.11 and p = 0.25, respectively). The overall results were similar under a stricter definition of carrier status (100% overlap of NDD CNV region) (Supplementary Table 4), whereas under a loose definition of carrier status (>0% overlap of NDD CNV region), four regions were FDR significant: the 8p23.1 del (beta = 0.0233, SE = 0.0078, p = 0.003, q = 0.04; OR = 1.271, 95% CI = [1.021, 1.582]), 15q11.2 BP1-BP2 del (beta = 0.0201, SE = 0.0056, p = 0.0003, q = 0.007; OR = 1.27, 95% CI = [1.090, 1.480]), 15q11.2-q12 Prader-Willi/Angelman syndrome del (beta = 0.0186, SE = 0.0053, p = 0.0004, q = 0.007; OR = 1.25, 95% CI = [1.080, 1.447]), and 22q11.2 dup (beta = 0.0216, SE = 0.0055, p = 8.3 x 10^{-3}, q = 0.0041; OR = 1.277, 95% CI = [1.128, 1.444]). We note that in this less restrictive analysis, the association of the 15q11.2-q12 Prader-Willi/Angelman syndrome was driven by the smaller 15q11.2 BP1-BP2 deletion and that no subjects in this study carried a deletion with a >50% overlap with the Prader-Willi/Angelman syndrome critical region.

Gene and gene-set analyses

We examined CNV association on the level of protein coding genes. 2880 genes harbored CNV with at least 0.01% frequency. We found that no gene was significant after multiple comparisons correction for the number of genes, in any strata (overall CNV, duplications, or deletions) (Supplementary Table 5). Following this we examined if CNV burden association with PTSD was enriched in any of 46 different gene-sets related to the brain and nervous system and 7 control gene-sets of non-brain tissue types. No control gene-set was significant. In contrast, 22 out of 46 sets related to the brain and nervous system were enriched in CNV deletions (FDR q < 0.05) (Supplementary Table 6). Many of the top ranked genes in the significant sets overlapped with NDD CNV regions (Supplementary Table 7). As a sensitivity analysis, we removed CNVs overlapping or nearby (within 1 million base pairs) the NDD CNV regions (Supplementary Table 8), finding that no gene-set remained significant after this adjustment (all FDR q > 0.05).

Comparisons with common variant genetics

We generated PTSD polygenic risk scores for our data based on the recent independent MVP PTSD GWAS. We included PRS and cumulative NDD CNV carrier burden in a regression model of PTSD symptoms. PTSD PRS was significantly associated with increased PTSD symptoms (beta = 0.011, SE = 0.0004, p = 9.8 x 10^{-15}; OR = 1.16, 95% CI = [1.151, 1.17]) and explained 0.5% of the total variation in PTSD symptoms (SE = 0.04%, p = 2.6 x 10^{-11}). NDD CNV burden was also significantly associated with PTSD symptoms (beta = 0.0027, SE = 0.0003, p = 7.7 x 10^{-8}; OR = 1.57, 95% CI = 1.31, 1.89), and explained an additional 0.034% (SE = 0.0001, p = 0.0017) of the variation in PTSD symptoms.

DISCUSSION

We identified an association between the cumulative burden of CNVs and PTSD, which was largely driven by CNVs overlapping previously implicated NDD CNV regions. Two recent studies of CNVs in major depression [9, 10] also reported associations with NDD CNV burden, with effect sizes comparable to ours. The modest to moderate effect sizes observed are consistent with PTSD and MDD being disorders with less severity of cognitive impairment, comparatively moderate heritability and a larger
environmental component. In terms of how CNV burden modifies depression risk, Kendell et al. [9] suggested that the majority of the total effect came from the direct effects of CNVs, with some evidence of additional mediated effects stemming from socio-demographic risk factors including physical health, smoking, alcohol consumption, educational attainment, and social deprivation. As PTSD has similar risk factors [31], NDD CNVs may influence PTSD risk via the same mediated mechanisms. We propose that some of the psychiatric and neurodevelopmental consequences of CNVs may also increase PTSD risk, as they represent PTSD risk factors [32] [33].

In examining the individual NDD CNVs, we observed a significant association of PTSD with the 15q11.2 BP1-BP2 microdeletion, one of the most frequently occurring pathogenic CNVs identified in humans [34]. This CNV is associated with alterations in brain morphology and cognition [35]. There is a wide variety of possible clinical manifestations, including developmental delays, intellectual disability, as well as behavioral and psychiatric problems, including ADHD, ASD and schizophrenia [36]. Under a less strict definition of NDD carrier status (>0% overlap with NDD CNV region), the 22q11.2 duplication region and 8p23.1 deletion regions were significant. The 22q11.2 duplication has a variety of deleterious impacts [37], but generally they are less severe than those observed in the 22q11.2 deletion [38]. The 8p23.1 deletion is associated with developmental delays, hyperactivity, and impulsivity [39]. Rather than any specific functional aspects of these CNVs having led to the significant associations that we observed, we suspect that their relatively high frequencies in the data made them among the most statistically powered to identify.

Pathway specific enrichment of brain regions and neurodevelopmental gene-sets has consistently been observed in genetic studies of psychiatric disorders [3, 40–42]. We have identified significant associations with several biological pathways related to the development of the brain and nervous system. Our pathway analysis was not significant once we removed the CNV overlapping NDD regions, possibly suggesting an outsized or central role of genes in NDD regions relative to other genes within the pathways. Genes in NDD regions are known affect the development of the brain and nervous system, likely through the disruption of core molecular pathways [43] [44].

The regions we have identified as significant in CNV analyses have not been implicated in GWAS of PTSD. These regions may represent a distinct element of the genetic contribution to PTSD risk that is not readily identified by common variant analyses, suggesting that rare variation analysis complements common variant analysis, as has been hypothesized for psychiatric phenotypes [4]. The effects of implicated CNVs were modest in magnitude, albeit higher than reported common variant effects, consistent with the hypothesis that rare variants have stronger effects than common ones [45].

In terms of population risk prediction, due to the limited number of CNV carriers, CNV burden predicted substantially less total variation in PTSD than PRS. The utility of determining carrier status, rather than population level prediction, is that CNV carriers may be a subset of individuals for whom a tailored health management strategy [46] applies. Indeed, CNV carrier status has been proposed as a tool in clinical decision making for psychiatric disorders, albeit one that will first require expansion of the clinical knowledge base of CNVs [47]. But it is unclear how much this will apply directly to PTSD, as we did not identify any highly penetrant CNVs.

CONCLUSIONS
We have performed, to our knowledge, the largest (N=114,383 participants) investigation of the influence of CNV burden on PTSD risk, and furthermore, are the first to identify significant associations. Risk was enriched in regions that crossed over known NDD regions. In particular, we have implicated the 15q11.2 BP1-BP2 microdeletion. Larger sample size data, better detection methods, and annotated resources of CNV are necessary to explore these relationships further.

CODE AVAILABILITY
Code is available from https://github.com/nievergeltlab/cnv_freeze1 or by request to the corresponding author.

REFERENCES
1. Duncan LE, Cooper BN, Shen H. Robust findings from 25 years of PTSD genetics research. Curr Psychiatry Rep. 2018;20:115–115.
2. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen C-Y, Choi KW, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat Commun. 2019;10:4558.
3. Stein MB, Levey DF, Cheng Z, Wendt FR, Harrington K, Pathak GA, et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. Nat Genet. 2021;53:174–84.
4. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Berglum AD, Breen G, et al. Psychiatric genomics: an update and an agenda. Am J Psychiatry. 2018;175:15–27.
5. Smoller JW. The genetics of stress-related disorders: PTSD, depression, and anxiety disorders. Neuropsychopharmacology. 2016;41:297–319.
Ask H, Cheesman R, Jamis ES, Levey DF, Purves KL, Weber H. Genetic contributions to anxiety disorders: where are we and where are we heading. Psychol Med. 2022;51:2231–46.

Gudmundsson OO, Walters GB, Ingason A, Johannsson S, Zayats T, Athanasiou L, et al. Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. Transl Psychiatry. 2019;9:258.

Reichardt L, Gernandt DA. Genetic causes and modifiers of autism spectrum disorder. Front. Cell Neurosci. 2019;13:385.

Kendall KM, Rees E, Bracher-Smith M, Legge S, Rignol L, Zammitt S, et al. Association of rare copy number variants with risk of depression. JAMA Psychiatry. 2019;76:188–25.

Bimbbaum R, Mahjani B, Loos RJF, Sharp AJ. Clinical characterization of copy number variants associated with neurodevelopmental disorders in a large-scale multicenter biobank. JAMA Psychiatry. 2022;79:250–9.

Saraiya LC, Cappi C, Simpson HB, Stein DJ, Viswanath B, van den Heuvel OA, et al. Cutting-edge genetics in obsessive-compulsive disorder. Fam Pract. 2020:39:20–30.

Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. 2017;49:27–35.

Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. npj Genom Med. 2019;4:26.

Balagüé-Dobón L, Cáceres A, González JR. Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure. Bioinform. 2022;23:bbac043.

Polgoudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. A hidden Markov model designed for high-resolution copy number variation detection using next-generation sequencing data: features and functional annotation. Nucleic Acids Res. 2016;44:D733–D745.

Pinto D, Pagamnetta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional and genetic causes and modifiers of autism spectrum disorder. Front. Cell Neurosci. 2019:13:385.

Breslau N. The epidemiology of trauma, PTSD, and other posttrauma disorders. Trauma, Violence, Abus. 2009;10:198–210.

O'Hare T, Shen C, Sherrer M. High-risk behaviors and drinking-to-cope as mediators of lifetime depression and PTSD symptoms in individuals with severe clinical mental illness. J Trauma Stress. 2010;23:255–63.

Rafi SK, Butler MG. The 15q11.2 BP1–BP2 microdeletion (Bursmide–Butler) syndrome: in silico analyses of the four coding genes reveal functional associations with neurodevelopmental disorders. Int J Mol Sci. 2020;21:2396.

Writing Committee for the E-CNVWG. Association of copy number variation of the 15q11.2 BP1 copy number variant with cortical and subcortical morphology and cognition. JAMA Psychiatry. 2020;77:420–30.

Cox DM, Butler MG. The 15q11.2 BP1–BP2 microdeletion syndrome: a review. Int J Mol Sci. 2015;16:4068–82.

Ou Z, Berg JS, Yonath H, Enciso VB, Miller DT, Picker J, et al. Multiduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. Genet Med. 2008;10:262–77.

Wenger TL, Miller JS, DePolio LM, de Marchena AB, Clements CC, Emanuel BS, et al. 22q11.2 duplication syndrome: elevated rate of autism spectrum disorder and need for medical screening. Mol Autism. 2016;7:27.

Digilio MC, Marino B, G Buccione P, Giannotti A, Mingarelli R, Dallapiccola B. Deletion 8p syndrome. Am J Med Genet. 1998;75:334–6.

Mullins N, Forster AJ, O’Connell KS, Coombes B, Coleman JR, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53:817–29.

Devekots D, Walters RK, Martin J, Matthiesen M, Als TD, Abberger E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

Paridaas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50:381–9.

Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental disorders: from genetics to functional pathways. Trends Neurosci. 2020:43:608–21.

Cardoso AR, Lopes-Marques M, Silva RM, Serrano C, Amorim A, Prata MJ, et al. Essential genetic findings in neurodevelopmental disorders. Hum Genom. 2019;13:31.

Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biol. 2017;18:77.

Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19:581–90.

Sullivan PF, Owen MJ. Increasing the clinical psychiatric knowledge base about polygenic risk. JAMA Psychiatry. 2020;77:204–9.

Coughlin CR, Scharer GH, Shakh TH. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. Genome Med. 2012;4:80.

Nowakowska B. Clinical interpretation of copy number variants in the human genome. J Appl Genet. 2017;58:449–57.

Zhou M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinform. 2013;14:51.

Kaiser J. 200,000 whole genomes made available for biomedical studies. Science. 2021;374:1036.

Rees E, Kirov G. Copy number variation and neuropsychiatric illness. Curr Opin Genet Dev. 2021;61:685–7.

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. Am J Epidemiol. 2017;186:1026–34.

Janch AE, Douard E, Moreau C, Van Dijck A, Passeggeri M, Kooy F, et al. Estimating the effect size of the 15Q11.2 BP1–BP2 deletion and its contribution to neurodevelopmental syndromes: recommendations for practice. J Med Genet. 2019;56:701.

Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95:5–22.

Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95:5–22.
Financial support for the PTSD PGC was provided by the Cohen Veterans Bioscience, Stanley Center for Psychiatric Research at the Broad Institute, and One Mind. 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 USA/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).

Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The views expressed in this article are those of the authors and do not reflect the official policy or position of the DoD, or the U.S. Government. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70-25.

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 Henry R. Kranzler for his critical input.

AUTHOR CONTRIBUTIONS

Conception and design of the work: A.X.M., C.M.N., J.S., K.C.K., K.J.R., M.B.S., R.A.S., R.M.S., T.W. Acquisition, analysis, or interpretation of data: A.B.A., A.C.B., A.D.-K., A.E., A.E.A., A.G., A.G.U., A.L.R., A.O.R., A.P.K., A.R., A.X.M., B.L., B.O.R., B.P.F.R., B.R.L., B.T., B.W.D., C.E.F., C.E.L., C.H.V., C.M.N., C.M.S., C.P.M., D.B., D.G.B., D.L.D., D.M., D.S., E.A., E.A.B., E.G., E.K., G.C., G.H., H.K.O., H.W., I.J., I.L., I.D., J.B., J.J.L., J.J.M., J.D.A., J.R., J.S., J.S.S., J.V., K.A.M., K.C.K., K.D., K.J.R., K.J.R., K.M.L., L.A.M., L.A.Z., M.A.S., M.B.S., M.H.T., M.I.L., M.J.-L., M.Ja., M.Je., M.K., M.L.K., M.P., M.P.B., M.R.M., M.U., M.X., N.C.F., O.S., P.B., PGC-CN, PGC-PTD, R.A.B., R.A.S., R.C.K., R.H., R.J.U., R.M., R.M.S., R.M.Y., R.Y., S.A.M., S.D.L., S.J., S.J.M., T.W., T.D.S., V.B.R., W., W.J.C., W.S., Z.S. Drafted the work or substantively revised it: A.X.M., B.T., E.K., G.H., J.R.M., M.J.-L., M.K., O.S., S.J., S.W.S., W.E., Z.S.

COMPETING INTERESTS

MBS has in the past 3 years received consulting income from Actelion, Acadia Pharmaceuticals, Aptinyx, Biomemics, BioKeli Therapeutics, Cleixo, EmpowerPharm, GW Pharmaceuticals, Janssen, Jazz Pharmaceuticals, and Roche/Genentech and has stock options in Oxeia Biopharmaceuticals and Epivario. In the past 3 years, RCK has been a consultant for Datstat, Inc., RallyPoint Networks, Inc., Sage Pharmaceuticals, and Takeda. MU has been a consultant for System Analytic. All other authors report no biomedical financial interests or potential conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01776-4.

Correspondence and requests for materials should be addressed to Adam X. Maihofer.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
PSYCHIATRIC GENOMICS CONSORTIUM CNV WORKING GROUP

Adam X. Maihofer,1,2,3,5, Elizabeth Ketema,1,3,13, Allison E. Aiello,14, Ananda B. Amstaefer,15, Esmina Avdibeovitc16, Dragan Babic17, Dewleen G. Baker18,19, Jonathan I. Bisson19, Marco R. Boks20,21, Elizabeth A. Bolger21,22, Richard A. Bryant21,23, Angela C. Bustamante24,25, Jose Miguel Caldas-de-Almeida26, Graça Cardoso27,28,29, Jurgen Deckert30,31, Douglas L. Delahanty32,33,34, Katharina Domschke35,36,37, Boadie W. Dunlop38,39, Alma Dzubur-Kulenovic40,41,42, Alexandra Evans43,44, Norah C. Feeny45, Carol E. Franz46, Aarti Gautam47, Elbert Geuze48,49, Rasha Hammamieh50,51,52, Anthony P. King53, William S. Kremen, Jr.54,55, Bruce R. Lawford56,57, Lauren A. M. Lebois58,59, Catrin Lewis60,61,62, Sarah D. Linnstaedt62,63, Bozo Lugonja64,65, Ian Jones66,67,68,69, Michael J. Lyons66,67,68,69, Matig R. Mavissakalian69,70, Katie A. McLaughlin71,72, Samuel A. McLean73,74,75, Divya Mehta76,77,78,79, Rebecca Mellor80,81,82, Charles Phillip Morris83,84, Seid Muhie85,86,87,88, Holly K. Orcutt89,90, Matthew Peverill91,92, Andrew Ratanatharathorn93,94, Victoria B. Risbrough95,96,97,98, Albert Rizzo99,100, Andrea L. Roberts101,102, Alex O. Rothbaum103, Barbara O. Rothbaum104, Peter Roy-Byrne105,106, Kenneth J. Ruggiero107,108, Bart P. F. Rutten109,110, Dick Schijven111,112, Julia S. Seng113,114, Christina M. Sheerin115,116,117, Michael A. Sorenson118,119, Martin H. Teicher120,121, Monica Uddin122,123, Robert J. Ursano124,125, Christiaan H. Vinkers126,127,128, Joanne Voisey129,130,131, Keivan H. Vose132,133,134, Miguel Xavier135,136, Ross McD Young137,138, Loni A. Zoellner139,140, Kerry J. Ressler141,142,143, Murray B. Stein144,145,146, Karestan C. Koenen147,148,149,200,201,202, and Caroline M. Nievergelt1,2,3,150.

PSYCHIATRIC GENOMICS CONSORTIUM PTSD WORKING GROUP

Adam X. Maihofer,1,2,3,5, Worrawat Engchuan4,5, Guillaume Huguet6, Marieke Klein7, Jeffrey R. MacDonald8, Omar Shanta9, Bhooma Thiruvahindrapuram10, Martineau Jean-louis11, Zohra Saci12, Sebastien Jacquemont13, Stephen W. Scherer14,15,16 and Jonathan Sebat17,18,19,20.

Department of Psychiatry, Medical University of South Carolina, Charleston, SC, USA. 1Department of Psychiatry and Neuropsychology, Maastricht Universitair Medisch Centrum, School for Mental Health and Neuroscience, Maastricht, Limburg, the Netherlands. 2University of Michigan, School of Nursing, Ann Arbor, MI, USA. 3Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI, USA. 4Department of Women's and Gender Studies, University of Michigan, Ann Arbor, MI, USA. 5University of Michigan, Institute for Research on Women and Gender, Ann Arbor, MI, USA. 6Developmental Biopsychiatry Research Program, McLean Hospital, Belmont, MA, USA. 7Genomics Program, College of Public Health, University of South Florida, Tampa, FL, USA. 8Department of Psychiatry, Uniformed Services University, Bethesda, MD, USA. 9Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep & Stress Program, Amsterdam, the Netherlands. 10Department of Psychiatry, Amsterdam UMC Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. 11Department of Anatomy and Neurosciences, Amsterdam UMC Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. 12Univesidadede Nova de Lisboa, Nova Medical School, Lisboa, Portugal. 13Queensland University of Technology, School of Clinical Sciences, Kelvin Grove, QLD, Australia. 14University of the Sunshine Coast, The Chancellory, Sippy Downs, QLD, Australia. 15University of California San Diego, School of Public Health, La Jolla, CA, USA. 16Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research, Cambridge, MA, USA. 17Department of Epidemiology, Harvard T. H. School of Public Health, Boston, MA, USA. 18Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. 19Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. 20Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. 21University of California San Diego, Herbert Wertheim School of Public Health and Human Longevity Science, La Jolla, CA, USA. 22Department of Epidemiology and Health Sciences, Naval Health Research Center, San Diego, CA, USA. 23Division of Epidemiology and Biostatistics, San Diego State University, School of Public Health, San Diego, CA, USA. 24University of California, San Diego, Moores Cancer Center, San Diego, CA, USA. 25These authors contributed equally: Jonathan Sebat, Caroline M. Nievergelt. Lists of authors and their affiliations appear at the end of the paper. 26email: amaihofer@health.ucsd.edu