Epigenome-wide association of PTSD from heterogeneous cohorts with a common multi-site analysis pipeline

Andrew Ratanatharathorn[1] | Marco P. Boks[2] | Adam X. Maihofer[3] | Allison E. Aiello[4] | Ananda B. Amstadter[5] | Allison E. Ashley-Koch[6] | Dewleen G. Baker[3,7,8] | Jean C. Beckham[9,10,11] | Evelyn Bromet[12] | Michelle Dennis[9,11] | Melanie E. Garrett[6] | Elbert Geuze[2,13] | Guia Guffanti[14,15] | Michael A. Hauser[6] | Varun Kilaru[16] | Nathan A. Kimbrel[9,10,11] | Karestan C. Koenen[17,18,19] | Pei-Fen Kuan[20] | Mark W. Logue[21,22] | Benjamin J. Luft[23] | Mark W. Miller[22,24] | Colter Mitchell[25] | Nicole R. Nugent[26] | Kerry J. Ressler[14,15,27] | Bart P. F. Rutten[28] | Murray B. Stein[3,7,29] | Eric Vermetten[2,30,31,32] | Christiaan H. Vinkers[2] | Nagy A. Youssef[33] | VA Mid-Atlantic MIRECC Workgroup[9] | PGC PTSD Epigenetics Workgroup | Monica Uddin[34,35] | Caroline M. Nievergelt[3,7,8] | Alicia K. Smith[16,27]

[1] Department of Epidemiology, Columbia University, New York, New York

[2] Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

[3] Department of Psychiatry, University of California San Diego, San Diego, California

[4] Department of Epidemiology, UNC Chapel Hill, Chapel Hill, North Carolina

[5] Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia

[6] Duke Molecular Physiology Institute, Durham, North Carolina

[7] Veterans Affairs San Diego Healthcare System, San Diego, California

[8] Veterans Affairs Center of Excellence for Stress and Mental Health, San Diego, California

[9] Mental Illness Research, Education, and Clinical Center, VA Mid-Atlantic, Durham, North Carolina

[10] Durham VA Medical Center, Durham, North Carolina

[11] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina

[12] Epidemiology Research Group, State University of New York, New York

[13] Military Mental Healthcare-Research Centre, Ministry of Defence, Utrecht, Netherlands

[14] Neurobiology of Fear Laboratory, McLean Hospital, Belmont, Massachusetts

[15] Department of Epidemiology and Massachusetts General Hospital, Department of Psychiatry, Harvard T.H. Chan School of Public Health, Belmont, Massachusetts

[16] Department of Gynecology and Obstetrics, Emory University, Atlanta, Georgia

[17] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

[18] Center for Human Genetic Research, and Department of Psychiatry, Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts

[19] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts

[20] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York

[21] Department of Medicine (Biomedical Genetics), Boston University, Boston, Massachusetts

[22] VA Boston Healthcare System, Boston, Massachusetts

[23] Department of Medicine, State University of New York, New York

[24] Department of Psychiatry, Boston University School of Medicine,, Boston, Massachusetts

[25] Institute for Social Research, University of Michigan, Ann Arbor, Michigan

[26] Department of Pediatric Research, Psychiatry and Human Behavior, Brown University, Providence, Rhode Island

[27] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia

[28] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht,, Limburg, Netherlands

[29] Department of Family Medicine and Public Health, University of California San Diego, San Diego, California

[30] Department of Psychiatry, Leiden University Medical Center, Utrecht, Netherlands

[31] Military Mental Healthcare, Ministry of Defence, Utrecht, Netherlands

[32] Arq Psychotrauma Expert Group, Diemen, Netherlands

[33] Department of Psychiatry and Human Behavior and Office of Academic Affairs, Medical College of Georgia at Augusta University, Augusta, Georgia

[34] Department of Psychology, University of Illinois Urbana-Champaign, Urbana, Illinois

[35] Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, Illinois

Correspondence
Alicia K. Smith, PhD, Department of Gynecology and Obstetrics, Emory University School of Medicine, 101 Woodruff Circle NE, Ste 4217, Atlanta 30322, GA.
Email: alicia.smith@emory.edu

Funding information
U.S. Department of Veterans Affairs, Grant numbers: 11S-RCS-009, 5I01CX000431, B9254-C, IK2CX000525; National Institute of Mental Health, Grant numbers: MH071537, MH096764, R01MH093500, R01MH108826, RC1MH088283, RO1MH079806; National Institute on Drug Abuse, Grant numbers: R01DA022720, R01DA022720-S1; Biomedical Laboratory Research and Development, VA Office of Research and Development, Grant numbers: 1I01BX002150, I01BX002577; Centers for Disease Control and Prevention, Grant number: U01OH010416, U01OH010718, and 200-2011-39410; Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant number: 916.11.086

Compelling evidence suggests that epigenetic mechanisms such as DNA methylation play a role in stress regulation and in the etiologic basis of stress related disorders such as Post traumatic Stress Disorder (PTSD). Here we describe the purpose and methods of an international consortium that was developed to study the role of epigenetics in PTSD. Inspired by the approach used in the Psychiatric Genomics Consortium, we brought together investigators representing seven cohorts with a collective sample size of $N = 1147$ that included detailed information on trauma exposure, PTSD symptoms, and genome-wide DNA methylation data. The objective of this consortium is to increase the analytical sample size by pooling data and combining expertise so that DNA methylation patterns associated with PTSD can be identified. Several quality control and analytical pipelines were evaluated for their control of genomic inflation and technical artifacts with a joint analysis procedure established to derive comparable data over the cohorts for meta-analysis. We propose methods to deal with ancestry population stratification and type I error inflation and discuss the advantages and disadvantages of applying robust error estimates. To evaluate our pipeline, we report results from an epigenome-wide association study (EWAS) of age, which is a well-characterized phenotype with known epigenetic associations. Overall, while EWAS are highly complex and subject to similar challenges as genome-wide association studies (GWAS), we demonstrate that an epigenetic meta-analysis with a relatively modest sample size can be well-powered to identify epigenetic associations. Our pipeline can be used as a framework for consortium efforts for EWAS.

KEYWORDS
EWAS, meta-analysis, stress, trauma

1 | INTRODUCTION

Traumatic events are reported by over 70% of individuals during their lifetime (Benjet et al., 2016). They have been associated with a number of deleterious outcomes, including posttraumatic stress disorder (PTSD), a psychiatric disorder characterized by cognitive intrusions, avoidance, negative alterations in thoughts and mood, and alterations in physiological arousal and reactivity (American Psychiatric Association, 2000). PTSD can be severe and disabling and is often associated with a range of comorbid psychiatric conditions such as depression and substance use disorders (Brady, Killeen, Brewerton, & Lucerini, 2000; Najt, Fusar-Poli, & Brambilla, 2011). PTSD has also been associated

with a 2.8-fold increase in suicidal thoughts and behaviors (Sareen, Houlahan, Cox, & Asmundson, 2005) as well as with a number of chronic medical conditions (Boscarino, 2008; Coughlin, 2011; David, Woodward, Esquenazi, & Mellman, 2004; Heppner et al., 2009; Jakovljevic et al., 2008; Kubzansky, Koenen, Jones, & Eaton, 2009; Kubzansky, Koenen, Vokonas, & Sparrow, 2007; Luft et al., 2012). While the risk of developing PTSD depends in part on the nature of the trauma (Kessler, 2000), only a minority of those exposed to trauma develop PTSD. As a result, despite the high prevalence of lifetime trauma, the overall lifetime prevalence of PTSD in the United States is 6.8%, (Breslau et al., 1998; Kessler et al., 2005; Resnick, Kilpatrick, Dansky, Saunders, & Best, 1993), which suggests there are individual differences in resilience that, if better understood, might inform the development of new approaches to prevention and treatment.

Genetic epidemiological studies suggest that both genetic and environmental factors contribute to PTSD risk. Twin studies estimate the heritability of PTSD to be between 30% and 70%, (Sartor et al., 2012; Sartor et al., 2011; Stein, Jang, Taylor, Vernon, & Livesley, 2002; True et al., 1993; Xian et al., 2000) with the remaining variance being attributed to environmental factors. Genetic research, based on both candidate gene and genome-wide association studies (GWAS), has provided support for the role of genetics in the development and severity of PTSD and has begun to identify variants that account for some of the genetic influence on PTSD. The genetic loci identified in the extant GWAS have been implicated in a variety of processes, including neuroprotection, actin polymerization, neuronal function, and immune function (Almli et al., 2014; Guffanti et al., 2013; Logue et al., 2013; Xie et al., 2013). Although promising, no robust genetic variants associated with PTSD have been identified and much work remains to be done to understand the biological basis of PTSD risk (Logue et al., 2015).

A growing body of work has explored the role of environmental influences on an individual's response to trauma. Given the dependence of PTSD development on exposure to environmental (i.e., traumatic) events, clarifying the ways in which environmental influences might affect biological function are critical to understanding the etiology of PTSD. In this regard, epigenetic mechanisms, which can mediate environmental influences on gene function, are particularly relevant. Epigenetic modifications, such as DNA methylation at cytosine-guanine dinucleotides (CpG sites), induce changes in gene expression in part through structural alterations of DNA that are maintained through each round of cell division; they respond to changes in the environment, are potentially reversible, and can be targeted for disease therapies (Feinberg, 2007). DNA methylation regulates gene expression by influencing the recruitment and binding of regulatory proteins to DNA. Typically, higher methylation at gene promoter regions correlates with decreased expression of that gene, while intragenic methylation can regulate alternative promoters and enhancers (Bonasio, Tu, & Reinberg, 2010; Maunakea et al., 2010).

Animal studies have demonstrated that epigenetic changes—particularly alterations in DNA methylation in response to nurturing—are related to altered responses to stress (Jirtle & Skinner, 2007; Weaver et al., 2004). Similar alterations have been reported in the human literature, in both central and peripheral tissues (McGowan et al., 2009; Tyrka, Price, Marsit, Walters, & Carpenter, 2012). Considering the influence of traumatic stress on DNA methylation seen in some studies (Vinkers et al., 2015), epigenetic-based investigations may extend genetic research findings. For example, research reporting an association of PTSD with a genetic variant in the PAC1 receptor (*ADCYAP1R1*; rs2267735) went beyond this finding to observe that PTSD severity was also correlated with methylation levels of the gene (Ressler et al., 2011). Other work has suggested an interactive effect between trauma burden and DNA methylation in the serotonin transporter locus (*SLC6A4*) on PTSD risk, independent of the widely studied length polymorphism at this same locus (Koenen et al., 2011) and a potentially interacting effect of genetic and epigenetic variation at the dopamine receptor (*SLC6A3*) on PTSD risk (Chang et al., 2012). Indeed, stress exposure itself has been shown to alter epigenetic patterns in both animal and human studies (Moser et al., 2015; Roth, Lubin, Funk, & Sweatt, 2009; Sipahi et al., 2014). In addition to candidate gene methylation studies, a small number of studies have examined genome-wide DNA methylation patterns in PTSD (Mehta et al., 2013; Smith et al., 2011; Uddin et al., 2010). In these first genome-wide studies of DNA methylation, immune dysregulation figured prominently among the biological networks associated with PTSD and, at a CpG site level, DNA methylation levels in several CpGs showed suggestive evidence of replication between these studies (Mehta et al., 2013; Smith et al., 2011; Uddin et al., 2010).

Although the emerging literature on epigenetic influences on PTSD is promising, the majority of research to date has been conducted with modest sample sizes, with inherent limited statistical power. Furthermore, studies of DNA methylation have been hampered by technical issues including batch effects (Harper, Peters, & Gamble, 2013) and blood cell composition (Houseman et al., 2012). Experience from large-scale genetic studies, such as the Psychiatric Genomics Consortium (PGC), underscores how large collaborative endeavors can provide the adequate sample sizes and the statistical power necessary to produce significant and replicable results (Logue et al., 2015). Based on this experience, the PGC-PTSD formed the epigenetics working group to organize an expansive network of investigators and their collection of samples with genome-wide DNA methylation data available for joint analyses.

Although there are many advantages to this collaborative approach, there are also challenges. First, methods for assessment of trauma exposure and PTSD symptoms/diagnosis differ among cohorts, requiring harmonization of the phenotypic data. Participating cohorts have assessed PTSD using clinical interviews and self-report measures with possible case diagnoses of lifetime PTSD, current PTSD, or "probable" PTSD available in across different studies. The larger PGC-PTSD has made substantial progress harmonizing phenotypes across studies and a similar approach will be adopted by the PGC-PTSD epigenetics group in its analyses. A second challenge of consortia is to address complications in the data sharing that is a prerequisite to increasing the power and reproducibility of the study. The trend in psychiatric genetics has been to establish large-scale consortia for the purpose of expanding sample sizes beyond what is possible based on

the data from any one group. Almost universally used in current large-scale genomic consortia is meta-analysis, through which the data from individual sites are analyzed separately and combined based on summary statistics. In many cases, meta-analysis of individual-level results yields results comparable to those of a "mega-analysis" of pooled data from different studies, (Lin & Zeng, 2010; Mathew & Nordstrom, 1999; Olkin & Sampson, 1998). However, there have been important distinctions in the way the meta-analysis strategy has been implemented across consortia, including variations in the degree to which the cleaning and analysis of the individual-level data occurs (i.e., centralized vs. distributed). The analysis of individual data in a centralized manner allows a high degree of control over the quality control (QC) process and an ability to quickly perform follow-up analyses, but poses difficulties of requiring a larger degree of computational resources and storage at the consortium level and permission from all groups to share data.

The PGC-PTSD epigenetics group uses many of the same protocols and tools developed by the PGC (Logue et al., 2015), with the difference that not all data are centrally stored and managed since some constituent samples that originate from US military, US Veteran (VA), or foreign countries are subject to additional regulatory oversight, which do not allow the sharing of individual-level genomic data. To enable participation for these studies, the PGC-PTSD epigenetics group follows a strategy similar to that of the ENIGMA consortium (Thompson et al., 2014), in which a set of protocols and scripts are created, in this case to implement standardized QC and analysis pipelines for the Illumina HumanMethylation450 BeadChip. These scripts are performed at each participating site and analysis results are submitted to the consortium where they are assessed, collated, and meta-analyzed. In this study, we compare the performance of two QC and two analytical pipelines to control for genomic inflation, present the final PGC-PTSD epigenetics pipeline, and assess the performance of the PGC-PTSD epigenetics pipeline in a meta-analysis of age.

2 | MATERIALS AND METHODS

2.1 | The PGC EWAS cohorts

The participating cohorts, presented in Table 1, consisted of four military cohorts (MRS, PRISMO, VA-M, and VA-NCPTSD) and three civilian cohorts (DNHS, GTP, and WTC) that all measured DNA methylation (DNAm) with the Illumina HumanMethylation450k BeadChip. Descriptions of the cohorts are in the supplemental information. Each cohort consists of PTSD cases as well as trauma-exposed controls. A total of 1,147 subjects (~50% cases) were selected for inclusion in the EWAS and were subjected to the quality control and analytical pipelines.

2.2 | Posttraumatic stress disorder assessment

Similar to other analyses conducted by the PGC-PTSD, our analysis required consistently defining and harmonizing PTSD diagnoses across cohorts that used different instruments and methods of diagnosis (Logue et al., 2015). We used a diagnosis of current PTSD based on the diagnostic criteria defined by each cohort's principal investigator (see supplemental information). Individuals with lifetime diagnoses of PTSD but not current PTSD were excluded from analysis.

2.3 | Quality control procedures

We tested two quality control protocols: the eventually proposed PGC pipeline and a Functional Normalization (Funnorm) pipeline. In the PGC pipeline (Supplemental Figure S1), study investigators first conducted a visual inspection of control-probes designed to report on each step of the Infinium protocol such as bisulfite conversion and hybridization efficiency. In addition, samples with probe detection call rates <90% and those with an average intensity value of either <50% of the experiment-wide sample mean or <2,000 arbitrary units (AU) were excluded. Probes with detection *p*-values > 0.001 or those based on less than three beads were set to missing as were probes that cross-hybridized between autosomes and sex chromosomes (Teschendorff et al., 2013). CpG sites with missing data for >10% of samples within cohorts were excluded from analysis. Probes containing single nucleotide polymorphisms (SNPs; based on 1000 Genomes) within 10 base pairs of the target CpG were maintained in each dataset, but flagged and tracked throughout the analysis pipeline. This decision was based on the growing recognition that sequence variants can influence DNA methylation patterns throughout the genome (Smith et al., 2014). Even if an associated CpG site is influenced by genetics, such as is the case for *SKA2* (Boks et al., 2016; Rice et al., 2008), maintaining such probes is informative to our overall goal of identifying genes important for PTSD (Gibbs et al., 2010; Guintivano et al., 2014; Heyn et al., 2013). Normalization of probe distribution and background differences between Type I and Type II probes was conducted using Beta Mixture Quantile Normalization (BMIQ) (Teschendorff et al., 2013) after background correction. We chose BMIQ after comparing distributions of BMIQ normalized Type II probes in the Detroit Neighborhood Health Study (DNHS) with the raw distributions and distributions after applying the DASEN procedure in the R package watermelon (Supplemental Figure S2) (Pidsley et al., 2013). Following normalization, batch effect removal as implemented in the ComBat procedure of the SVA package in bioconductor was used to account for sources of technical variations including batch and positional effects, which can cause spurious associations (Johnson, Li, & Rabinovic, 2007). Individual cohorts also controlled for additional covariates that may not have been balanced within chips but that were of interest in downstream analyses, such as case designation and sex (if relevant). Following completion of this QC pipeline, each cohort confirmed that there were no remaining sources of technical variation by examining the association of PCs of the methylation levels with chip and position using multivariate linear regression, bar plots, and heat maps.

The second QC protocol used the Functional Normalization procedure implemented in the R package minfi, which has been reported to remove technical variation more effectively than ComBat or other supervised methods (Aryee et al., 2014; Fortin et al., 2014).

TABLE 1 PGC-PTSD EWAS Cohorts

		Civilian			Military			
	Total	DNHS	GTP	WTC	MRS	PRISMO	VA-M	VA-NCP
N	1147	100	270	180	126	62	176	233
Mean age (SD)	38.0 (–)	53.6 (14.0)	41.7 (12.4)	49.7 (8.3)	22.2 (3.0)	27.1 (9.2)	34.9 (9.9)	32.0 (8.4)
Current PTSD (%)	50	40	39	47	50	50	49	69
Male (%)	73	38	30	100	100	100	78	90
Race (%)								
White	56	15	5	76	57	100	100	74
Hispanic	6	0	0	0	25	0	0	14
Black	33	85	94	4	8	0	0	9
Asian	1	0	0	0	3	0	0	2
Other	3	0	0	20	0	0	0	0

Briefly, raw IDATs were loaded into R and Functional Normalization applied using the default of two principal components (PCs) as covariates. The resulting normalized beta matrix was then extracted and used for analysis.

2.4 | Cellular heterogeneity

DNA methylation is known to vary by cell type which impacts the measured whole blood methylation as a result of the amalgamation of the cell type proportions in each individual's sample. To control for possible confounding by individuals' underlying cell type heterogeneity, proportions of CD8, CD4, NK, B cells, monocytes, and granulocytes were estimated using each individual's DNA methylation data, publicly available reference data (GSE36069), and the method described by Jaffe and Irizarry (2014) (Reinius et al., 2012), based on the Houseman algorithm (Houseman et al., 2012). CD8, CD4, NK, B cell, and monocyte cell proportions were included as covariates in our statistical analyses.

2.5 | Ancestry

Accounting for population stratification has become routine practice for genetic association studies, and most recently has also been shown to be of importance in DNA methylation studies (Barfield et al., 2014; Nielsen et al., 2010). GWAS methods such as principal components (PCs) derived from SNPs can be incorporated into EWAS, but were not always available for all cohorts, or all samples within a cohort. An alternative based on methylation probes that proxy nearby SNPs was developed by Barfield et al. (2014) for use in European and African American subjects. Here we evaluated and extended this approach to other ancestral populations as part of the PGC-PTSD EWAS pipeline.

A subset of ancestry-diverse subjects ($N = 128$, including European Americans, African Americans, Latinos/Native Americans, and "others" including East Asians) from the Marine Resiliency study (MRS) were selected based on available genome-wide genotype data (Illumina HumanOmniExpressExome array) and matching Illumina 450K methylation data (Nievergelt et al., 2015). Ancestry using GWAS data was inferred as described in Nievergelt et al. (2013). In brief, genotypes of 1,783 ancestry-informative markers (AIMs) were used to determine a subject's ancestry at the continental level using STRUCTUREv2.3.2.1, including prior population information of the HGDP reference set (Falush, Stephens, & Pritchard, 2003; Li et al., 2008). Based on these ancestry estimates, subjects were placed into one of four groups: European Americans, African Americans, Latinos/Native Americans, and "Others." PCs were derived using Eigenstrat (Price et al., 2006).

Ancestry estimates using methylation data were derived using subsets of methylation probes in close proximity to SNPs identified by Barfield et al. (2014). Probe sets with 0 bp distance ($N = 7,703$ CpG probes), within 1 bp distance (17,995 CpG probes), and within 10 bp distance ($N = 50,319$ CpG probes) were compared. GWAS-derived PCs were visually compared to methylation-probe derived PCs and genotypes of SNPs in proximity of CpG sites were compared with respective CpG methylation values using Pearson correlation (r).

2.6 | Statistical analysis

Within each cohort, logit transformed β values (M-values) (Du et al., 2010) were modeled by linear regression as a function of PTSD, adjusting for sex, age, the estimated cell proportions, and ancestry using PCs. For cohorts with available GWAS data, the first three PCs from the GWAS were used. For cohorts without GWAS data, the method described by Barfield et al. was used to generate ancestry PCs directly from the EWAS data. Consistent with the original paper and our analysis (full results below), the second through fourth PCs were used as covariates in the model to control for ancestry. Note that while ancestry is a primary source for variation in GWAS, other potentially confounding factors such as cellular heterogeneity are a primary source for variation in EWAS data. Comparison with SNP data showed that ancestry inference is strongest when excluding EWAS-derived PC1. QQ-plots of the PTSD p-values were examined for evidence of genomic inflation due to unaccounted technical variation or other

confounders. In addition, the genomic inflation factors (λ) were calculated for each study. Two adjustments were considered to improve the precision of the estimated variances. First, moderated *t*-statistics were calculated using the empirical Bayes method implemented in the R package limma (Smyth, 2005). Second, HC3 robust standard errors, which have been shown to be the most effective in samples smaller than 250, were calculated using the R package sandwich (Long & Ervin, 2000; Zeileis, 2004).

Cohort-level analysis results were combined using the inverse normal method (Marot, Foulley, Mayer, & Jaffrézic, 2009). Briefly, one-sided *p*-values for each CpG site in each study were calculated from the *t*-statistics. Next *z*-scores were calculated from the one-sided *p*-values and weighted by the number of subjects in each study relative to the total in the meta-analysis. Two-sided *p*-values of the *z*-score were then calculated and genomic inflation examined. Finally, *p*-values were adjusted for multiple-testing using the False Discovery Rate (FDR) procedure at the type I error rate level of 5% (Benjamini & Hochberg, 1995).

2.7 | Sensitivity analysis

Numerous robust associations between age and DNAm have been reported (Bocklandt et al., 2011; Hannum et al., 2013; Horvath, 2013; Horvath et al., 2012; Weidner et al., 2014). Because we expect that age will be associated similarly with DNAm in each cohort, we can leverage this highly reproducible relationship to evaluate the pipeline's performance. In each cohort, we modeled methylation as a function of age along with covariates for ancestry and gender (if applicable) and meta-analyzed the results as outlined above. In addition, we measured concordance between studies by estimating the correlation between the *t*-statistics of the age variable.

2.8 | Power

Power for EWAS is more favorable compared to GWAS as a result of the continuous nature of the DNA methylation measures, but instead suffers from poor distribution including low variances and heteroscedasticity (Du et al., 2010). The power to detect a differentially methylated CpG site depends on the percent difference in methylation between cases and controls, the pooled variation in methylation (σ) across CpG sites, and the number of cases and controls (Liu & Hwang, 2007). We conducted simulations to test the smallest mean difference between PTSD cases and controls we could detect based on our projected sample size and a $\sigma = 0.43$, which represents the highest variation reported in one of our cohorts, and thus a very conservative estimate (Orr & Liu, 2009).

3 | RESULTS

3.1 | Participating cohorts

Sample characteristics for studies that have contributed data to this first PGC-PTSD EWAS study are listed in Table 1 ($N = 1{,}147$). Three of the seven studies are composed of civilians, while the remaining studies include active duty and veteran military populations. The majority of participants, especially from the military cohorts, were male (73%) and of European American (EA) descent (56%). All participants were exposed to trauma and half of participants suffered from current PTSD (50%). Data collection occurred across the United States (e.g., Atlanta, Detroit, San Diego, Durham, Boston, and New York) and Europe. While a few studies used clinical interviews, the majority of studies used self-report ratings of PTSD symptoms that relied on established cutoffs to assign caseness. A detailed description of participating cohorts is provided in the supplementary information.

3.2 | Power

The power analysis shows that with our sample of 573 cases and 574 controls ($N = 1147$), we are sufficiently powered to find at least one CpG site with a mean methylation difference of 0.08 between cases and controls (Figure 1).

3.3 | Ancestry

We investigated the utility of DNA methylation-based ancestry estimates based on CpGs with nearby genetic variants in cis as proposed by Barfield et al. (2014). A comparison of CpG probes with SNPs within 1 bp distance ($N = 17{,}995$) and CpG probes with SNPs within 10 bp distance ($N = 50{,}319$) showed a higher genotype-methylation correlation for the 1 bp probes ($r = 0.29$, $p = 1.8 \times 10^{-15}$) than the 10 bp probes ($r = 0.06$, $p = 0.0015$). Figure 2 shows a SNP derived PC plot based on available GWAS data including PCs 1 and 2 (panel A), PCs 2 and 3 from the methylation-based CpGs with a SNP 1 bp from the probes (panel B), and the PCs from CpG probes within

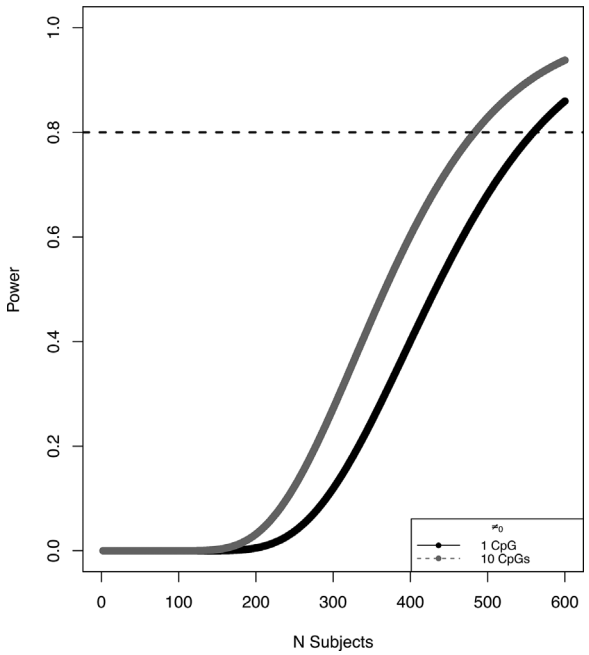

FIGURE 1 Sample size vs. power to to detect differentially methylated CpG sites. The black curve indicates the number of cases and controls necessary to find a differentially methylated if only one CpG site exists, while the gray line indicates the size necessary if 10 differentially methylated sites exist

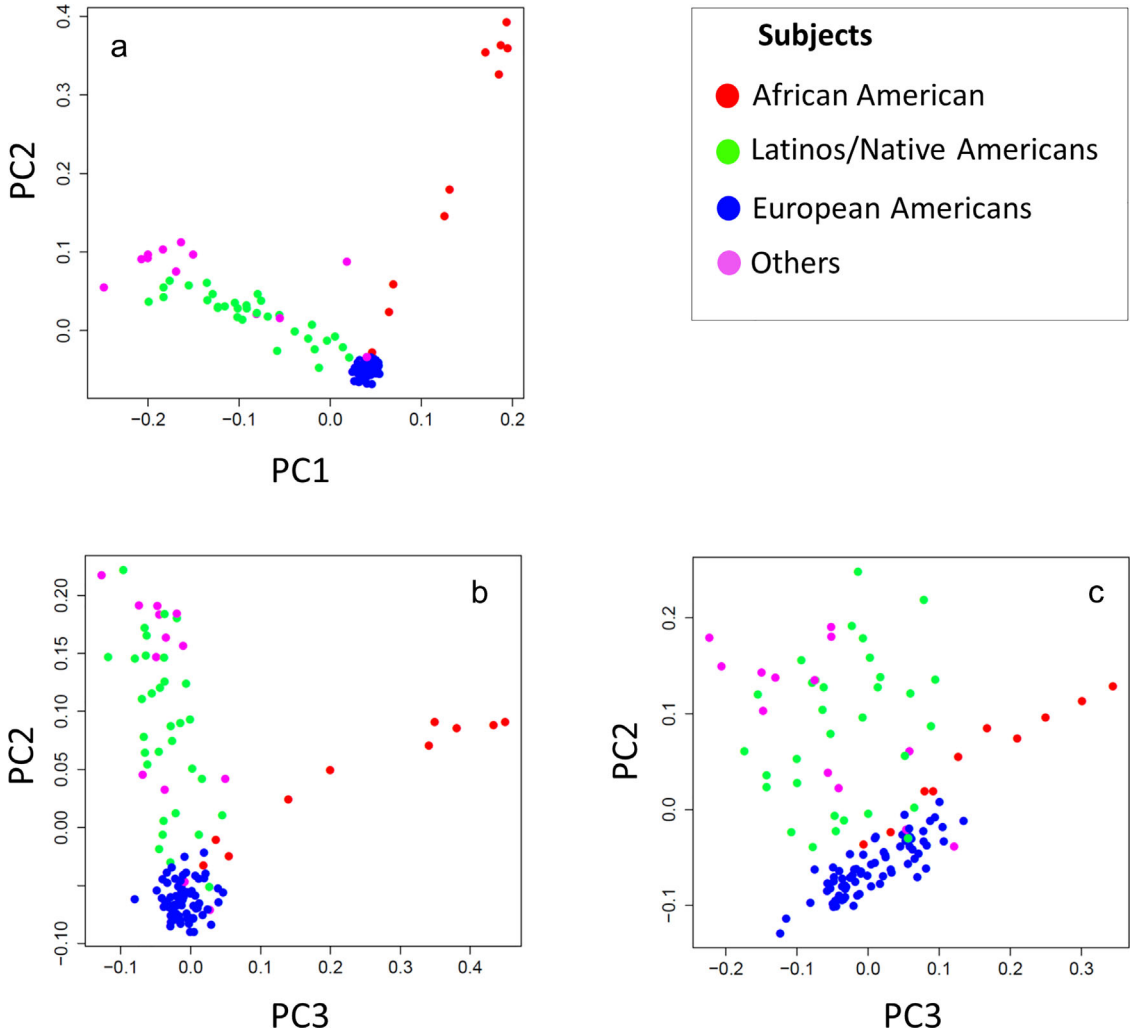

FIGURE 2 Ancestry inference using SNPs versus methylation probes in 128 participants of the Marine Resiliency Study (MRS). (a) Principal component (PC) plot showing ancestry inferred using SNPs from a genome-wide association study (GWAS). PC plots based on CpG probes with SNPs within 1 bp distance (b) and with SNPs within 10 bp distance (c), respectively. Subject are placed into four ancestral groups based on ancestry estimates using ancestry-informative SNPs and a reference panel (see methods).

10 bp of a SNP (panel C). Supplemental Figure S5 shows PCs 2, 3, 4, and PCs 2, 3, and 6 for, respectively, CpGs within 1 bp and 10 bp of a SNP. These results along with the genotype-methylation correlations show that the use of DNA methylation ancestry estimates (PCs 2–4) using probes within 1 bp of a SNP provide reliable results and are suitable as ancestry covariates in our analyses.

3.4 | Quality control results

The number of samples and probes, not including cross-reactive probes) removed in our proposed PGC-EWAS pipeline ranged from 677 to 10,218 across studies (Supplemental Table S3). Figure 3 presents the genomic inflation factors from the analysis of PTSD for each individual study using two different quality control methods as well as two different analysis pipelines: (1) our proposed PGC-EWAS pipeline, described in detail in the Methods section above, as well as (2) the Functional Normalization (Funnorm) QC pipeline. Resulting data from each QC pipeline were then subjected to linear regression analysis, performed with empirical Bayes and HC3 standard errors, respectively. In studies using the Funnorm pipeline there were large variations in the genomic inflation factor with two studies showing high inflation (DNHS, GTP) and two studies showing substantial deflation (MRS, VA-NCP) regardless of whether empirical Bayes or HC3 standard errors were used. Using the PGC-EWAS pipeline and HC3 standard errors, six of the seven studies showed genomic deflation with ($\lambda < 1.0$), while only one study was deflated when using empirical Bayes standard errors (Supplemental Table S1). These results indicate that the PTSD-PGC EWAS pipeline, combined with empirical Bayes standard errors, is the preferred method for cohorts participating in our meta-analysis.

3.5 | Sensitivity analysis: Age meta-analysis results

Results for the age analysis using our pipeline and no standard error adjustment are presented in Table 2. All studies reported numerous

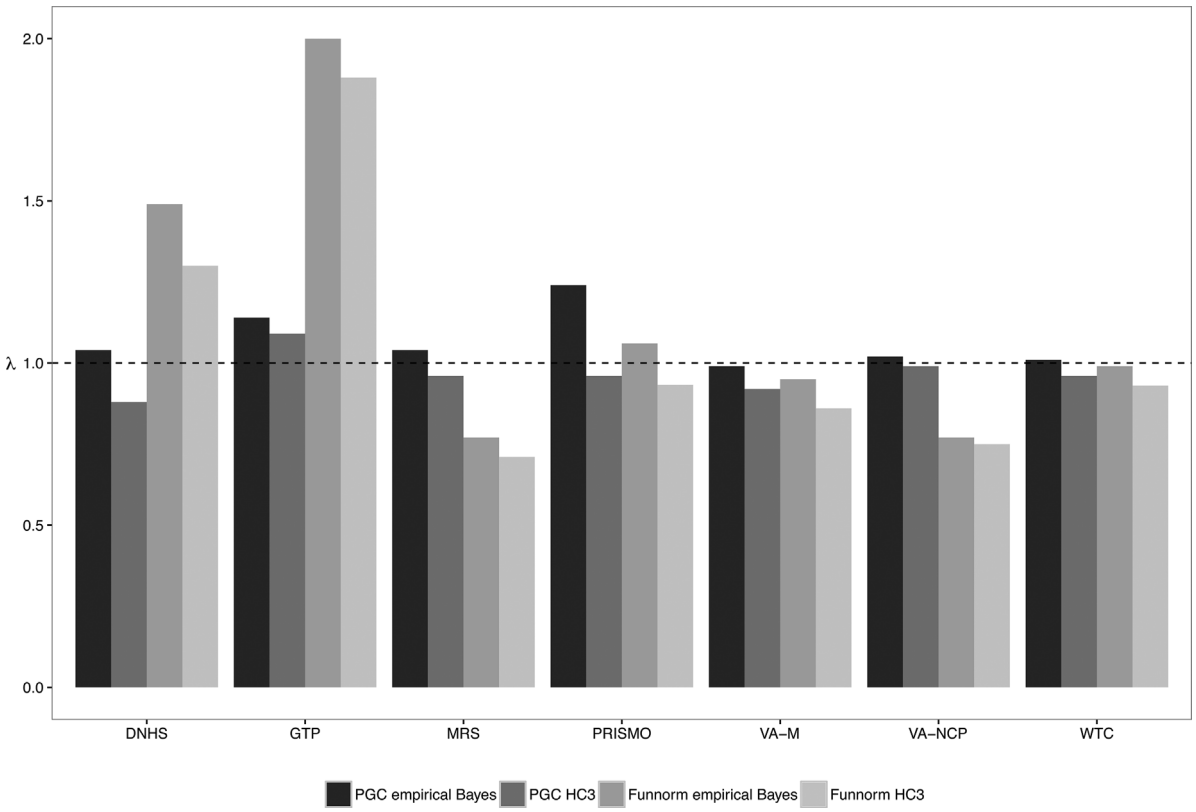

FIGURE 3 PTSD genomic inflation factors (λ) by quality control pipeline (PGC vs. Funnorm) and standard error adjustment method (empirical Bayes vs. HC3)

FDR-significant CpG sites but substantially more significant results were reported for the combined meta-analysis. MRS and PRISMO reported the fewest significant sites. However the age range for participants in these studies was narrower as both studies included active military personnel. The correlations of the *t*-statistics ranged from 0.171 to 0.692 when all sites were analyzed and from 0.441 to 0.886 among the FDR significant sites (Supplemental Figure S3). The strong correlations of the most significant sites indicate that each cohort retained the biological signal of age after QC. Of the 494 CpG sites reported to have been associated with age, 326 were significant after FDR-correction (Supplemental Table S2). In addition, a forest plot of the most significant CpG site representative of the FDR significant sites, shows a consistent direction of effect in each study (Supplemental Figure S4).

4 | DISCUSSION

PTSD is unique among psychiatric disorders in that its occurrence requires exposure to a significant traumatic event. With an environmental exposure embedded into the etiology of the disorder, the PTSD diagnosis affords an unusual opportunity to identify individual differences in the biological response to trauma to increase risk for, or resilience to, the disorder. Here we have introduced an international

TABLE 2 Age associations using the PGC-PTSD epigenetics QC and analysis pipeline

Study	Sites	Sites with (FDR < 0.05)	Sites with ($p < 5 \times 10^{-5}$)	Sites with ($p < 5 \times 10^{-6}$)	Sites with ($p < 5 \times 10^{-7}$)
DNHS	455,079	4,766	1,744	678	299
GTP	453,351	59,100	21,562	14,299	9,586
MRS	455,601	210	311	99	34
PRISMO	446,688	246	316	121	41
VA-M	455,641	42,474	12,913	7,213	4,159
VA-NCPTSD	453,747	35,217	10,522	6,331	3,991
WTC	455,340	14,239	5,013	2,730	1,525
Meta-analysis	444,164	119,308	57,332	46,629	38,656

collaboration that has been established to identify epigenetic associations—specifically, DNA methylation—related to risk for, or resilience to, PTSD. We presented the development of a consistent pipeline for processing and quality-control of epigenome-wide association data comparing two quality control approaches and statistical pipelines. In our analysis of PTSD, we found that our proposed PGC-EWAS pipeline controlled for genomic inflation and deflation more consistently than Functional Normalization, regardless of the standard error correction used. In light of these findings, we encourage consortia with epigenome-wide methylation data to implement our quality control pipeline including checks for genomic inflation and strengthened associations with age before meta-analyzing across studies.

Through these collaborative efforts to analyze existing DNA methylation data from blood obtained from both military and civilian cohorts, we are poised to collectively address one of the main challenges of psychiatric genomics, namely the need for large, harmonized samples to adequately power genome-scale analyses. The current collaborative dataset allows detection of methylation differences around seven percent, larger than most reported methylation differences, (Vinkers et al., 2015). Additional EWAS datasets that are forthcoming will likely prove essential to detecting PTSD-associated DNA methylation differences in our planned meta-analyses. Moving forward, we anticipate that our collaborative efforts will grow to include additional cohorts from around the world; indeed, in the last year alone, several new studies have expressed interest in participating in future EWAS analyses as their data become available. In addition to the DNA methylation analysis the close allegiance with the PGC-PTSD group has laid the foundation for integrating data from GWAS, EWAS, and gene expression/transcriptome analyses. In combination with other biological measures and coordinated neuro-imaging efforts (Logue et al., 2015) that may become accessible through this collaboration, these system-wide integrations will facilitate a more complete understanding of the molecular architecture and biological underpinnings of PTSD.

The harmonization of some study characteristics paired with the demographic and clinical diversity of the samples, including the differences between military and civilian trauma, allows us the opportunity to identify DNA methylation patterns predictive for specific groups of individuals and types of trauma. This will not only provide insight into the heterogeneity of PTSD, but may also help explain mechanisms for the variation in conditional effects of different types of trauma on PTSD (Wisco et al., 2014). Additionally, it will also provide a framework from which DNA methylation may be informative for early risk prediction and treatment stratification.

Looking ahead, we are optimistic that our PTSD EWAS collaboration will identify blood-based DNA methylation signatures that associate reliably with PTSD. Identification of robust peripheral biomarkers is an important first step and has potential for early detection and prevention. The ultimate goal is to provide new insights into the etiology of PTSD. To truly understand the mechanistic basis of PTSD, it will be critical to compare our blood-derived epigenetic biomarkers with those from other tissues, in particular brain tissue. As a first pass, DNA methylation-based biomarkers that associate with PTSD at particular CpG sites in blood can, at this time, be compared to CpG site derived from brain tissues, thanks to the Epigenomic Roadmap datasets (Bernstein et al., 2010; Kundaje et al., 2015). However, these comparisons will be limited to a comparison of DNA methylation levels in brain and blood in general, as the current data are not representative of PTSD. Over time, however, the development of PTSD brain biobanks of brain and other tissue types including blood cells, will help us to pinpoint whether blood-derived, DNA methylation biomarkers of PTSD reflect similar alterations in brain tissue, as recent work suggests may be possible for certain pathways (Daskalakis, Cohen, Cai, Buxbaum, & Yehuda, 2014). Collectively, these cross-tissue efforts will provide insight into the biological pathways underlying PTSD vulnerability and will ultimately facilitate new treatment and modes of prevention.

ACKNOWLEDGMENTS

This work was supported by the U.S. Army Medical Research and Materiel Command and the National Institute of Mental Health (NIMH; R01MH108826) as well as the Biomedical and Laboratory Research and Development (#I01BX002577). We appreciate the technical support of all of the staff, volunteers and participants from the Grady Trauma Project, supported by the National Institutes of Mental Health (MH096764 and MH071537). DNHS, which is grateful to all of the participants and staff for their contributions, was funded by NIH Awards R01DA022720, R01DA022720-S1, and RC1MH088283. The Marine Corps, Navy Bureau of Medicine and Surgery (BUMED) and VA Health Research and Development (HSR&D) provided funding for MRS data collection and analysis and NIH R01MH093500 funded the GWAS assays and analysis. Acknowledged are Victoria B. Risbrough Ph.D (VA San Diego Healthcare System & UCSD), Mark A. Geyer (UCSD), Daniel T. O'Connor (UCSD), all MRS investigators, and all MRS investigators, as well as the MRS administrative core and data collection staff listed in the Methods article (Baker et al. Prev Chronic Dis. 2012;9(10):E97). The authors also thank the Marine and Navy Corpsmen volunteers for military service and participation in MRS. Data collection of PRISMO was funded by the Dutch Ministry of Defence, and DNA methylation analyses were funded by the VENI Award fellowship from the Netherlands Organisation for Scientific Research (NWO, grant number 916.11.086). The VA Boston-National Center for PTSD Study research was supported in part by National Institute of Mental Health Award RO1MH079806, Department of Veterans Affairs, Clinical Science Research & Development Program Award 5I01CX000431-02, Department of Veterans Affairs, Biomedical Laboratory Research & Development Program Award 1I01BX002150-01, the Translational Research Center for TBI and Stress Disorders (TRACTS), a VA Rehabilitation Research and Development Traumatic Brain Injury Center of Excellence (B9254-C), and the Cooperative Studies Program, Department of Veterans Affairs. This research is the result of work supported with resources and the use of facilities at the Pharmacogenomics Analysis Laboratory, Research and Development Service, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas. This work was also

supported by a Career Development Award to E. J. Wolf from the Department of Veterans Affairs, Clinical Sciences Research, and Development Program. Dr. Kimbrel was supported by a Career Development Award (#IK2CX000525) from the Clinical Science Research and Development (CSR&D). Dr. Beckham was supported by a Research Career Scientist Award (#11S-RCS-009) from the CSR&D Service of VA ORD. This research was also supported, in part, by a Merit Award (#I01BX002577) from the Biomedical Laboratory Research and Development (BLR&D) Service of VA ORD. The VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center Workgroup includes John A. Fairbank, Mira Brancu, Patrick S. Calhoun, Eric A. Dedert, Eric B. Elbogen, Kimberly T. Green, Robin A. Hurley, Angela C. Kirby, Jason D. Kilts, Christine E. Marx, Gregory McCarthy, Scott D. McDonald, Marinell Miller-Mumford, Scott D. Moore, Rajendra A. Morey, Jennifer C. Naylor, Treven C. Pickett, Jared Rowland, Jennifer J. Runnals, Cindy Swinkels, Steven T. Szabo, Katherine H. Taber, Larry A. Tupler, Elizabeth E. Van Voorhees, H. Ryan Wagner, Richard D. Weiner, and Ruth Yoash-Gantz. Data collection from the World Trade Center (WTC) cohort was supported by CDC awards U01OH010416, U01OH010718, and 200-2011-39410. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the VA, NIH, or the United States government.

CONFLICTS OF INTEREST

Dr. Youssef's disclosures include Speaker CME honoraria from the Georgia Department of Behavioral Health and Developmental Disabilities (DHBDD). Dr. Youssef received research support from the Department of Veteran Affairs and The Augusta Biomedical Research Corporation. Dr. Stein has in the past 3 years received payments for editorial work from UpToDate, Biological Psychiatry, and Depression and Anxiety. He has also in the past 3 years been paid as a consultant for Actelion Pharmaceuticals, Janssen, Pfizer, Resilience Therapeutics, and Tonix Pharmaceuticals. No other author declares any conflict of interest.

REFERENCES

Almli, L. M., Srivastava, A., Fani, N., Kerley, K., Mercer, K. B., Feng, H., . . . Ressler, K. J. (2014). Follow-up and extension of a prior genome-wide association study of posttraumatic stress disorder: Gene x environment associations and structural magnetic resonance imaging in a highly traumatized African-American civilian population. *Biological Psychiatry*, *76*(4), e3.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363–1369.

American Psychiatric Association. (2000). DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision. Washington, DC: American Psychiatric Association 75.

Barfield, R. T., Almli, L. M., Kilaru, V., Smith, A. K., Mercer, K. B., Duncan, R., . . . Conneely, K. N. (2014). Accounting for population stratification in DNA methylation studies. *Genetic Epidemiology*, *38*(3), 231–241.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*(1), 289–300.

Benjet, C., Bromet, E., Karam, E. G., Kessler, R. C., McLaughlin, K. A., Ruscio, A. M., . . . Koenen, K. C. (2016). The epidemiology of traumatic event exposure worldwide: Results from the World Mental Health Survey Consortium. *Psychological Medicine*, *46*(2), 327–343.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., . . . Thomson, J. A. (2010). The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology*, *28*(10), 1045–1048.

Bocklandt S., Lin W., Sehl M. E., Sánchez F. J., Sinsheimer J. S., Horvath S., & Vilain E. (2011). Epigenetic predictor of age.

Boks, M. P., Rutten, B. P., Geuze, E., Houtepen, L. C., Vermetten, E., Kaminsky, Z., & Vinkers, C. H. (2016). SKA2 methylation is involved in cortisol stress reactivity and predicts the development of post-traumatic stress disorder (PTSD) after military deployment. *Neuropsychopharmacology*, *41*(5), 1350–1356.

Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular signals of epigenetic states. *Science*, *330*(6004), 612–616.

Boscarino, J. A. (2008). A prospective study of PTSD and early-age heart disease mortality among Vietnam veterans: Implications for surveillance and prevention. *Psychosomatic Medicine*, *70*(6), 668–676.

Brady, K. T., Killeen, T. K., Brewerton, T., & Lucerini, S. (2000). Comorbidity of psychiatric disorders and posttraumatic stress disorder. *Journal of Clinical Psychiatry*, *61*(Suppl7), 22–32.

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry*, *55*(7), 626–632.

Chang, S.-C., Koenen, K. C., Galea, S., Aiello, A. E., Soliven, R., Wildman, D. E., & Uddin, M. (2012). Molecular variation at the SLC6A3 locus predicts lifetime risk of PTSD in the Detroit Neighborhood Health Study. *PLoS ONE*, *7*(6), e39184.

Coughlin, S. S. (2011). Post-traumatic stress disorder and cardiovascular disease. *The Open Cardiovascular Medicine Journal*, *5*, 164–170.

Daskalakis, N. P., Cohen, H., Cai, G., Buxbaum, J. D., & Yehuda, R. (2014). Expression profiling associates blood and brain glucocorticoid receptor signaling with trauma-related individual differences in both sexes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(37), 13529–13534.

David, D., Woodward, C., Esquenazi, J., & Mellman, T. A. (2004). Comparison of comorbid physical illnesses among veterans with PTSD and veterans with alcohol dependence. *Psychiatric Services*, *55*(1), 82–85.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*, *11*(1), 587.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature*, *447*(7143), 433–440.

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., . . . Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, *15*(11), 503.

Gibbs, J. R., van der Brug, M. P., Hernandez, D. G., Traynor, B. J., Nalls, M. A., Lai, S.-L., . . . Troncoso, J. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genetics*, *6*(5), e1000952.

Guffanti, G., Galea, S., Yan, L., Roberts, A. L., Solovieff, N., Aiello, A. E., . . . Koenen, K. C. (2013). Genome-wide association study implicates a

novel RNA gene, the lincRNA AC068718.1, as a risk factor for post-traumatic stress disorder in women. *Psychoneuroendocrinology*, *38*(12), 3029–3038.

Guintivano, J., Brown, T., Newcomer, A., Jones, M., Cox, O., Maher, B. S., … Kaminsky, Z. A. (2014). Identification and replication of a combined epigenetic and genetic biomarker predicting suicide and suicidal behaviors. *American Journal of Psychiatry*, *171*(12), 1287–1296.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., … Gao, Y. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*(2), 359–367.

Harper, K. N., Peters, B. A., & Gamble, M. V. (2013). Batch effects and pathway analysis: Two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiology Biomarkers & Prevention*, *22*(6), 1052–1060.

Heppner, P. S., Crawford, E. F., Haji, U. A., Afari, N., Hauger, R. L., Dashevsky, B. A., … Baker, D. G. (2009). The association of posttraumatic stress disorder and metabolic syndrome: A study of increased health risk in veterans. *BMC Medicine*, *7*, 1.

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., … Wang, L. (2013). DNA methylation contributes to natural human variation. *Genome Research*, *23*(9), 1363–1372.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115.

Horvath, S., Zhang, Y., Langfelder, P., Kahn, R. S., Boks, M., van Eijk, K., … Ophoff, R. A. (2012). Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biology*, *13*(10), R97.

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., … Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*(1), 1.

Jaffe, A. E., Irizarry, R. A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology*, *15*(2), R31.

Jakovljevic, M., Babic, D., Crncevic, Z., Martinac, M., Maslov, B., & Topic, R. (2008). Metabolic syndrome and depression in war veterans with post-traumatic stress disorder. *Psychiatria Danubina*, *20*(3), 406–410.

Jirtle, R. L., Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility. *Nature Reviews: Genetics*, *8*(4), 253–262.

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127.

Kessler, R. C. (2000). Posttraumatic stress disorder: The burden to the individual and to society. *The Journal of Clinical Psychiatry*, *61*(suppl 5), 4–12.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602.

Koenen, K. C., Uddin, M., Chang, S. C., Aiello, A. E., Wildman, D. E., Goldmann, E., & Galea, S. (2011). SLC6A4 methylation modifies the effect of the number of traumatic events on risk for posttraumatic stress disorder. *Depression and Anxiety*, *28*(8), 639–647.

Kubzansky, L. D., Koenen, K. C., Jones, C., & Eaton, W. W. (2009). A prospective study of posttraumatic stress disorder symptoms and coronary heart disease in women. *Health Psychology*, *28*(1), 125–130.

Kubzansky, L. D., Koenen, K. C., Spiro, A., 3rd, Vokonas, P. S., & Sparrow, D. (2007). Prospective study of posttraumatic stress disorder symptoms and coronary heart disease in the Normative Aging Study. *Archives of General Psychiatry*, *64*(1), 109–116.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., … Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, *319*(5866), 1100–1104.

Lin, D. Y., Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*, *97*(2), 321–332.

Liu, P., Hwang, J. G. (2007). Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics*, *23*(6), 739–746.

Logue, M. W., Amstadter, A. B., Baker, D. G., Duncan, L., Koenen, K. C., Liberzon, I., … Ressler, K. J. (2015). The psychiatric genomics consortium posttraumatic stress disorder workgroup: Posttraumatic stress disorder enters the age of large-scale genomic collaboration. *Neuropsychopharmacology*, *40*(10), 2287–2297.

Logue, M. W., Baldwin, C., Guffanti, G., Melista, E., Wolf, E. J., Reardon, A. F., … Koenen, K. C. (2013). A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. *Molecular Psychiatry*, *18*(8), 937–942.

Long, J. S., Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217–224.

Luft, B., Schechter, C., Kotov, R., Broihier, J., Reissman, D., Guerrera, K., … Friedman-Jimenez, G. (2012). Exposure, probable PTSD and lower respiratory illness among World Trade Center rescue, recovery and clean-up workers. *Psychological Medicine*, *42*(05), 1069–1079.

Marot, G., Foulley, J.-L., Mayer, C.-D., & Jaffrézic, F. (2009). Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics*, *25*(20), 2692–2699.

Mathew, T., Nordstrom, K. (1999). On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics*, *55*(4), 1221–1223.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., … Zhao, Y. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, *466*(7303), 253–257.

McGowan, P. O., Sasaki, A., D'Alessio, A. C., Dymov, S., Labonté, B., Szyf, M., … Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, *12*(3), 342–348.

Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K., Altmann, A., Pace, T. W., … Mercer, K. B. (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proceedings of the National Academy of Sciences*, *110*(20), 8302–8307.

Moser, D. A., Paoloni-Giacobino, A., Stenz, L., Adouan, W., Manini, A., Suardi, F., … Rusconi-Serpa, S. (2015). BDNF methylation and maternal brain activity in a violence-related sample. *PLoS ONE*, *10*(12).

Najt, P., Fusar-Poli, P., & Brambilla, P. (2011). Co-occurring mental and substance abuse disorders: A review on the potential predictors and clinical outcomes. *Psychiatry Research*, *186*(2), 159–164.

Nielsen, D. A., Hamon, S., Yuferov, V., Jackson, C., Ho, A., Ott, J., & Kreek, M. J. (2010). Ethnic diversity of DNA methylation in the OPRM1 promoter region in lymphocytes of heroin addicts. *Human Genetics*, *127*(6), 639–649.

Nievergelt, C. M., Maihofer, A. X., Mustapic, M., Yurgil, K. A., Schork, N. J., Miller, M. W., … O'Connor, D. T. (2015). Genomic predictors of combat stress vulnerability and resilience in US Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. *Psychoneuroendocrinology*, *51*, 459–471.

Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K., & Kidd, J. R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative Genetics*, *4*(1), 13.

Olkin, I., Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, *54*(1), 317–322.

Orr, M., Liu, P. (2009). Sample size estimation while controlling false discovery rate for microarray experiments using the ssize. fdr package. *R Journal*, *1*, 47–53.

Pidsley, R., Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, *14*(1), 293.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909.

Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.-E., Greco, D., … Kere, J. (2012). Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS ONE*, *7*(7), 41361.

Resnick, H. S., Kilpatrick, D. G., Dansky, B. S., Saunders, B. E., & Best, C. L. (1993). Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. *Journal of Consulting and Clinical Psychology*, *61*(6), 984.

Ressler, K. J., Mercer, K. B., Bradley, B., Jovanovic, T., Mahan, A., Kerley, K., … May, V. (2011). Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. *Nature*, *470*(7335), 492–497.

Rice, L., Waters, C. E., Eccles, J., Garside, H., Sommer, P., Kay, P., … Stratford, I. (2008). Identification and functional analysis of SKA2 interaction with the glucocorticoid receptor. *Journal of Endocrinology*, *198*(3), 499–509.

Roth, T. L., Lubin, F. D., Funk, A. J., & Sweatt, J. D. (2009). Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biological Psychiatry*, *65*(9), 760–769.

Sareen, J., Houlahan, T., Cox, B. J., & Asmundson, G. J. (2005). Anxiety disorders associated with suicidal ideation and suicide attempts in the National Comorbidity Survey. *The Journal of Nervous and Mental Disease*, *193*(7), 450–454.

Sartor, C. E., Grant, J. D., Lynskey, M. T., McCutcheon, V. V., Waldron, M., Statham, D. J., … Nelson, E. C. (2012). Common heritable contributions to low-risk trauma, high-risk trauma, posttraumatic stress disorder, and major depression. *Archives of General Psychiatry*, *69*(3), 293–299.

Sartor, C. E., McCutcheon, V. V., Pommer, N. E., Nelson, E. C., Grant, J. D., Duncan, A. E., … Heath, A. C. (2011). Common genetic and environmental contributions to post-traumatic stress disorder and alcohol dependence in young women. *Psychological Medicine*, *41*(7), 1497–1505.

Sipahi, L., Wildman, D. E., Aiello, A. E., Koenen, K. C., Galea, S., Abbas, A., & Uddin, M. (2014). Longitudinal epigenetic variation of DNA methyltransferase genes is associated with vulnerability to post-traumatic stress disorder. *Psychological Medicine*, *44*(15), 3165–3179.

Smith, A. K., Conneely, K. N., Kilaru, V., Mercer, K. B., Weiss, T. E., Bradley, B., … Ressler, K. J. (2011). Differential immune system DNA methylation and cytokine regulation in post-traumatic stress disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *156*(6), 700–708.

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., … Conneely, K. N. (2014). Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, *15*(1), 145.

Smyth, G. K. (2005). Limma: Linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, (pp. 397–420).

Stein, M. B., Jang, K. L., Taylor, S., Vernon, P. A., & Livesley, W. J. (2002). Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: A twin study. *American Journal of Psychiatry*, *159*(10), 1675–1681.

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics*, *29*(2), 189–196.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., … Alzheimer's Disease Neuroimaging Initiative ECICSYSG(2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, *8*(2), 153–182.

True, W. R., Rice, J., Eisen, S. A., Heath, A. C., Goldberg, J., Lyons, M. J., & Nowak, J. (1993). A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. *Archives of General Psychiatry*, *50*(4), 257–264.

Tyrka, A. R., Price, L. H., Marsit, C., Walters, O. C., & Carpenter, L. L. (2012). Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: Preliminary findings in healthy adults. *PLoS ONE*, *7*(1), 30148.

Uddin, M., Aiello, A. E., Wildman, D. E., Koenen, K. C., Pawelec, G., de Los Santos, R., … Galea, S. (2010). Epigenetic and immune function profiles associated with posttraumatic stress disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9470–9475.

Vinkers, C. H., Kalafateli, A. L., Rutten, B. P., Kas, M. J., Kaminsky, Z., Turner, J. D., & Boks, M. P. (2015). Traumatic stress and human DNA methylation: A critical review. *Epigenomics*, *7*(4), 593–608.

Weaver, I. C. G., Cervoni, N., D'Alessio, A. C., Champagne, F. A., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Epigenetic programming through maternal behavior. *Nature Neuroscience*, *7*, 847–854.

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., … Mühleisen, T. W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, *15*(2), 1.

Wisco, B. E., Marx, B. P., Wolf, E. J., Miller, M. W., Southwick, S. M., & Pietrzak, R. H. (2014). Posttraumatic stress disorder in the US veteran population: Results from the National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, *75*(12), 1338–1346.

Xian, H., Chantarujikapong, S. I., Scherrer, J. F., Eisen, S. A., Lyons, M. J., Goldberg, J., … True, W. R. (2000). Genetic and environmental influences on posttraumatic stress disorder, alcohol and drug dependence in twin pairs. *Drug and Alcohol Dependence*, *61*(1), 95–102.

Xie, P., Kranzler, H. R., Yang, C., Zhao, H., Farrer, L. A., & Gelernter, J. (2013). Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biological Psychiatry*, *74*(9), 656–663.

Zeileis A. (2004). Econometric computing with HC and HAC covariance matrix estimators.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ratanatharathorn A, Boks MP, Maihofer AX, et al. Epigenome-wide association of PTSD from heterogeneous cohorts with a common multi-site analysis pipeline. *Am J Med Genet Part B*. 2017;174B:619–630. https://doi.org/10.1002/ajmg.b.32568