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

Andrew Ratanatharathorn, Columbia University
Marco P Boks, University Medical Center Utrecht
Adam X Maihofer, University of California
Allison E Aiello, University of North Carolina
Ananda B Amstadter, Virginia Commonwealth University
Allison Ashley-Koch, Duke University
Dewleen G Baker, University of California
Jean C Beckham Beckham, VA Mid-Atlantic Healthcare Network
Evelyn Bromet, State University of New York
Michelle Dennis, VA Mid-Atlantic Healthcare Network

Only first 10 authors above; see publication for full author list.

Journal Title: American Journal of Medical Genetics Part B: Neuropsychiatric Genetics
Volume: Volume 174, Number 6
Publisher: Wiley: 12 months | 2017-09-01, Pages 619-630
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1002/ajmg.b.32568
Permanent URL: https://pid.emory.edu/ark:/25593/td9pb

Final published version: http://dx.doi.org/10.1002/ajmg.b.32568

Copyright information:
© 2017 Wiley Periodicals, Inc.

Accessed February 13, 2020 10:17 AM EST
Epigenome-wide Association of PTSD from Heterogeneous Cohorts with a Common Multi-Site Analysis Pipeline


1Columbia University, Department of Epidemiology
2University Medical Center Utrecht, Brain Center Rudolf Magnus
3University of California San Diego, Department of Psychiatry
4UNC Chapel Hill, Department of Epidemiology
5Virginia Commonwealth University, Department of Psychiatry
6Duke Molecular Physiology Institute
7Veterans Affairs San Diego Healthcare System
8Veterans Affairs Center of Excellence for Stress and Mental Health
9VA Mid-Atlantic, Mental Illness Research, Education, and Clinical Center
10Durham VA Medical Center
11Duke University Medical Center, Department of Psychiatry and Behavioral Sciences
12State University of New York, Epidemiology Research Group
13Military Mental Healthcare- Research Centre, Ministry of Defence
14McLean Hospital, Neurobiology of Fear Laboratory
15Harvard T.H. Chan School of Public Health, Department of Epidemiology and Massachusetts General Hospital, Department of Psychiatry

Corresponding Author, Alicia K. Smith, Ph.D., Emory University School of Medicine, 101 Woodruff Circle NE, Ste 4217, Atlanta, GA 30322, 404-712-5006 (phone), alicia.smith@emory.edu.

Conflict 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.
Abstract

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 Posttraumatic 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; trauma; stress

**Introduction**

Traumatic events are reported by over 70% of individuals during their lifetime [Benjet and others 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 [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 and others 2000; Najt and others 2011]. PTSD has also been associated with a 2.8-fold increase in suicidal thoughts and behaviors [Sareen and others 2005] as well as with a number of chronic medical conditions [Boscarino 2008; Coughlin 2011; David and others 2014; Heppner and others 2009; Jakovljevic and others 2008; Kubzansky and others 2009; Kubzansky and others 2007; Luft and others 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 and others 1998; Kessler and others 2005; Resnick and others 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 30–70%, [Sartor and others 2012; Sartor and others 2011; Stein and others 2002; True and others 1993; Xian and others 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 and others 2014; Guffanti and others 2013; Logue and others 2013; Xie and others 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 and others 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 and others 2010; Maunakea and others 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 and Skinner 2007; Weaver and others 2004]. Similar alterations have been reported in the human literature, in both central and peripheral tissues [McGowan and others 2009; Tyrka and others 2012]. Considering the influence of traumatic stress on DNA methylation seen in some studies [Vinkers and others 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 and others 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 and others 2011] and a potentially interacting effect of genetic and epigenetic variation at the dopamine receptor (SLC6A3) on PTSD risk [Chang and others 2012]. Indeed, stress exposure itself has been shown to alter epigenetic patterns in both animal and human studies [Moser and others 2015; Roth and others 2009; Sipahi and others 2014]. In addition to candidate gene methylation studies, a small number of studies have examined genome-wide DNA methylation patterns in PTSD [Mehta and others 2013; Smith and others 2011; Uddin and others 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 and others 2013; Smith and others 2011; Uddin and others 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 and others 2013] and blood cell composition [Houseman and others 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 and others 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 and Zeng 2010; Mathew and Nordstrom 1999; Olkin and 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 versus 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 and others 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 and others 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.
Materials and Methods

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.

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 and others 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.

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 1), 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 and others 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 and others 2014]. Even if an associated CpG site is influenced by genetics, such as is the case for SKA2 [Boks and others 2016; Rice and others 2008], maintaining such probes is informative to our overall goal of identifying genes important for PTSD [Gibbs and others 2010; Guintivano and others 2014; Heyn and others 2013]. Normalization of probe distribution and background differences between Type I and Type II probes was conducted using Beta Mixture Quantile Normalization (BMIQ) [Teschendorff and others 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 2) [Pidsley and others 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 and others 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 and others 2014; Fortin and others 2014]. 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.

**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 colleagues [Jaffe and Irizarry 2014; Reinius and others 2012], based on the Houseman algorithm [Houseman and others 2012]. CD8, CD4, NK, B cell, and monocyte cell proportions were included as covariates in our statistical analyses.

**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 and others 2014; Nielsen and others 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. for use in European and African American subjects [Barfield and others 2014]. 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 and others 2015]. Ancestry using GWAS data was inferred as described in Nievergelt et al. [Nievergelt and others 2013]. In brief, genotypes of 1783 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 and others 2003; Li and others 2008]. Based on these ancestry estimates, subjects were placed into one of 4 groups: European Americans, African Americans, Latinos/Native Americans and ‘Others’. PCs were derived using Eigenstrat [Price and others 2006].

Ancestry estimates using methylation data were derived using subsets of methylation probes in close proximity to SNPs identified by Barfield et al. [Barfield and others 2014]. Probe sets with 0bp distance (N=7,703 CpG probes), within 1bp distance (17,995 CpG probes), and within 10bp 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).

**Statistical Analysis**

Within each cohort, logit transformed β values (M-values) [Du and others 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 and colleagues 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 and Ervin 2000; Zeileis 2004].

Cohort-level analysis results were combined using the inverse normal method [Marot and others 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 percent [Benjamini and Hochberg 1995].

**Sensitivity analysis**

Numerous robust associations between age and DNAm have been reported [Bocklandt and others 2011; Hannum and others 2013; Horvath 2013; Horvath and others 2012; Weidner
and others 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.

**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 and others 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 and 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 σ=0.43, which represents the highest variation reported in one of our cohorts, and thus a very conservative estimate [Orr and Liu 2009].

**Results**

**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.

**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).

**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. [Barfield and others 2014]. A comparison of CpG probes with SNPs within 1bp distance (N= 17,995) and CpG probes with SNPs within 10bp distance (N= 50,319) showed a higher genotype-methylation correlation for the 1bp probes (r=0.29, p=1.8 × 10^-15) than the 10bp probes (r=0.06, p=0.0015). Figure 3 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 1bp from
the probes (panel B), and the PCs from CpG probes within 10bp of a SNP (panel C). Supplemental Figure 5 shows PCs 2, 3, 4, and PCs 2, 3 and 6 for, respectively, CpGs within 1bp and 10bp 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 1bp of a SNP provide reliable results and are suitable as ancestry covariates in our analyses.

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 3). Figure 2 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: (i) our proposed PGC-EWAS pipeline, described in detail in the Methods section above, as well as (ii) 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 1). 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.

Sensitivity analysis: Age Meta-Analysis Results

Results for the age analysis using our pipeline and no standard error adjustment are presented in Table 3. All studies reported numerous 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 3). 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 2). 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 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 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 and others 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 neuroimaging efforts [Logue and others 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 and others 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 and others 2010; Kundaje and others 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 and others 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.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**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 (R01BX002577). 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 R01MB093500 funded the GWAS assays and analysis. Acknowledged are Victoria B. Rishborough 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 was supported in part by National Institute of Mental Health Award RO1MB079806, Department of Veterans Affairs, Clinical Science Research & Development Program Award 510CX005431-02, Department of Veterans Affairs, Biomedical Laboratory Research & Development Program Award 1101BX002150-01, 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 was supported under an agreement between the Department of Veterans Affairs and the Pharmacogenomics Research Network. This work was also supported, in part, by a Merit Award (#R01BX002577) 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. Deredt, 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-Munford, 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. Tupper, 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.

**References**


Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2018 September 01.


Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC genomics. 2014; 15(1):145. [PubMed: 24555763]


Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. 2004
Figure 1.
Sample size vs. power 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).
Figure 3.
PTSD genomic inflation factors (λ) by quality control pipeline (PGC vs. Funnorm) and standard error adjustment method (empirical Bayes vs. HC3).
### Table 1

PGC-PTSD EWAS Cohorts

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

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

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