

June 2021

## Analysis of Post-traumatic Stress Disorder Gene Expression Profiles in a Prospective, Community-based Cohort

Jan Dahrendorff  
*University of South Florida*

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>

 Part of the [Genetics Commons](#)

---

### Scholar Commons Citation

Dahrendorff, Jan, "Analysis of Post-traumatic Stress Disorder Gene Expression Profiles in a Prospective, Community-based Cohort" (2021). *Graduate Theses and Dissertations*.  
<https://digitalcommons.usf.edu/etd/9095>

This Thesis is brought to you for free and open access by the Graduate School at Digital Commons @ University of South Florida. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Analysis of Post-traumatic Stress Disorder Gene Expression Profiles in a Prospective,  
Community-based Cohort

by

Jan Dahrendorff

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science in Public Health  
with a concentration in Genomics  
College of Public Health  
University of South Florida

Major Professor: Monica Uddin, PhD.  
Derek Wildman, PhD.  
Agaz Wani, PhD

Date of Approval  
June 25th, 2021

Keywords: PTSD, gene expression, RNA-sequencing, transcriptome

Copyright © 2021, Jan Dahrendorff

## **Acknowledgements**

I would like to express my sincerest appreciation to Monica Uddin, Ph.D., for her ongoing support and guidance throughout this project. It has been a great pleasure and honor to work under Dr. Uddin on this project and I am grateful for the countless opportunities Dr. Uddin has provided to me in order to continue my academic curiosity. My deepest gratitude to my committee member Agaz Wani, Ph.D., for his patience and constant support throughout this project. Without his guidance and his vast computational expertise, the completion of this project would not have been possible. Additionally, I would like to acknowledge my committee member Derek Wildman, Ph.D., for his expertise, directions, and support during this project.

Furthermore, I would like to thank my fellow lab members Charley Wang, Ph.D., Sarah Burgan, Zachary Graham and Janelle Donglasan for their professional support during this project. Lastly I would like to acknowledge Thomas Keller, Ph.D along with the other members of the genomics hub who have always been a reliable resource to seek help guidance or motivation.

## Table of Contents

List of Tables .....	ii
List of Figures .....	iii
Abstract .....	iv
Introduction .....	1
Post-traumatic Stress Disorder: Etiology, Epidemiology and Risk Factors .....	1
RNA-Sequencing studies of PTSD .....	4
Lack of Prospective, Longitudinal Transcriptomic Assessments of PTSD .....	8
Objective .....	8
Materials and Methods .....	10
Study Participants .....	10
Assessment of PTSD and Demographics .....	11
RNA-Sequencing .....	12
Filtering and Removing Low Counts .....	12
Blood Cell-Type Deconvolution .....	13
Transformation of Cell Proportions .....	15
Differential Gene Expression Analysis .....	15
Specifying the Form Model for Dream Analysis .....	16
Gene Set Enrichment Analysis (GSEA) .....	18
Validation of Differentially Expressed Signatures in the Human Brain .....	18
Results .....	20
Demographic and PTSD-related Variables in the Study Sample .....	20
Outcome of RNA processing, gene mapping and gene assignment .....	23
Estimation of the Blood Cell Type Proportions .....	23
Differential Gene Expression .....	26
Gene Set Enrichment Analysis .....	31
Discussion .....	33
References .....	36

## **List of Tables**

Table 1: Participant Characteristics of PTSD and Control Samples.....	20
Table 2: Downregulated Genes According to Lifetime PTSD Development.....	29
Table 3: Upregulated Genes According to Lifetime PTSD Development.....	30

## List of Figures

Figure 1: Breakdown of Number of Participants with Lifetime PTSD .....	21
Figure 2: Demographic breakdown of study participants.....	22
Figure 3: Cell Proportions at the Baseline Measure .....	<b>Error! Bookmark not defined.</b>
Figure 4: Cell Fractions of the baseline Measure displayed in a Violin Plot.....	25
Figure 5: Cell Proportions of the Follow-up Measure .....	25
Figure 6: Cell Fractions of the Follow-up Measure displayed in a Violin Plot.....	26
Figure 7: Differentially Expressed Genes According to Lifetime PTSD Development.....	28
Figure 8: Top Gene Ontology Terms Associated with PTSD Development.....	31
Figure 9: Dotplot of Gene Ontology Terms in Relation to PTSD Development.....	32

## Abstract

Post-traumatic stress disorder (PTSD) is a common and debilitating psychiatric disorder that may occur in individuals exposed to traumatic events such as accidents, interpersonal violence, war, combat-exposures or natural disasters. PTSD is a significant public health issue with a high disease burden associated with substantial health care costs and several comorbidities negatively affecting an individual's quality of life. The biological underpinnings of the disorder are not well understood. Gene expression studies can shed light into the complex physiology of PTSD. However, to date, studies employing a hypothesis-free approach examining the whole transcriptome are scarce and are limited to assessment of only one timepoint. Here we applied a transcriptome-wide gene expression screen with RNA-sequencing to whole blood samples in order to elucidate the gene expression signatures associated with the development of PTSD. The study participants (n=72, of which 21 eventually developed PTSD) are a subsample of participants enrolled in a longitudinal and prospective cohort study of adults living in Detroit, Michigan called Detroit Neighborhood Health Study. PTSD was assessed in a structured telephone interview and whole blood samples were taken both before and after trauma exposure. A linear mixed model was used in order to analyze the gene expression data optimized for repeated measures. We found 45 differentially expressed genes in our cohort of individuals who developed PTSD with an estimated  $\log_2$  fold change  $> 1.5$  at a nominal p value ( $p < 0.05$ ). Seven of our upregulated genes including *PAX6*, *TSPAN7*, *PXDN*, *VWC2*, *SULF1*, *PAX3* and *NFATC4* were also expressed in the brain, while four of our down regulated genes including *UGT2A3*, *ACTP1*, *HSPB7* and *FMOD* were also found to be expressed in the brain. Gene set enrichment

analysis of the differentially expressed genes implicated several pathways relating to brain functions and immune functioning to be enriched in individuals developing PTSD. No differentially expressed genes were observed after multiple test correction. However, the longitudinal sampling of participants prior to the development of PTSD provides a promising mean to gain insight into the pathophysiology underlying PTSD development in future studies.



## **Introduction**

### **Post-traumatic Stress Disorder: Etiology, Epidemiology and Risk Factors**

Post-traumatic stress disorder (PTSD) is a pervasive and debilitating mental disorder with a lifetime prevalence of approximately 6.8-7.3 % in the United States general population (Roberts et al., 2011; Kessler & Wang, 2008), while an increased rate on the order of 25-30% has been reported in segments of the population exposed to persistent psychological stress including combat exposed veterans, victims of assault or refugees (National Institute of Mental Health, 2017; Roberts et al., 2011; Kessler & Wang, 2008; McLaughlin et al., 2015, Fulton et al., 2015). The disorder manifests itself through a range of symptoms including the recurrent and intrusive recollection of the traumatic experience through hallucinations and nightmares, changes in arousal and emotional reactions, alterations in mood and cognition, avoidance of stimuli reminding of the trauma and impairment of functioning in social and occupational contexts (American Psychiatric Association, 2013). PTSD can have a significant impact on an individual's quality of life caused by the condition itself or other medical comorbidities associated with the disorder including chronic inflammation, substance abuse psychological disorders and suicide (Kessler et al., 2005).

A range of risk factors for PTSD have been previously identified including the level of previous stress exposures, low socio-economic status, extent of social support and female sex (Yehuda & LeDoux, 2007; Brewin et al., 2000; Lowe et al., 2014). Trauma exposure is the etiological component of PTSD and exposure to traumatic events is ubiquitous, although it is not deterministic as not everyone experiencing a trauma develops PTSD (Resnick et al., 1993;

Kessler et al. 1995, Mills et al., 2011; Kessler et al.,2017). In fact, only a small proportion of individuals exposed to traumatic events develop PTSD while the majority remains resilient sometimes even after repeated exposure (Zohar et al., 2009). Recent assessments of the World health organization's world mental health surveys spanning over 24 countries with 68,894 respondents suggest 70.4 % have experienced at least one lifetime trauma with the most frequently experienced types of trauma in the United States including interpersonal violence and sexual violence in women, while accidents or combat exposure are the most prevalent exposures in men (Benjet et al., 2016). On a global scale, natural disasters, wars and other humanitarian emergencies account for the majority of traumatic experiences (Kessler et al., 2017).

The significant individual differences in response to trauma and subsequent PTSD might be explained by the heterogeneity at the molecular level that can impact someone's response to trauma. Consequently, identifying the molecular factors influencing psychiatric risk and resilience factors in PTSD has emerged as priority in research of the disorder. An enhanced understanding of the complex mechanisms underlying a maladaptive trauma response and the development of PTSD can help to mitigate the vast morbidity and burden associated with the disorder and can also inform efforts of treatment development, interventions and preventive strategies.

Over the last decade significant progress has been made in characterizing the etiologic factors of PTSD. A large body of literature spanning across twin studies, targeted candidate gene analysis, genome and epigenome-wide association studies, as well as methylation and gene expression studies have implicated the role of genetics and epigenetics in the vulnerability to the development of PTSD (Logue et al., 2013; Mehta et al., 2013; Sarapas et al., 2011; Xie et al., 2013). Several putative risk factors and risk loci associated with PTSD have been identified,

however, discrete diagnostic biomarkers underlying PTSD etiology and the pathways leading to the disorder remain to be elucidated. Emerging literature acknowledges altered gene expression as an important genomic factor in the susceptibility to PTSD (Mehta et al., 2020). The assessment of gene expression can provide valuable insights into the complex pathophysiology of psychiatric disorders like PTSD.

Initial reports by Segman and colleagues reported differences in parts of the transcriptome of patients surviving trauma with PTSD, implicating transcriptional enhancers and immune activation genes to be dysregulated in peripheral blood (Segman et al., 2005). Previous research investigating the mechanisms underlying PTSD have converged on the functioning of the immune system and hypothalamic–pituitary–adrenal (HPA) axis regulation (Daskalakis et al., 2016; Logue et al., 2015). Several of these studies focused on the examination of glucocorticoid signaling and pro-inflammatory cytokines in individuals with PTSD. Genes playing a role in the regulation of the glucocorticoid receptor (GR) such as *FKBP5* (Binder et al., 2008; Menke et al., 2013) have been particularly well documented. The GR plays a key role in the regulation of the HPA axis and the stress response. Further, the GR is involved with regulating the immune system and immune responses which can be triggered by stressors, (Slavich et al., 2014; Song et al., 1999) therefore playing an important role in etiology of stress-related psychopathologies. In individuals with PTSD, alterations in glucocorticoid signaling have been identified as a potential vulnerability factor for the development of PTSD (Raison et al., 2003; van Zuiden et al., 2013). Additionally, a large body of literature has documented the association of PTSD with altered immune system functioning, with elevated levels of pro- and anti-inflammatory cytokines and C-reactive protein among the most well-documented (Passos et al., 2015; Michopolous et al., 2017; Rosen et al., 2017). Accordingly, several genes involved in the regulation of GR functioning

such as *FKBP5* are also involved with the immune response and have been identified to be differentially expressed in PTSD afflicted individuals (Sarapas et al., 2011; Yehuda et al., 2009; Mehta et al., 2011). Collectively, past literature has converged on the differential expression of genes involved with the immune and glucocorticoid responses induced by stress in individuals with PTSD. The majority of these studies however, applied a gene expression microarray approach (Zieker et al., 2007; Yehuda et al., 2009; Neylan et al., 2011; Sarapas et al., 2011; Mehta et al., 2011; Pace et al., 2012; Glatt et al., 2013, Mehta et al., 2013; Logue et al., 2015; Guardado et al., 2016), limiting the identification of novel transcripts and pathways involved in PTSD pathophysiology and making it difficult to elucidate mechanistic connections between the genes and system level insights.

Previous single time-point transcriptomic wide screens have implicated alterations in glucocorticoid signaling and immune regulation pathways in conferring risk for PTSD in individuals who have been exposed to trauma (Segman et al., 2005; Yehuda et al., 2009; Sarapas et al., 2011; Mehta et al., 2011; Bam et al., 2016). However, genome-wide differential expression analysis of longitudinal cohorts capturing two or more timepoints might be more appropriate in order to elucidate resiliency signatures and expression factors conferring (or mitigating) psychiatric risk in the development of PTSD (Breen et al., 2015).

### **RNA-Sequencing studies of PTSD**

More recently, transcriptomic-wide screens have been able to extend findings from prior studies adopting a microarray approach. The application of transcriptome-wide RNA-sequencing platforms enables a comprehensive and hypothesis-free interrogation of expression signatures without having to rely on a priori knowledge of gene targets previously identified as risk factors (Kukurba & Montgomery, 2015). To date transcriptome-wide gene expression assessments of

PTSD utilizing RNA-Seq are scarce and inconclusive (Bam et al., 2016; Breen et al., 2015; Kuan et al., 2017; Boscarino et al., 2019; Pattinson et al., 2020) with the majority of implicated biomarker candidates having yet to be replicated in further studies.

Bam et al combined RNA sequencing and microRNA microarray approaches in their study of peripheral blood samples from Veterans with PTSD and controls (N=10) (Bam et al., 2016). Their analysis found 326 genes and 190 MicroRNAs to be significantly differentially expressed among the PTSD group (Bam et al., 2016). A subsequent functional enrichment analysis revealed several dysregulated immune system pathways in participants with PTSD. Additionally, the authors observed alteration to the cell differentiation pathway characterized by the differential expression of *STAT4* and *TBX21*, with both of those genes having crucial roles in the regulation of T cells.

Kuan and colleagues conducted a comprehensive gene expression study analyzing RNA sequencing data derived from whole blood in World trade Center responders comprising of individuals with current and past PTSD as well as controls who never had PTSD in their lifetime (Kuan et al., 2017). After randomly dividing samples of 324 responders into a discovery (N= 195) and a replication (N= 87) cohort the authors identified 448 differentially expressed genes in their discovery cohort of individuals with current PTSD with 99 of the genes remaining significant in the replication cohort at FDR < 0.05. In their analysis the genes *NDUFA1*, *CCDC85B*, *SNORD54*, *FKBP5* and *SNORD46* showed the absolute fold change > 1.2 in individuals currently suffering from PTSD. While *NDUFA1*, *CCDC85B*, *SNORD46* and *SNORD54* were downregulated, *FKBP5* was up regulated in individuals with current PTSD. Subsequent pathway analysis identified the top five pathways were located in the glucocorticoid receptor signaling pathway, pathways involved with macrophages, endothelial cells, fibroblast,

the action cytoskeleton signaling pathway, granzyme A signaling pathway as well as the NGF signaling pathway. The construction of a polygenic risk score from 30 differentially expressed genes showed differences in risk scores shared between current and Past PTSD in comparison to the controls. Interestingly, Current and Past PTSD displayed comparable polygenic expression scores. Their findings also confirmed candidate genes *FKBP5*, *CASP2*, *SOD1*, *BBC3*, and *C9orf84* found to be differentially expressed at  $FDR < 0.05$  as in prior studies of PTSD (Logue et al., 2015). Overall, the findings from this study support the findings from early work implicating pathways involved with immune system and glucocorticoid functioning in PTSD (Kuan et al., 2017).

Similarly, Boscarino et al conducted a differential gene expression analysis of peripheral blood from 85 Canadian male military personnel previously deployed to Afghanistan using RNA-sequencing (Boscarino et al., 2019). Their transcriptomic analysis of 27 participants currently presenting with symptoms of PTSD and 58 negative controls revealed higher expression of *LRP8* (or *ApoE*) and *GOLM1* in individuals with PTSD while *CYP2C8* was found to be significantly downregulated in that group compared to the controls. *LRP8* is highly expressed in the hippocampus and the amygdala, involved with the activation of synaptic plasticity genes and therefore important for learning and memory functioning as well as fear conditioning, a core pathway implicated in PTSD pathogenesis (Li et al., 2016). *LRP8* has also been reported to play a role in the maintenance of cholesterol homeostasis which is important in synapse formation and development. The isoforms E2 and E4 of *LRP8* have been implicated in an increased risk for cardiovascular disease and neurodegenerative conditions like Alzheimer's disease suggesting a link between PTSD and the risk for the development of cardiovascular disease as well as neurological disorders including Alzheimer's disease (Weiner et al., 2014).

A recent RNA-sequencing study conducted by Pattinson and colleagues examined the gene expression profiles in peripheral blood samples of male military personnel with PTSD and excessive daytime sleepiness (EDS) along with their respective controls (Pattinson et al., 2020). Their study included 46 active-service military personnel or veterans with PTSD and 61 PTSD-absent controls. The participants were assessed for EDS and sorted into three groups according to the severity of presented symptoms and presence of excessive daytime sleepiness. In the first group consisting of participants with PTSD not having excessive daytime sleepiness only two genes, *RAP1 GTPase Activating Protein* and *TBC1 Domain Family Member 3E* were significantly differentially expressed in comparison to their non-PTSD counterparts at an adjusted p-value. In the group of participants with PTSD as well as symptoms of excessive daytime sleepiness, over 251 genes were found to be differentially expressed compared to the controls. A subsequent candidate network analysis identified 10 significant networks having significant dysregulation, most notably those involved with organismal injury and abnormalities and a network centered on cellular function and maintenance. Genes in those networks were similarly associated with the immune response, ubiquitination, ATP modulation and regulation of the P13K-Akt pathway. In the third group comparing participants with PTSD and EDS to those with PTSD and without EDS had a differential gene expression of 1,873 genes. Of note, the study did not control for the covariates age, other demographic variables collected in the study as well as cell-type proportions. The networks identified by IPA Network analysis for this group focused on the networks mostly associated with PTSD and EDS. After selecting the Carbohydrate Metabolism network and the RNA Posttranscriptional Modification many differentially expressed genes identified were associated with neurodegeneration, ubiquitination, mitochondria functioning, glucocorticoid regulation, sleep problems like insomnia and circadian

regulation (Pattinson et al., 2020). Interestingly the limited dysregulation of only two genes between the PTSD non-EDS and PTSD-Absent group contrasts findings from an earlier study the reported significant gene dysregulation accompanies the presence of PTSD (Kuan et al., 2017).

### **Lack of Prospective, Longitudinal Transcriptomic Assessments of PTSD**

Another common theme of the majority of the aforementioned studies is that they applied a single-timepoint evaluation on individuals already diagnosed with PTSD, therefore lacking a prospective design, with only one notable exception. Breen and colleagues investigated the transcriptome wide gene expression RNA-seq data in peripheral blood of 94 U.S Marines pre- and post-deployment constructing co-expression networks associated with PTSD (Breen et al., 2015). Their analysis identified modules with an overexpression of genes that were enriched for processes involved with innate immunity and interferon signaling to be associated with PTSD development (Breen et al., 2015). To our knowledge, longitudinal designs like this have yet to be applied to a community-based cohort. Application of these longitudinal study designs would be especially informative, as it is unclear whether gene expression patterns are reflective of current symptoms or are innately posing as an enduring susceptibility prior to trauma exposure. The assessment of gene expression patterns longitudinally is therefore important for the characterization of the molecular underpinnings of PTSD and the discovery of biomarkers with clinical utility.

### **Objective**

This study aims to interrogate the gene expression profiles of individuals with PTSD in a community dwelling cohort. The analysis is based on exploring the transcriptomic profiles in



whole blood of PTSD diagnoses based on two different variables focused on lifetime PTSD and also a PTS symptom severity construct (explained in the methods below) potentially capable of a more nuanced distinction between cases and controls. Of particular interest is the investigation of differential gene expression signatures associated with PTSD development. The application of a linear mixed model will determine the gene expression changes of the participants from baseline to the follow up measure, flexibly modeling the repeated measure gene expression data while increasing power and decreasing false positives.

## **Materials and Methods**

### **Study Participants**

The study participants were enrolled in a five yearlong study called the Detroit Neighborhood Health Study (DNHS). The DNHS is a prospective and longitudinal cohort study of adults residing in Detroit, Michigan (Uddin et al., 2010). The initial aim of the study was to evaluate to which extent ecologic stressors including concentrated disadvantage, distribution of income and residential segregation contribute to PTSD risk. Data collection for the DNHS began in 2008 with each wave of measurement including a phone interview over 40 minutes. The participants were compensated for their participation with \$25 after each wave. Biospecimen collection was compensated with an additional \$25. Data collection for the baseline wave occurred 2009-2010 with approximately annual follow ups until 2014. At the beginning of each interview and when a biological specimen was collected informed consent was obtained. The study protocol has been reviewed and approved by the Institutional Review Board of the University of Michigan.

A subset of DNHS participants consented to provide a biospecimen. The biospecimens were acquired from the participants during a visit in their homes by a phlebotomist. The samples were subsequently shipped and processed at Wayne State University in Michigan (Uddin et al., 2010). From this biospecimen subset 72 participants, out of which 21 developed PTSD, were identified along with 144 available whole blood samples taken both before and after trauma exposure (e.g from wave2 and wave 4 respectively). Total RNA was isolated from the leukocytes

of the blood samples with Leukolock kits (Ambion, Austin, TX), as previously described (Logue et al., 2013).

### **Assessment of PTSD and Demographics**

Lifetime PTSD among the participants was assessed using PTSD Checklist (PCL-C) via a structured telephone interview, as previously described (Weathers & Ford, 1996; Uddin et al., 2010). The PCL-C is a 17- item measurement of PTSD symptoms that uses the PTSD symptom criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychological Association [APA], 1994). The telephone interview also included a range of additional questions regarding the duration, time and impairment mediated by the symptoms to identify cases of PTSD according to DSM-IV criteria. Additionally, a list of specific traumatic events was presented from which the participants were asked which of them were experienced in the past (Breslau et al., 1998). If all six DSM-IV criteria were met by participants, they were considered to be affected by lifetime PTSD. In addition to the assessment of PTSD, demographic variables including race, sex and age were collected.

### **PTS Symptom Severity Measure**

A PTS symptom severity measure was also explored for differential gene expression analysis as a measure of traumatic stress potentially delivering more nuanced expression signatures regarding PTSD development. Using the PTS symptom severity might offer more statistical power and individuals who suffer from post-traumatic stress but do not meet all the diagnostic criteria for PTSD due to different numbers of symptoms they present across the different symptom clusters, are here included. An increase of statistical power is achieved due to

the fact that PTS symptom severity is a continuous variable while Lifetime PTSD is a dichotomous variable based on the fulfillment of all PTSD diagnostic criteria.

## **RNA-Sequencing**

RNA-sequencing (RNA-seq) was performed on all samples (n=144). The ribosomal RNAs were removed with Ribozero Human/Mouse rat from Illumina. The RNAseq libraries were prepared with TruSeq stranded total RNAsequencing Sample Preparation kit by Illumina. The libraries were quantified by qPCR and subsequently sequenced on 8 lanes for 101 cycles from each end of the fragments on a HiSeq 4000 sequencing system. Fastq data was generated, and the resulting data was quality controlled (see details below). After alignment of the reads the quantification of expression on the gene level was facilitated with the use of the Featurecounts function of the R-package Subreads (Liao, Smyth & Shi , 2019).

## **Filtering and Removing Low Counts**

The first step includes the standard RNA-seq processing in which the reads from the counts table are read in. The filtering by the number of counts and the removal of low read counts prior to downstream analysis is important from a biological as well as a statistical perspective. Insufficient evidence for differential expression is provided by genes with significantly low gene counts across all libraries. From a biological standpoint a gene must be expressed at a minimal threshold in order for it to be translated into a protein or to be of biological significance. From a statistical point of view, genes with only very low counts are unlikely to be considered differentially expressed due to the fact that insufficient statistical evidence for differential gene expression is given. The inclusion of such genes would also interfere with subsequent multiple testing when estimating false discovery rate, ultimately

reducing power to detect genes that are differentially expressed. The removal of such genes will not result in a loss of information and can therefore be done (Chen, Lun & Smyth, 2016).

For our data we have selected to filter on the basis of counts-per-million (CPM) retaining genes if they are expressed at CPM above 0.5 in our samples. Retaining genes expressed at this particular threshold has been recommended in previous literature of PTSD and the authors of the edgeR program (Kuan et al., 2017; Robinson, McCarthy & Smyth, 2010).

The subsequent object `geneExpr = DGEList( countMatrix[isexpr,] )` is then subset in order to only retain the genes filtered by the chosen threshold. The `calcNormFactors` function of the open-source Bioconductor project package edgeR is then, after the filtering step, used to calculate the normalization factors used to scale between the libraries (Robinson, McCarthy & Smyth, 2010).

### **Blood Cell-Type Deconvolution**

The cell-type-deconvolution process refers to the separation of heterogenous mixture signals into its constituent cell components by a range of experimental and computational approaches. Differences in cell type proportions and composition as well as their contributions to gene expression profiles is an important determinant of phenotypic variation capable of influencing disease states (Sharma et al., 2019). The identification of discrete cell populations is also important in analysis of differential gene expression as it can be substantially confounded by differences in the cell proportions in each sample.

Towards this end, the deconvolution of various blood cell types was performed with Cibersortx, a software applying a computational framework based on support vector machine

regression to infer different cell types and cell type specific expression profiles in bulk RNA samples without relying on physical cell isolation (Newman et al., 2019). In order to enumerate the fractions of various cell populations in the RNA-sequencing data with Cibersortx two input files were required. The first required input is termed “signature matrix” a knowledgebase representing marker genes enabling the discrimination between different cell types and subsets. Further, a tab-delimited text file called “mixture file” was utilized consisting of the gene expression profile counts of the samples.

The reference signature matrix “Non-small cell lung cancer PBMCs” is based on single-cell RNA-seq analysis derived from peripheral blood mononuclear cells previously isolated from tissues of patients suffering from non-small cell lung cancer generated on the 10x Genomics Chromium v2 assay (Chen et al., 2020). The matrix consists of profiles of six leukocyte subsets including B Cells, CD4T & CD8T Cells, monocytes, natural killer cells and natural killer T cells. This reference signature matrix has been previously used in another RNA-seq study investigating multisystem inflammatory syndrome in children (Beckmann et al.,2020).

## **Transformation of Cell Proportions**

Given a computational approach was used to deconvolute the cell fractions of the five cell types, the issue that our five cell type fractions add up to 1.0 for the samples has the implication that the cell fractions are only spanning four dimensions precluding knowing how four fractions determines the fifth (Hoffman et al., 2021).

This issue was addressed by transforming the five cell fractions into four variables with the use of the isometric log ratio of the R package compositions (Van den Boogaart & Tolosana-Delgado, 2013) and creating covariates Cellfraction\_1, Cellfraction\_2, Cellfraction\_3, cellFrac\_ilr\_4. Transforming into the cell fractions has the additional advantage of invariance to scaling and the changing the order of the variables.

## **Differential Gene Expression Analysis**

In order to gain insights into the pathophysiology associated with PTSD development, two differential gene expression analyses were performed with the differential expression for repeated measures (dream) function available in the variancePartition R package (Hoffman and Schadt, 2020). Dream employs a linear model capable of increasing statistical power while minimizing false positives by joining multiple statistical concepts into a single statistical model implemented by its software (Hoffman and Roussos, 2020). The dream analysis is built onto the widely applied workflow for differential expression analysis of limma/voom (Law et al. 2014).

The model includes several statistical concepts into its workflow including the modeling of repeated measures, precision weights capable of modeling measurement errors in the RNA-

sequencing counts and the ability to include several random effects. Apart from the ability to model a range of random effects, there is a separate estimation of random effects for each single gene.

### **Specifying the Form Model for Dream Analysis**

Next, the form was specified in which the variable to be tested must be a fixed effect. For our analysis several metadata variables were included and assessed for their contribution to variation in differential expression. Those variables include disease status “PTSD”, Sex, age, time point (wave) as well as a range of cell proportions commonly considered in transcriptomic studies of whole blood samples from individuals with mental disorders including B cells, CD4T cells, CD8T cells, monocytes and Neutral killer cells.

The first analysis was focused on the development of PTSD between the wave 2 and wave 4 measure.

In the first analysis, gene expression between cases vs controls was tested. Subsequently it was tested if the PTSD coefficient is different from zero in order to investigate if there is a PTSD effect after the removal of variation across wave and accounting for the repeated measurement of donors including a random effect for the participant.

For this first step the following regression formula was used:

$$\sim \text{ptsd} + (1|\text{resp}) + \text{wave} + \text{Sex} + \text{Age} + \text{Cellfraction\_1} + \text{Cellfraction\_2} + \text{Cellfraction\_3} + \text{Cellfraction\_4}$$

In which “ptsd” is the lifetime PTSD status at the end of wave 4, resp is the identifier for the individual participant, wave is the timepoint (w2= baseline, wave4= follow up) and Cellfraction\_1 - Cellfraction\_4 are the log transformed cell proportion estimates.



Additionally, cases vs controls were tested without the inclusion of the wave with the following formula:

```
~ ptsd + (1|resp) + Sex + Age + Cellfraction_1 + Cellfraction_2 + Cellfraction_3 +  
Cellfraction_4
```

Furthermore, it was explored if there is a wave-by-disease interaction using formula:

```
~ ptsd + wave + ptsd:wave + (1|resp) + Sex + Age + Cellfraction_1 + Cellfraction_2 +  
Cellfraction_3 + Cellfraction_4
```

Subsequently, the coefficient PTSD:wave was tested for the association between PTSD and gene expression changes between the two waves by examining whether it was significantly different from 0 for each gene.

The second analysis included the PCL score measuring PTS symptom severity instead of using the lifetime PTSD measure with the following formula:

```
form <- ~ (1|resp) + Age + Sex + Cellfraction_1 + Cellfraction_2 + Cellfraction_3 +  
Cellfraction_4 + PCL_score_combined
```

Differential expression in our data considered was evaluated at a nominally significant p-value < 0.05 with a fold change of 1.5 or greater and also at an adjusted p-value < 0.05 using the Benjamini-Hochberg false discovery rate estimation (Hochberg and Benjamini, 1990).

## **Gene Set Enrichment Analysis (GSEA).**

In order to elucidate the biological processes underlying and accompanying PTSD development a gene set enrichment analysis (GSEA) was performed on the identified differentially expressed genes. A GSEA can facilitate the identification of classes of genes found to be over-represented in a large gene set that can be subsequently associated with a disease. The R package ‘clusterProfiler’ is capable of automating the process of biological-term classification and the enrichment analysis of gene clusters. We have applied the commonly used Gene Ontology (GO) (Ashburner et al., 2000) that provides annotations of genes to a range of biological processes, molecular functions, and different cell components. Differentially expressed signatures at a nominally significant p-value  $< 0.05$  were tested for enrichment.

## **Validation of Differentially Expressed Signatures in the Human Brain**

Many complex diseases underly the dysfunction of multiple tissues and organs. This dysfunction is often characterized by tissue specific transcriptional changes mediating causal links between the genotype and disease (Cookson et al., 2009). Assessing tissue specific gene expression is an important consideration given that the assessment of psychiatric disorders can be challenging because the organ central to the disease-the brain is not available to be assessed in living individuals. Investigating if the differentially expressed genes identified in blood samples are also expressed in the brain can help evaluate how reflective blood transcriptomic profiles are of the disorders of the brain. To this end, we verified the expression of the in blood found differentially expressed genes if they are also expressed in the human brain by conducting a

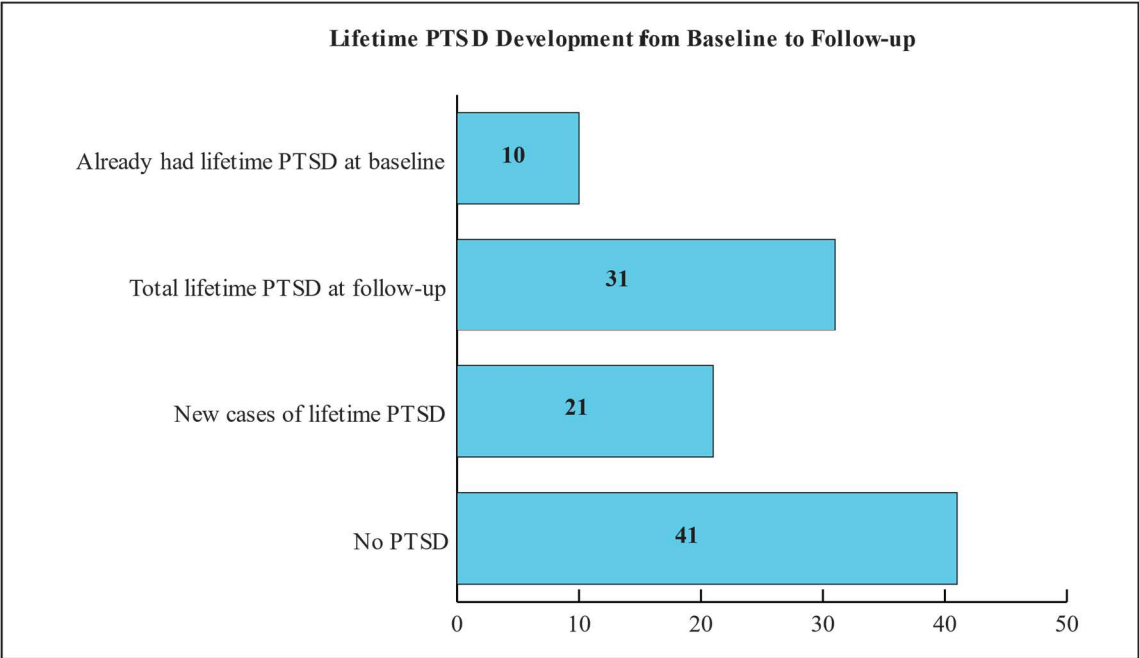
search on the publicly available Human Protein Atlas (HPA) database(<http://www.proteinatlas.org/>).

## Results

### Demographic and PTSD-related Variables in the Study Sample

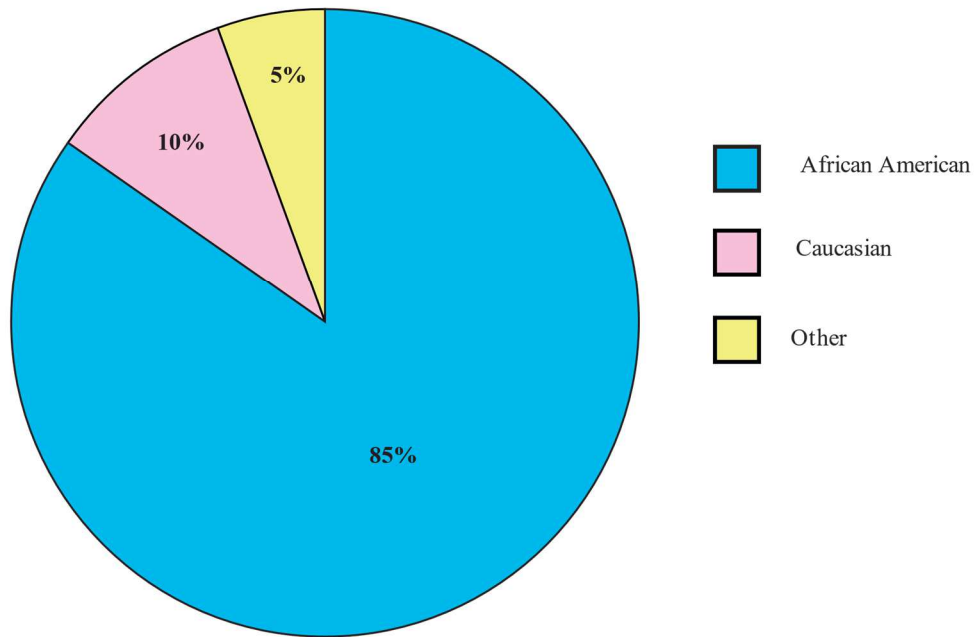
**Table 1.** Participant Characteristics of PTSD and Control Samples

<i>Characteristic</i>	<i>All participants (N = 72)</i>		<i>Lifetime PTSD at Baseline (N = 10)</i>		<i>Lifetime PTSD at follow-up (N = 31)</i>		<i>New Case of lifetime PTSD (N = 21)</i>		<i>No lifetime PTSD by follow-up (N = 41)</i>		<i>PTS symptom score at Baseline** (Min-Max:6.0-72.0)</i>	<i>PTS symptom score by follow-up** (Min-Max:17.0-73.0)</i>
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>		
<i>Age*</i>	56.66	12.78	56.98	12.80	56.49	12.94	54.38	13.91	56.97	12.10		
<i>Female</i>	39	54	6	60	20	64	14	67	19	46	31.27	33.25
<i>Male</i>	33	46	4	40	11	36	7	33	22	54	32.46	33.83
<i>African American</i>	61	85	9	90	24	77	15	71	37	90	32.46	33.83
<i>Caucasian</i>	7	10	1	10	5	16	4	19	2	5	31.16	31.13
<i>Other</i>	4	5	0	0	2	7	2	10	2	5	30.66	32.21



**Figure 1.** Breakdown of Number of Participants with Lifetime PTSD

**Race/Ethnicity Distribution**



**Figure 2.** Demographic breakdown of study participants

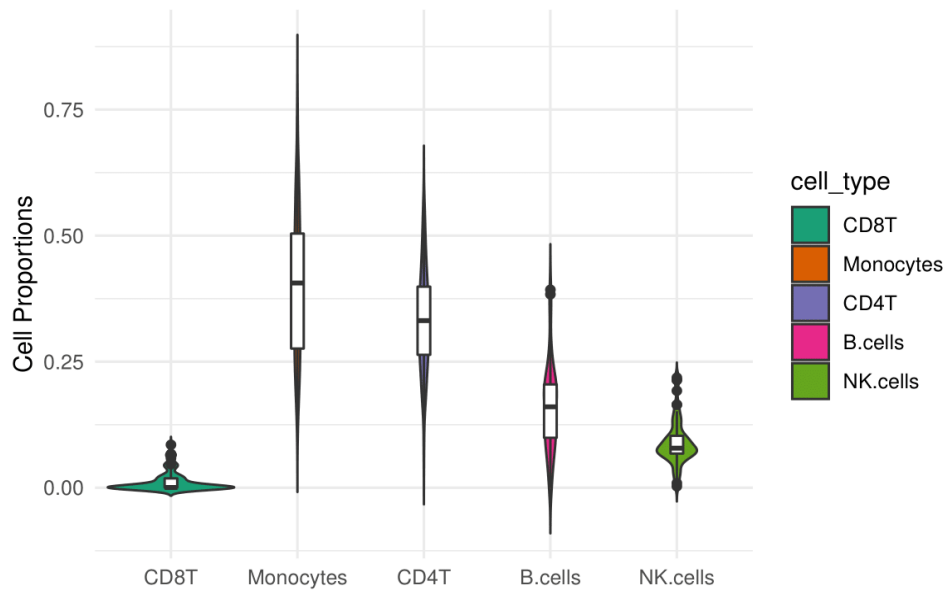
## **Outcome of RNA processing, gene mapping and gene assignment**

From the original RNAseq experiment, a total of 320,197,33 reads were obtained mapping to 58,884 transcripts. 75% of obtained reads successfully mapped to the GRCh38 reference genome

([ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode\\_human/release\\_24/GRCh38.primary\\_assembly.genome.fa.gz](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_24/GRCh38.primary_assembly.genome.fa.gz)). 87% of reads were assigned to a total of 26,659 individual genes after mapping with the use of the FeatureCounts function of the R-package Subreads (Liao, Smyth & Shi , 2019).

## **Estimation of the Blood Cell Type Proportions**

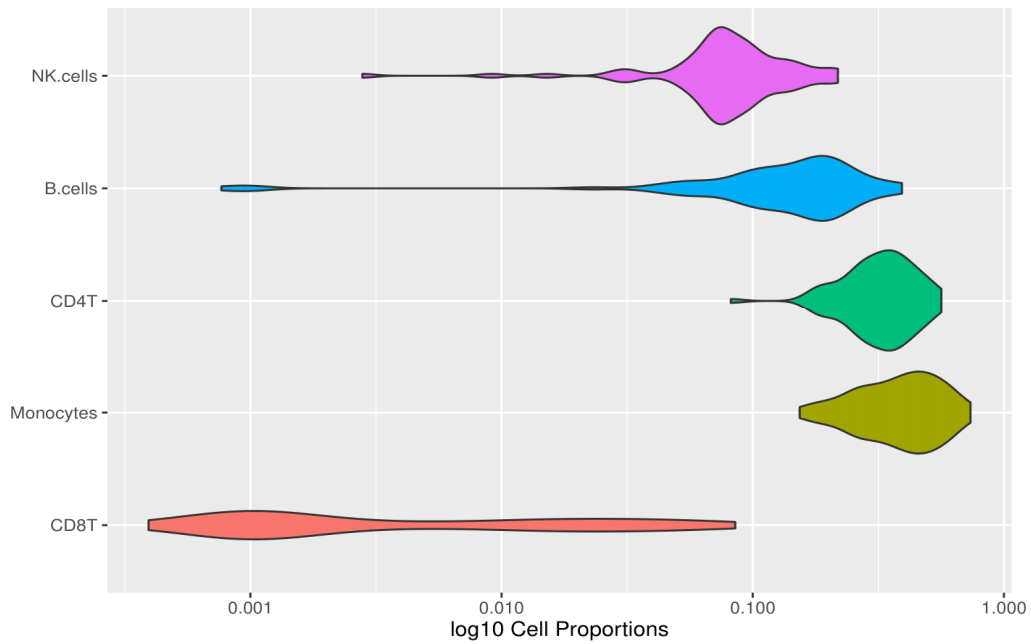
Using quality-controlled data, Cibersortx was used to determine the cell-type composition of the heterogenous blood samples including the baseline and the follow-up transcriptomic measures. For both timepoints the quantified proportions included B cells, CD4T, CD8T, Monocytes and NK cells for each participant. Natural killer T cell estimates were provided in the Cibersortx output but not included in the exploration of covariates among the cell proportions because they were only represented in less than 4% of the samples.



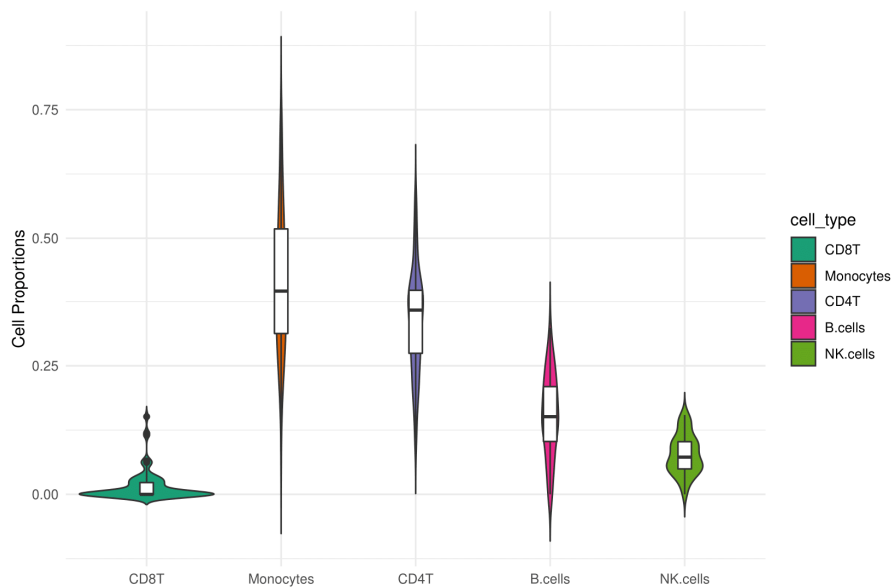
**Figure 3.** Cell Proportions at the Baseline Measure

Cell proportions of the 72 samples included in the study at the baseline measure including Cell fractions of B cells, CD4T, CD8T, Monocytes, NK cells.



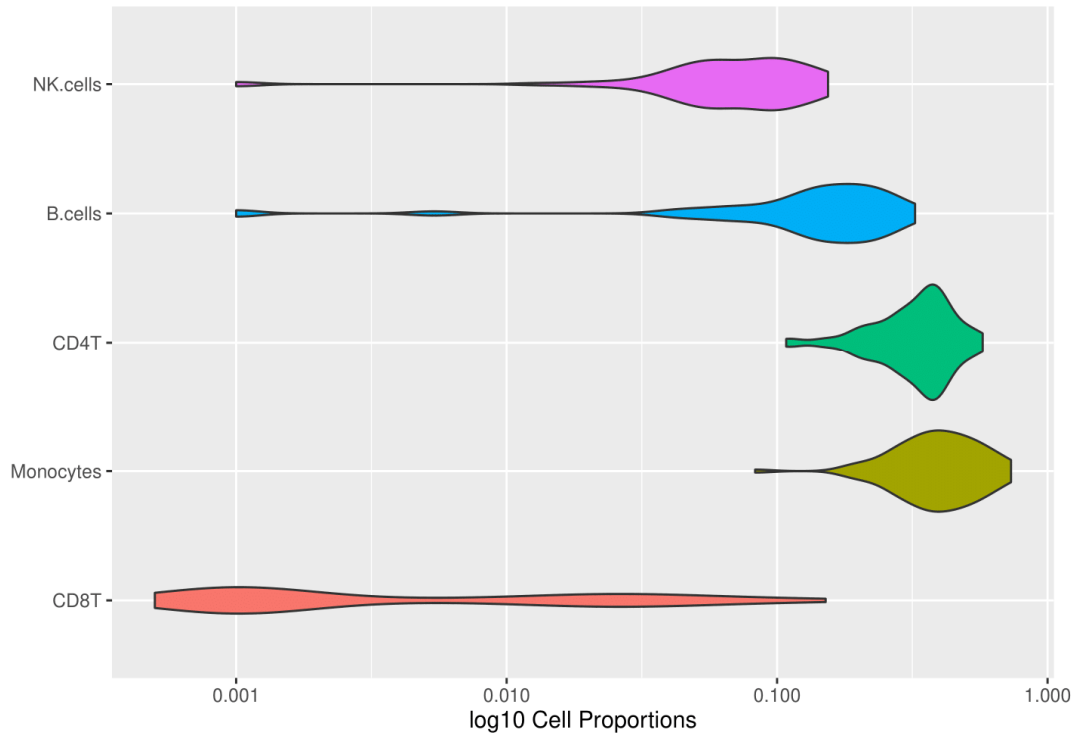


**Figure 4.** Cell Fractions of the baseline Measure displayed in a Violin Plot



**Figure 5.** Cell Proportions of the Follow-up Measure

Cell proportions of the 72 samples included in the study at the follow-up measure including Cell fractions of B cells, CD4T, CD8T, Monocytes, NK cells.



**Figure 6.** Cell Fractions of the Follow-up Measure displayed in a Violin Plot

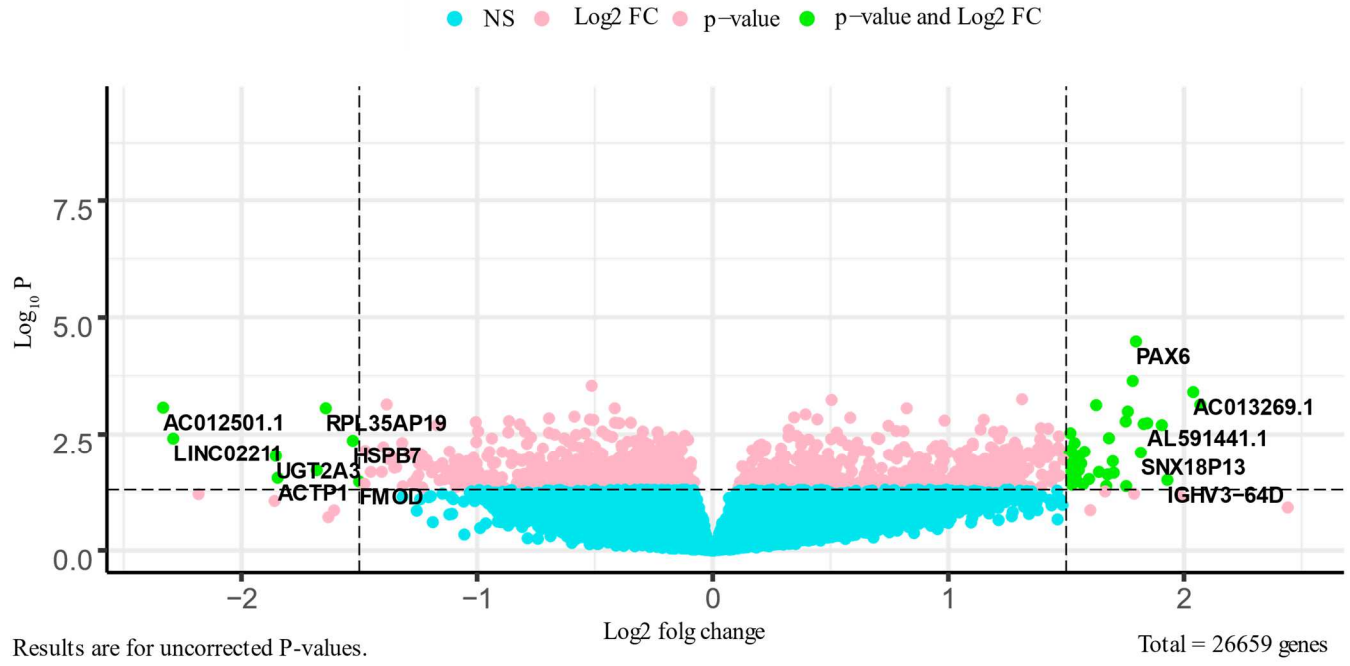
### Differential Gene Expression

To gain insights into the gene expression signatures associated with PTSD development in our community-based cohort, a differential gene expression analysis was performed testing multiple coefficients of interest. Retention of power and the controlling of the false positivity rate was ensured by accounting for the repeated measures of our PTSD cases and controls with the dream function of the variancePartition R package (Hoffman et al., 2020), which enables specification of random effect(s). In the dream analysis the effects of the individual on the gene expression are modeled as a random effect for each individual gene with the use of a linear mixed model. Two different DE analyses were performed. For the models, multiple test

correction was performed taking the number of genes tested into an account with the use of the Benjamini and Hochberg method (Hochberg and Benjamini, 1990).

In our first two analyses the gene expression between cases vs controls was tested with and without the inclusion of the wave covariate. Subsequently it was tested if the PTSD coefficient is different from zero in order to investigate if there is a PTSD effect. Neither of these analyses suggested any differentially expressed genes with a  $\log_2$ fold change greater than 1.5 at an unadjusted p-value  $< 0.05$ . Similarly the model, testing for gene expression that included the PCL score measuring PTS symptom severity instead of using the lifetime PTSD, measure did not suggest any differentially expressed genes even at the nominal p-value  $< 0.05$ . In the analysis testing for a wave-by-disease interaction and subsequently testing the coefficient “PTSD:wave” for the association between PTSD and gene expression changes between the two waves there were several differentially expressed genes with a  $\log_2$ fold change greater than 1.5 at an unadjusted p-value  $< 0.05$ . The volcano plot (Figure 7) shows the global gene expression patterns according to lifetime PTSD development, indicating a larger number of upregulated genes compared to a few downregulated genes associated with PTSD development. In total, 45 genes have been identified to be differentially expressed at a nominal p-value  $< 0.05$ , however, none of the differentially expressed genes survived multiple test correction at an adjusted p-value  $< 0.05$ .

Top Down and Upregulated Genes According to the Onset of Lifetime PTSD (PTSD:Wave Coefficient)



**Figure 7.** Volcano Plot displaying Differentially Expressed Genes According to Lifetime PTSD Development

Eight genes including *AC012501.1*, *LINC02211*, *UGT2A3*, *ACTP1*, *LINC02228*, *RPL35AP19*, *HSPB7* and *FMOD* were identified as downregulated according to lifetime PTSD development (Table 2).

**Table 2.** Downregulated Genes According to Lifetime PTSD Development

*Downregulated at Nominal p-value < 0.05 in PTSD:wave (log fold change 1.5)*

<i>Ensemblid</i>	Gene	LogFC	AveExpr	t	P.Value	Adj.P.Val	Z.std
<i>ENSG00000227400</i>	<i>AC012501.1</i>	-2.33236	-2.16611	-3.41908	0.000845	0.952575	-3.33756
<i>ENSG00000245662</i>	<i>LINC02211</i>	-2.28962	-3.46875	-2.94785	0.00389	0.952575	-2.88695
<i>ENSG00000135220</i>	<i>UGT2A3</i>	-1.85419	-4.07665	-2.64266	0.009397	0.952575	-2.59725
<i>ENSG00000229890</i>	<i>ACTP1</i>	-1.84626	-4.4253	-2.22633	0.027948	0.952575	-2.19802
<i>ENSG00000251273</i>	<i>LINC02228</i>	-1.68067	-4.02054	-2.37971	0.019097	0.952575	-2.34364
<i>ENSG00000239872</i>	<i>RPL35AP19</i>	-1.64212	-4.78866	-3.41661	0.000874	0.952575	-3.32825
<i>ENSG00000173641</i>	<i>HSPB7</i>	-1.52756	-4.55159	-2.90137	0.004345	0.952575	-2.85199
<i>ENSG00000122176</i>	<i>FMOD</i>	-1.50034	-4.45216	-2.15001	0.033614	0.952575	-2.12468

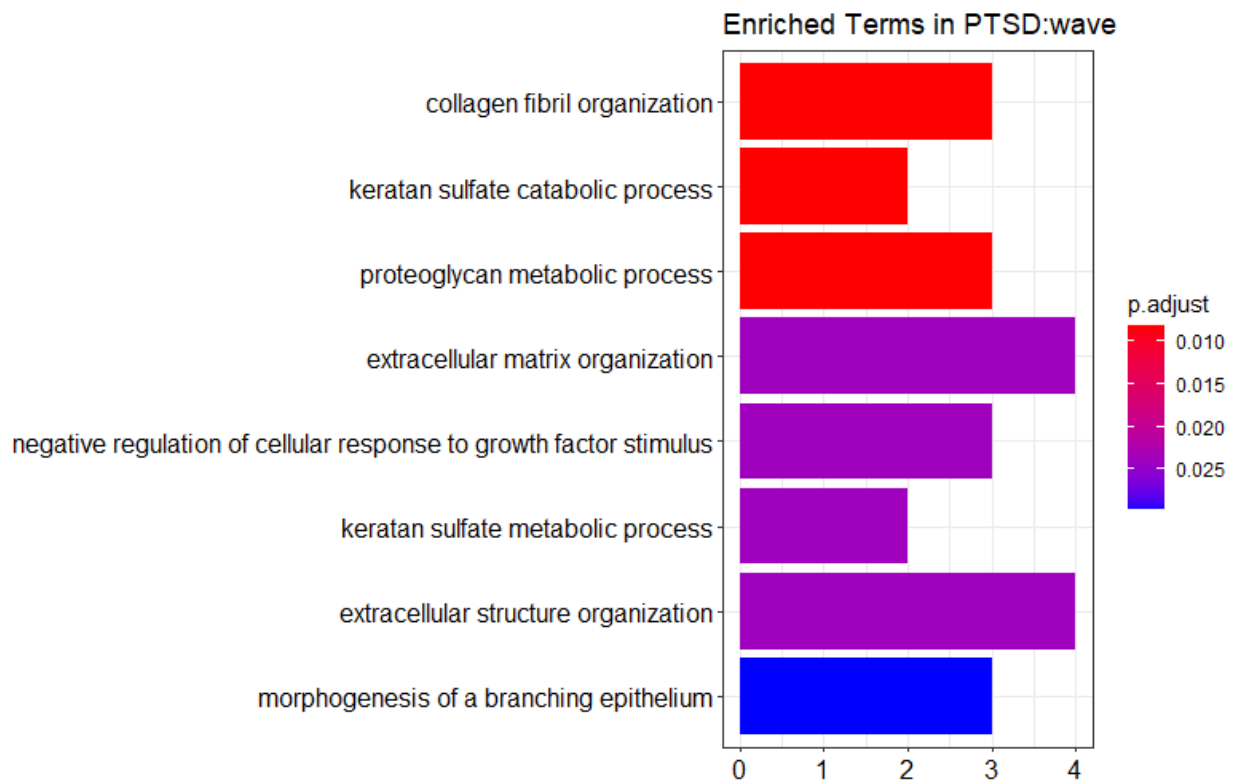
Furthermore, there were a total of 37 upregulated genes associated with lifetime PTSD development at a nominal p-value < 0.05 and log<sub>2</sub> (fold change) above 1.5 (Table 3). Among the top differentially expressed genes with the highest fold change were *AL136317.2*, *AC013269.1*, *IGHV3-64D*, *AF130359.1*, *AL591441.1*, *TSPAN7*, *SNX18P13*, *PAX6*, *AC008543.5*, *Z69706.1* and *PXDN*.

**Table 3** Upregulated Genes According to Lifetime PTSD Development*Upregulated at Nominal p-value < 0.05 in PTSD:wave (log fold change 1.5)*

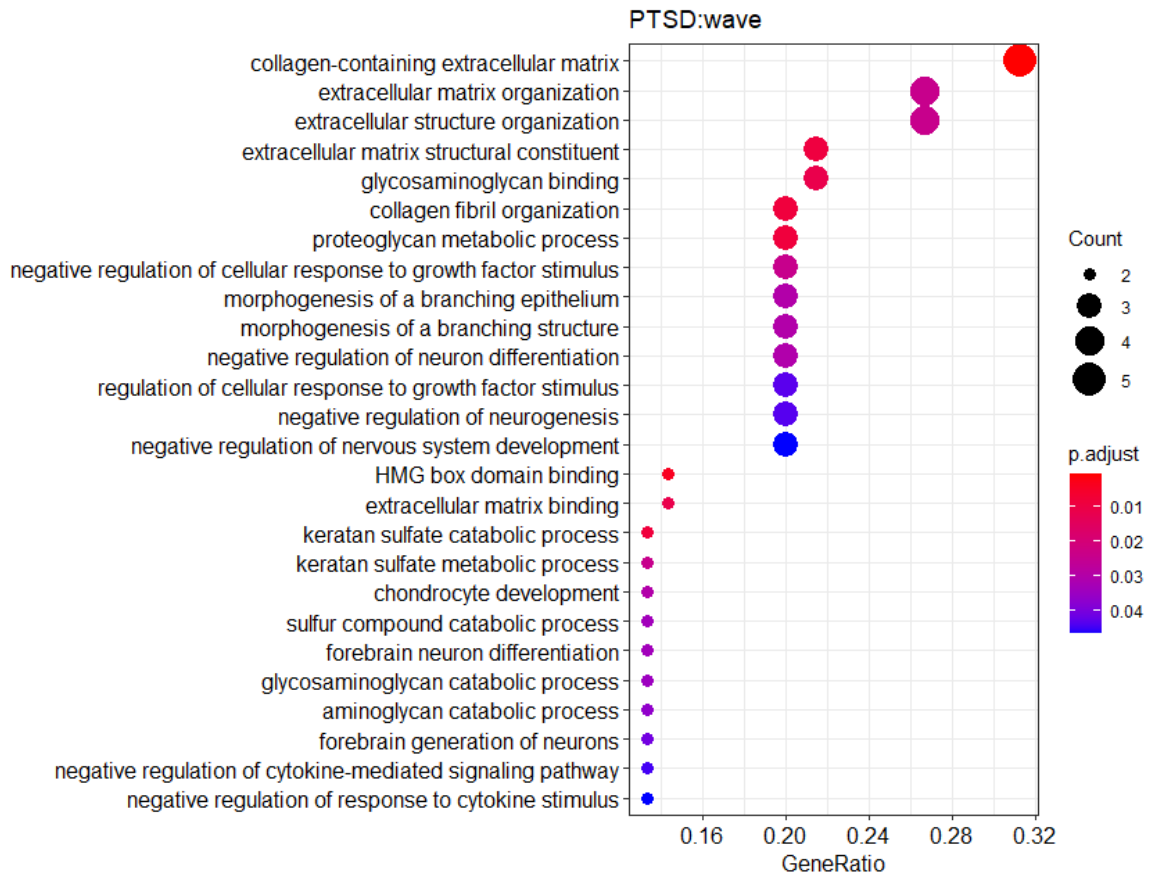
<i>Ensemblid</i>	<i>Gene</i>	<i>LogFC</i>	<i>AveExpr</i>	<i>t</i>	<i>P.Value</i>	<i>Adj.P.Val</i>	<i>Z.std</i>
<i>ENSG00000274893</i>	<i>AL136317.2</i>	2.069513	-3.80093	3.476319	0.00073	0.952575	3.37795
<i>ENSG00000225341</i>	<i>AC013269.1</i>	2.039292	-4.14782	3.665617	0.000393	0.952575	3.54454
<i>ENSG00000282639</i>	<i>IGHV3-64D</i>	1.929707	-3.0084	2.187479	0.031035	0.952575	2.156621
<i>ENSG00000237735</i>	<i>AF130359.1</i>	1.906208	-4.58473	3.175578	0.002008	0.952575	3.089076
<i>ENSG00000276128</i>	<i>AL591441.1</i>	1.844134	-4.40663	3.198495	0.00182	0.952575	3.118174
<i>ENSG00000156298</i>	<i>TSPAN7</i>	1.82917	-3.79308	3.170037	0.001889	0.952575	3.107136
<i>ENSG00000230965</i>	<i>SNX18P13</i>	1.816912	-2.16646	2.708481	0.007875	0.952575	2.657394
<i>ENSG00000007372</i>	<i>PAX6</i>	1.796315	-3.00469	4.301555	3.24E-05	0.877348	4.155555
<i>ENSG00000267646</i>	<i>AC008543.5</i>	1.782542	-3.37977	3.818061	0.000227	0.952575	3.686855
<i>ENSG00000268836</i>	<i>Z69706.1</i>	1.761723	-3.34508	3.357925	0.001023	0.952575	3.284242
<i>ENSG00000130508</i>	<i>PXDN</i>	1.754969	-1.95644	2.052409	0.042077	0.952575	2.032763
<i>ENSG00000285079</i>	<i>AL513493.1</i>	1.752864	-3.2731	3.208111	0.001672	0.952575	3.143115
<i>ENSG00000184761</i>	<i>PRSS40B</i>	1.702262	-3.44557	2.328637	0.021705	0.952575	2.295487
<i>ENSG00000187135</i>	<i>VSTM2B</i>	1.698038	-3.62639	2.540431	0.012212	0.952575	2.505954
<i>ENSG00000229847</i>	<i>EMX2OS</i>	1.681767	-1.7885	2.961896	0.003813	0.952575	2.893211
<i>ENSG00000185736</i>	<i>ADARB2</i>	1.676024	0.873709	2.288869	0.023651	0.952575	2.262747
<i>ENSG00000261348</i>	<i>LINC02136</i>	1.670426	-4.42673	2.06117	0.04176	0.952575	2.035902
<i>ENSG00000170153</i>	<i>RNF150</i>	1.639384	-1.0087	2.346386	0.020885	0.952575	2.310054
<i>ENSG00000226676</i>	<i>BX248123.1</i>	1.627546	-3.69832	3.487307	0.000746	0.952575	3.372131
<i>ENSG00000235726</i>	<i>AC010148.1</i>	1.597349	-1.52826	2.203848	0.029654	0.952575	2.174677
<i>ENSG00000286099</i>	<i>1.592427</i>	-2.53084	2.166485	0.032043	0.952575	2.143873	
<i>ENSG00000188730</i>	<i>VWC2</i>	1.578	-4.36536	2.713234	0.007538	0.952575	2.672072
<i>ENSG00000145147</i>	<i>SLIT2</i>	1.571715	-0.69046	2.099633	0.03817	0.952575	2.07303
<i>ENSG00000228314</i>	<i>CYP4F29P</i>	1.563679	0.328895	2.503854	0.013714	0.952575	2.464657
<i>ENSG00000233048</i>	<i>LINC01722</i>	1.555737	-3.63935	2.646673	0.009102	0.952575	2.608181
<i>ENSG00000137573</i>	<i>SULF1</i>	1.553857	-4.42013	2.379621	0.019074	0.952575	2.344088
<i>ENSG00000228971</i>	<i>AL356479.1</i>	1.540334	-4.04196	2.219536	0.028132	0.952575	2.195446
<i>ENSG00000228791</i>	<i>THRB-AS1</i>	1.53563	-3.00011	2.878841	0.004867	0.952575	2.815727
<i>ENSG00000225258</i>	<i>AC009478.1</i>	1.529383	-4.83787	2.666998	0.008854	0.952575	2.617629
<i>ENSG00000232006</i>	<i>AC005537.1</i>	1.525101	-2.33867	2.087177	0.03932	0.952575	2.060819
<i>ENSG00000276484</i>	<i>AC126755.5</i>	1.524739	-3.30051	2.088198	0.038673	0.952575	2.067652
<i>ENSG00000227766</i>	<i>HCG4P5</i>	1.522675	-5.30751	2.497207	0.013944	0.952575	2.45871
<i>ENSG00000060718</i>	<i>COL11A1</i>	1.521923	-4.74859	2.361928	0.020029	0.952575	2.325796
<i>ENSG00000135903</i>	<i>PAX3</i>	1.521438	-4.30189	2.17104	0.032099	0.952575	2.143173
<i>ENSG00000100968</i>	<i>NFATC4</i>	1.519876	-3.85836	2.744275	0.006896	0.952575	2.701825
<i>ENSG00000213614</i>	<i>HEXA</i>	1.517953	-4.52251	3.040019	0.002968	0.952575	2.97099
<i>ENSG00000168122</i>	<i>ZNF355P</i>	1.515112	-4.04418	2.693571	0.008234	0.952575	2.642316

## Gene Set Enrichment Analysis

In order to identify the biological processes associated with PTSD development, a gene set enrichment analysis was performed. Investigating the pathway enrichment for the 37 upregulated and eight downregulated differential expression signatures at the nominal significant  $p$ -value  $< 0.05$  we found some gene sets to be significantly enriched. The top gene ontology terms associated with PTSD development in regard to the PTSD:wave measure (Fig. 8) were found to be involved in cellular organization and functioning.



**Figure 8.** Top Gene Ontology Terms Associated with PTSD Development



**Figure 9.** Dotplot of Gene Ontology Terms in Relation to PTSD Development



## Discussion

Despite the contributions of several RNA-sequencing studies of PTSD, longitudinal analysis that characterizes the gene expression changes associated with the development of the disorder are lacking. In order to address this gap in knowledge, we analyzed RNA-seq data in whole blood from 72 community dwelling individuals from the Detroit Neighborhood Health Study. Among them, 21 developed PTSD from the baseline to the follow-up time point. A linear mixed model optimized for repeated measures was applied in order to analyze the gene expression data. We found 45 differentially expressed genes in our cohort of individuals who developed PTSD with an estimated  $\log_2$ fold change  $> 1.5$  at a nominal p value; however, none of these survived correction for multiple hypothesis testing.

Several of our differentially expressed genes identified in blood are also expressed in the brain (Human Protein Atlas (HPA) database (<http://www.proteinatlas.org/>)) With regard to the upregulated genes, *PAX6*, *TSPAN7*, *PXDN*, *VWC2*, *SULF1*, *PAX3* and *NFATC4* were also expressed in the brain. Out of the downregulated genes, *UGT2A3*, *ACTP1*, *HSPB7* and *FMOD* were also found to be expressed in the brain. *PAX6* is an important transcriptional factor for normal brain development and is involved in the regulation of the expression of several other genes coordinating cortical development (Osumi et al., 2008; Ypsilanti & Rubenstein, 2016). *TSPAN7* has been found to differentially expressed in a previous microarray study investigating the gene expression profiles of military personnel who sustained closed-head injuries from explosions during deployment (Heinzelmann et al., 2014). *TSPAN7* has also been associated with cognitive defects mediated by downstream alterations to the *TM4SF2* gene (Penzes et al., 2013).

*VWC2* has been previously found to be differentially expressed (downregulated) in individuals with PTSD and mild cognitive impairment in a targeted study of World Trade Center responders (Kuan et al., 2020). *NFATC4* is among a family of five distinct *nuclear factor of activated T cells* genes regulating the transcriptional induction of several genes involved with immune activators and modulators including a range of adaptive immunity and pro-inflammatory cytokines (Kipanyula, Kimaro & Etet, 2016). Dysregulation of other *NFATC* loci has previously been implicated in PTSD in peripheral tissues (Breen et al., 2019). Collectively, our results confirm the dysregulation of genes and the enrichment of pathways involved with immune functioning implicated in other PTSD studies.

Gene set enrichment analysis of the differentially expressed genes implicated several pathways to be enriched in individuals developing PTSD. Among the top enriched pathways were multiple pathways were involved with brain functions such as forebrain neuron differentiation and the regulation of neurogenesis. Additionally, two of the top enriched pathways were involved with the negative regulation of cytokine-mediated signaling and the negative regulation of the response to cytokine stimulus, immune related pathways that have been implicated in PTSD in previous studies (Breen et al., 2015; Bam et al., 2016; Kuan et al., 2019). PTSD has been consistently linked to increased inflammation in the body, which is in line with the findings of some of our immune-related enriched pathways (Michopolous et al., 2017; Rosen et al., 2017).

No differentially expressed genes were observed after multiple test correction. This outcome is not surprising considering the efforts to prevent false positives by accounting for several covariates and our small sample size. Although no statistically significant results were achieved in our cohort, the longitudinal sampling of participants prior to developing PTSD can

be important in identifying expression signatures underlying PTSD development in future studies. A strength of our study is the assessment of a community-based cohort as opposed to accessing PTSD in veterans, active duty military or first responders. Additionally, this work might be of use in larger meta-analysis efforts that leverage larger sample sizes in order to identify transcriptomic signatures underlying or accompanying the development of PTSD. Further, the implicated genes associated with PTSD development could be validated in future genome wide association studies.

## References

- Bam, M., Yang, X., Zumbun, E. E., Zhong, Y., Zhou, J., Ginsberg, J. P., . . . Nagarkatti, M. (2016). Dysregulated immune system networks in war veterans with PTSD is an outcome of altered miRNA expression and DNA methylation. *Sci Rep*, 6, 31209. doi:10.1038/srep31209
- 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. *Psychol Med*, 46(2), 327-343. doi:10.1017/S0033291715001981
- Bonanno, G. A. (2004). Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am Psychol*, 59(1), 20-28. doi:10.1037/0003-066X.59.1.20
- Boscarino, C., Nalpathamkalam, T., Pellicchia, G., Li, W., Thiruvahindrapuram, B., & Merico, D. (2019). Using Next-Generation Sequencing Transcriptomics To Determine Markers of Post-traumatic Symptoms: Preliminary Findings from a Post-deployment Cohort of Soldiers. *G3 (Bethesda)*, 9(2), 463-471. doi:10.1534/g3.118.200516
- Breen, M. S., Maihofer, A. X., Glatt, S. J., Tylee, D. S., Chandler, S. D., Tsuang, M. T., . . . Woelk, C. H. (2015). Gene networks specific for innate immunity define post-traumatic stress disorder. *Mol Psychiatry*, 20(12), 1538-1545. doi:10.1038/mp.2015.9

- Breen, M. S., Bierer, L. M., Daskalakis, N. P., Bader, H. N., Makotkine, I., Chattopadhyay, M., Xu, C., Grice, A. B., Tocheva, A. S., Flory, J. D., Buxbaum, J. D., Meaney, M. J., Brennard, K., & Yehuda, R. (2020). Correction: Differential transcriptional response following glucocorticoid activation in cultured blood immune cells: a novel approach to PTSD biomarker development. *Translational psychiatry*, *10*(1), 1.  
<https://doi.org/10.1038/s41398-019-0665-5>
- 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. *Arch Gen Psychiatry*, *55*(7), 626-632. doi:10.1001/archpsyc.55.7.626
- Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol*, *68*(5), 748-766. doi:10.1037//0022-006x.68.5.748
- Cattaneo, A., Cattane, N., Begni, V., Pariante, C. M., & Riva, M. A. (2016). The human BDNF gene: peripheral gene expression and protein levels as biomarkers for psychiatric disorders. *Transl Psychiatry*, *6*(11), e958. doi:10.1038/tp.2016.214
- 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.  
doi:10.1371/journal.pone.0039184
- Chen, J., Tan, Y., Sun, F., Hou, L., Zhang, C., Ge, T., . . . Zhang, P. (2020). Single-cell transcriptome and antigen-immunoglobulin analysis reveals the diversity of B cells in non-small cell lung cancer. *Genome Biol*, *21*(1), 152. doi:10.1186/s13059-020-02064-6

- Cookson, W., Liang, L., Abecasis, G., Moffatt, M., & Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nature reviews. Genetics*, *10*(3), 184–194. <https://doi.org/10.1038/nrg2537>
- Fulton, J. J., Calhoun, P. S., Wagner, H. R., Schry, A. R., Hair, L. P., Feeling, N., . . . Beckham, J. C. (2015). The prevalence of posttraumatic stress disorder in Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) Veterans: a meta-analysis. *J Anxiety Disord*, *31*, 98-107. doi:10.1016/j.janxdis.2015.02.003
- Guardado, P., Olivera, A., Rusch, H. L., Roy, M., Martin, C., Lejbman, N., . . . Gill, J. M. (2016). Altered gene expression of the innate immune, neuroendocrine, and nuclear factor-kappa B (NF-kappaB) systems is associated with posttraumatic stress disorder in military personnel. *J Anxiety Disord*, *38*, 9-20. doi:10.1016/j.janxdis.2015.12.004
- Heinzelmann, M., Reddy, S. Y., French, L. M., Wang, D., Lee, H., Barr, T., Baxter, T., Mysliwicz, V., & Gill, J. (2014). Military personnel with chronic symptoms following blast traumatic brain injury have differential expression of neuronal recovery and epidermal growth factor receptor genes. *Frontiers in neurology*, *5*, 198. <https://doi.org/10.3389/fneur.2014.00198>
- Hochberg, Y., & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in medicine*, *9*(7), 811–818. <https://doi.org/10.1002/sim.4780090710>
- Hoffman, G. E. & Roussos, P. (2020). “dream: Powerful differential expression analysis for repeated measures designs.” *Bioinformatics*. doi: 10.1093/bioinformatics/btaa687.

- Hoffman, G. E. & Schadt, E. E. (2016). “variancePartition: Interpreting drivers of variation in complex gene expression studies.” *BMC Bioinformatics*, **17**. doi: 10.1186/s12859-016-1323-z.
- Hoffman, G. E., Ma, Y., Montgomery, K. S., Bendl, J., Jaiswal, M. K., Kozlenkov, A., . . . Roussos, P. (2020). Sex differences in the human brain transcriptome of cases with schizophrenia. bioRxiv, 2020.2010.2005.326405. doi:10.1101/2020.10.05.326405
- Hoge, C. W., Terhakopian, A., Castro, C. A., Messer, S. C., & Engel, C. C. (2007). Association of posttraumatic stress disorder with somatic symptoms, health care visits, and absenteeism among Iraq war veterans. *Am J Psychiatry*, *164*(1), 150-153.  
doi:10.1176/ajp.2007.164.1.150
- Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., . . . Koenen, K. C. (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *Eur J Psychotraumatol*, *8*(sup5), 1353383. doi:10.1080/20008198.2017.1353383
- 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. *Arch Gen Psychiatry*, *62*(6), 593-602.  
doi:10.1001/archpsyc.62.6.593
- Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, *62*(6), 617-627. doi:10.1001/archpsyc.62.6.617

- Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Arch Gen Psychiatry*, 52(12), 1048-1060. doi:10.1001/archpsyc.1995.03950240066012
- Kipanyula, M. J., Kimaro, W. H., & Seke Etet, P. F. (2016). The Emerging Roles of the Calcineurin-Nuclear Factor of Activated T-Lymphocytes Pathway in Nervous System Functions and Diseases. *Journal of aging research*, 2016, 5081021. <https://doi.org/10.1155/2016/5081021>
- Koenen, K. C., Amstadter, A. B., & Nugent, N. R. (2009). Gene-environment interaction in posttraumatic stress disorder: an update. *J Trauma Stress*, 22(5), 416-426. doi:10.1002/jts.20435
- Kuan, P. F., Waszczuk, M. A., Kotov, R., Clouston, S., Yang, X., Singh, P. K., . . . Luft, B. J. (2017). Gene expression associated with PTSD in World Trade Center responders: An RNA sequencing study. *Transl Psychiatry*, 7(12), 1297. doi:10.1038/s41398-017-0050-1
- Kuan, P. F., Clouston, S., Yang, X., Kotov, R., Bromet, E., & Luft, B. J. (2020). Molecular linkage between post-traumatic stress disorder and cognitive impairment: a targeted proteomics study of World Trade Center responders. *Translational psychiatry*, 10(1), 269. <https://doi.org/10.1038/s41398-020-00958-4>
- Li, M., Huang, L., Grigoriou-Serbanescu, M., Bergen, S. E., Landen, M., Hultman, C. M., . . . Swedish Bipolar Study, G. (2016). Convergent Lines of Evidence Support LRP8 as a Susceptibility Gene for Psychosis. *Mol Neurobiol*, 53(10), 6608-6619. doi:10.1007/s12035-015-9559-6



- Logue, M. W., Baldwin, C., Guffanti, G., Melista, E., Wolf, E. J., Reardon, A. F., . . . Miller, M. W. (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. *Mol Psychiatry*, 18(8), 937-942. doi:10.1038/mp.2012.113
- Logue, M. W., Smith, A. K., Baldwin, C., Wolf, E. J., Guffanti, G., Ratanatharathorn, A., . . . Miller, M. W. (2015). An analysis of gene expression in PTSD implicates genes involved in the glucocorticoid receptor pathway and neural responses to stress. *Psychoneuroendocrinology*, 57, 1-13. doi:10.1016/j.psyneuen.2015.03.016
- Lowe, S. R., Galea, S., Uddin, M., & Koenen, K. C. (2014). Trajectories of posttraumatic stress among urban residents. *Am J Community Psychol*, 53(1-2), 159-172. doi:10.1007/s10464-014-9634-6
- McLaughlin, K. A., Alvarez, K., Fillbrunn, M., Green, J. G., Jackson, J. S., Kessler, R. C., . . . Alegria, M. (2019). Racial/ethnic variation in trauma-related psychopathology in the United States: a population-based study. *Psychol Med*, 49(13), 2215-2226. doi:10.1017/S0033291718003082
- McLaughlin, K. A., Koenen, K. C., Friedman, M. J., Ruscio, A. M., Karam, E. G., Shahly, V., . . . Kessler, R. C. (2015). Subthreshold posttraumatic stress disorder in the world health organization world mental health surveys. *Biol Psychiatry*, 77(4), 375-384. doi:10.1016/j.biopsych.2014.03.028
- Mehta, D., Gonik, M., Klengel, T., Rex-Haffner, M., Menke, A., Rubel, J., . . . Binder, E. B. (2011). Using polymorphisms in FKBP5 to define biologically distinct subtypes of

- posttraumatic stress disorder: evidence from endocrine and gene expression studies. *Arch Gen Psychiatry*, 68(9), 901-910. doi:10.1001/archgenpsychiatry.2011.50
- Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K., Altmann, A., Pace, T. W., . . . Binder, E. B. (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc Natl Acad Sci U S A*, 110(20), 8302-8307. doi:10.1073/pnas.1217750110
- Mehta, D., Miller, O., Bruenig, D., David, G., & Shakespeare-Finch, J. (2020). A Systematic Review of DNA Methylation and Gene Expression Studies in Posttraumatic Stress Disorder, Posttraumatic Growth, and Resilience. *J Trauma Stress*, 33(2), 171-180. doi:10.1002/jts.22472
- Michopoulos, V., Powers, A., Gillespie, C. F., Ressler, K. J., & Jovanovic, T. (2017). Inflammation in Fear- and Anxiety-Based Disorders: PTSD, GAD, and Beyond. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 42(1), 254–270. <https://doi.org/10.1038/npp.2016.146>
- Mills, K. L., McFarlane, A. C., Slade, T., Creamer, M., Silove, D., Teesson, M., & Bryant, R. (2011). Assessing the prevalence of trauma exposure in epidemiological surveys. *Aust N Z J Psychiatry*, 45(5), 407-415. doi:10.3109/00048674.2010.543654
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., . . . Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 12(5), 453-457. doi:10.1038/nmeth.3337
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., . . . Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk

tissues with digital cytometry. *Nat Biotechnol*, 37(7), 773-782. doi:10.1038/s41587-019-0114-2

Osumi, N., Shinohara, H., Numayama-Tsuruta, K., & Maekawa, M. (2008). Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. *Stem cells (Dayton, Ohio)*, 26(7), 1663–1672. <https://doi.org/10.1634/stemcells.2007-0884>

Pattinson, C. L., Guedes, V. A., Edwards, K., Mithani, S., Yun, S., Taylor, P., . . . Gill, J. M. (2020). Excessive daytime sleepiness is associated with altered gene expression in military personnel and veterans with posttraumatic stress disorder: an RNA sequencing study. *Sleep*, 43(9). doi:10.1093/sleep/zsaa036

Penzes, P., Buonanno, A., Passafaro, M., Sala, C., & Sweet, R. A. (2013). Developmental vulnerability of synapses and circuits associated with neuropsychiatric disorders. *Journal of neurochemistry*, 126(2), 165–182. <https://doi.org/10.1111/jnc.12261>

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. *J Consult Clin Psychol*, 61(6), 984-991. doi:10.1037//0022-006x.61.6.984

Rosen, R. L., Levy-Carrick, N., Reibman, J., Xu, N., Shao, Y., Liu, M., Ferri, L., Kazeros, A., Caplan-Shaw, C. E., Pradhan, D. R., Marmor, M., & Galatzer-Levy, I. R. (2017). Elevated C-reactive protein and posttraumatic stress pathology among survivors of the 9/11 World Trade Center attacks. *Journal of psychiatric research*, 89, 14–21. <https://doi.org/10.1016/j.jpsychires.2017.01.007>

- Roberts, A. L., Gilman, S. E., Breslau, J., Breslau, N., & Koenen, K. C. (2011). Race/ethnic differences in exposure to traumatic events, development of post-traumatic stress disorder, and treatment-seeking for post-traumatic stress disorder in the United States. *Psychol Med*, 41(1), 71-83. doi:10.1017/S0033291710000401
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rusch, H. L., Robinson, J., Yun, S., Osier, N. D., Martin, C., Brewin, C. R., & Gill, J. M. (2019). Gene expression differences in PTSD are uniquely related to the intrusion symptom cluster: A transcriptome-wide analysis in military service members. *Brain Behav Immun*, 80, 904-908. doi:10.1016/j.bbi.2019.04.039
- Sarapas, C., Cai, G., Bierer, L. M., Golier, J. A., Galea, S., Ising, M., . . . Yehuda, R. (2011). Genetic markers for PTSD risk and resilience among survivors of the World Trade Center attacks. *Dis Markers*, 30(2-3), 101-110. doi:10.3233/DMA-2011-0764
- Segman, R. H., Shefi, N., Goltser-Dubner, T., Friedman, N., Kaminski, N., & Shalev, A. Y. (2005). Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. *Mol Psychiatry*, 10(5), 500-513, 425. doi:10.1038/sj.mp.4001636
- Sharma, A., Merritt, E., Hu, X., Cruz, A., Jiang, C., Sarkodie, H., Zhou, Z., Malhotra, J., Riedlinger, G. M., & De, S. (2019). Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors. *Cell reports*, 29(8), 2164–2174.e5. <https://doi.org/10.1016/j.celrep.2019.10.045>

- Tylee, D. S., Chandler, S. D., Nievergelt, C. M., Liu, X., Pazol, J., Woelk, C. H., . . . Marine Resiliency Study, I. (2015). Blood-based gene-expression biomarkers of post-traumatic stress disorder among deployed marines: A pilot study. *Psychoneuroendocrinology*, 51, 472-494. doi:10.1016/j.psyneuen.2014.09.024
- Uddin, M., Aiello, A. E., Wildman, D. E., Koenen, K. C., Pawelec, G., de Los Santos, R., Goldmann, E., & 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.  
<https://doi.org/10.1073/pnas.0910794107>
- Watson, P. (2019). PTSD as a Public Mental Health Priority. *Curr Psychiatry Rep*, 21(7), 61.  
doi:10.1007/s11920-019-1032-1
- Weiner, M. W., Veitch, D. P., Hayes, J., Neylan, T., Grafman, J., Aisen, P. S., . . . Department of Defense Alzheimer's Disease Neuroimaging, I. (2014). Effects of traumatic brain injury and posttraumatic stress disorder on Alzheimer's disease in veterans, using the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement*, 10(3 Suppl), S226-235. doi:10.1016/j.jalz.2014.04.005
- Yehuda, R., Cai, G., Golier, J. A., Sarapas, C., Galea, S., Ising, M., . . . Buxbaum, J. D. (2009). Gene expression patterns associated with posttraumatic stress disorder following exposure to the World Trade Center attacks. *Biol Psychiatry*, 66(7), 708-711.  
doi:10.1016/j.biopsych.2009.02.034

Yehuda, R., Hoge, C. W., McFarlane, A. C., Vermetten, E., Lanius, R. A., Nievergelt, C. M., . . .

Hyman, S. E. (2015). Post-traumatic stress disorder. *Nat Rev Dis Primers*, 1, 15057.

doi:10.1038/nrdp.2015.57

Yehuda, R., & LeDoux, J. (2007). Response variation following trauma: a translational

neuroscience approach to understanding PTSD. *Neuron*, 56(1), 19-32.

doi:10.1016/j.neuron.2007.09.006

Ypsilanti, A. R., & Rubenstein, J. L. (2016). Transcriptional and epigenetic mechanisms of early

cortical development: An examination of how Pax6 coordinates cortical

development. *The Journal of comparative neurology*, 524(3), 609–629.

<https://doi.org/10.1002/cne.23866>

Zieker, J., Zieker, D., Jatzko, A., Dietzsch, J., Nieselt, K., Schmitt, A., . . . Gebicke-Haerter, P. J.

(2007). Differential gene expression in peripheral blood of patients suffering from post-

traumatic stress disorder. *Mol Psychiatry*, 12(2), 116-118. doi:10.1038/sj.mp.4001905

Zohar, J., Fostick, L., Cohen, A., Bleich, A., Dolfín, D., Weissman, Z., . . . Israeli Consortium

on, P. (2009). Risk factors for the development of posttraumatic stress disorder following

combat trauma: a semiprospective study. *J Clin Psychiatry*, 70(12), 1629-1635.

doi:10.4088/JCP.08m04378