

SUPPLEMENTARY FIGURES

Title: PTSD blood transcriptome mega-analysis:
Shared inflammatory pathways across biological sex and modes of trauma

Authors: Michael S. Breen,^{1,2*} Daniel S. Tylee,³ Adam X. Maihofer⁴,
Thomas C. Neylan^{5,6}, Divya Mehta⁷, Elizabeth Binder^{8,9}, Sharon D. Chandler⁴,
Jonathan L. Hess³, William S. Kremen^{4,10}, Victoria B. Risbrough^{4,10}, Christopher H.
Woelk¹¹, Dewleen G. Baker^{4,10}, Caroline M. Nievergelt^{4,10}, Ming T. Tsuang^{4,10},
Joseph D. Buxbaum^{1,2} and Stephen J. Glatt³

Affiliations: ¹Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ²Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; ³Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab); Departments of Psychiatry and Behavioral Sciences & Neuroscience and Physiology; SUNY Upstate Medical University; Syracuse, NY, USA; ⁴Department of Psychiatry, University of California San Diego, California, USA; ⁵Department of Psychiatry, University of California, San Francisco; ⁶San Francisco Veterans Affairs Medical Center; ⁷School of Psychology and Counseling, Faculty of Health, Queensland University of Technology, Kelvin Grove, Queensland, Australia; ⁸Department of Translational Research in Psychiatry, Max-Planck Institute of Psychiatry, Munich, Germany; ⁹Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA; ¹⁰Veterans Affairs Center of Excellence for Stress and Mental Health, San Diego, California, USA; ¹¹Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK.

* To whom correspondence should be addressed:

Email: michael.breen@mssm.edu

Supplementary Figure 1. Multi-step analytic workflow used in the current study.

Supplementary Figure 2. Mega-analytic differential expression and unsupervised clustering.

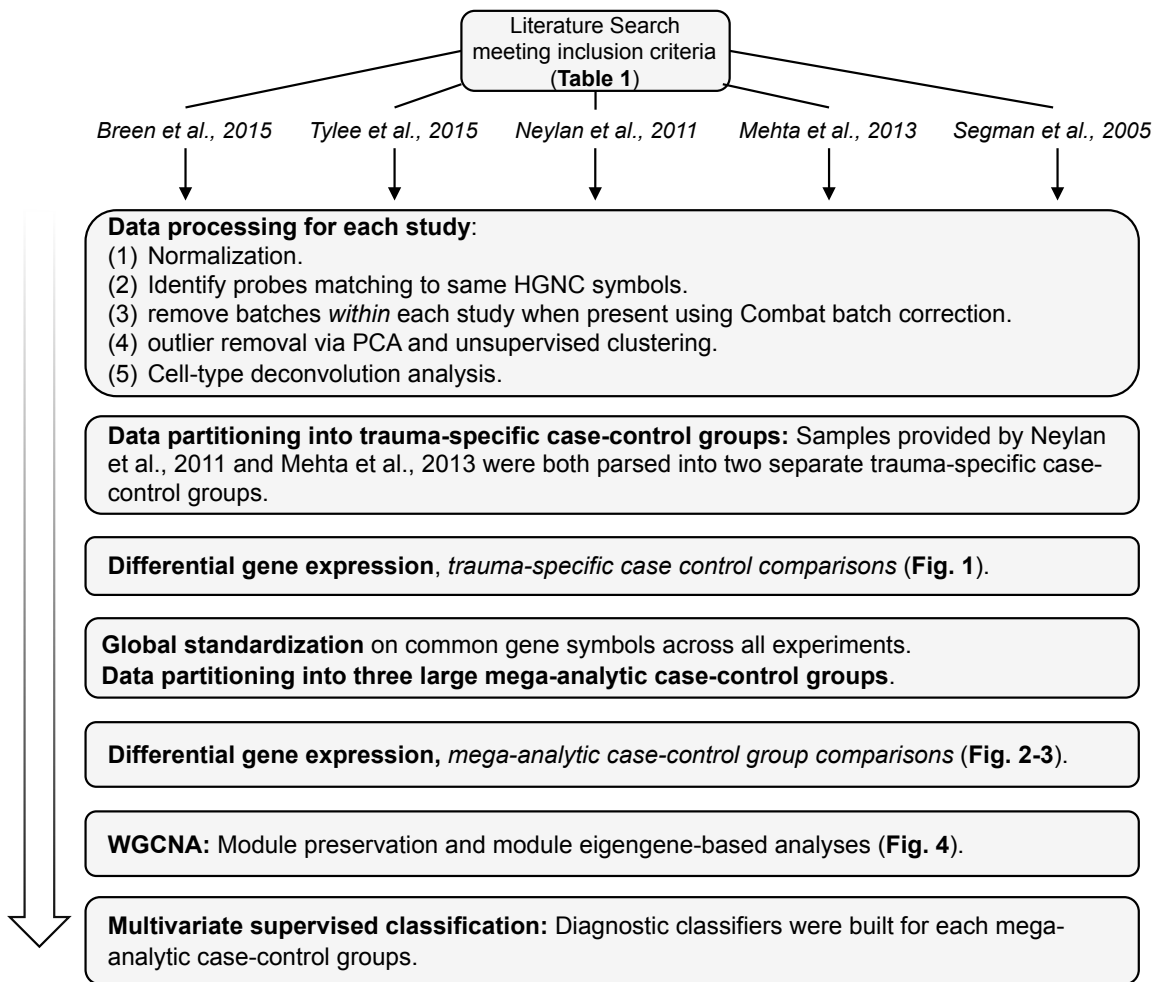
Supplementary Figure 3. STRING protein-protein interaction (PPI) network.

Supplementary Figure 4. Gene and gene-ontology cross-disorder overlaps.

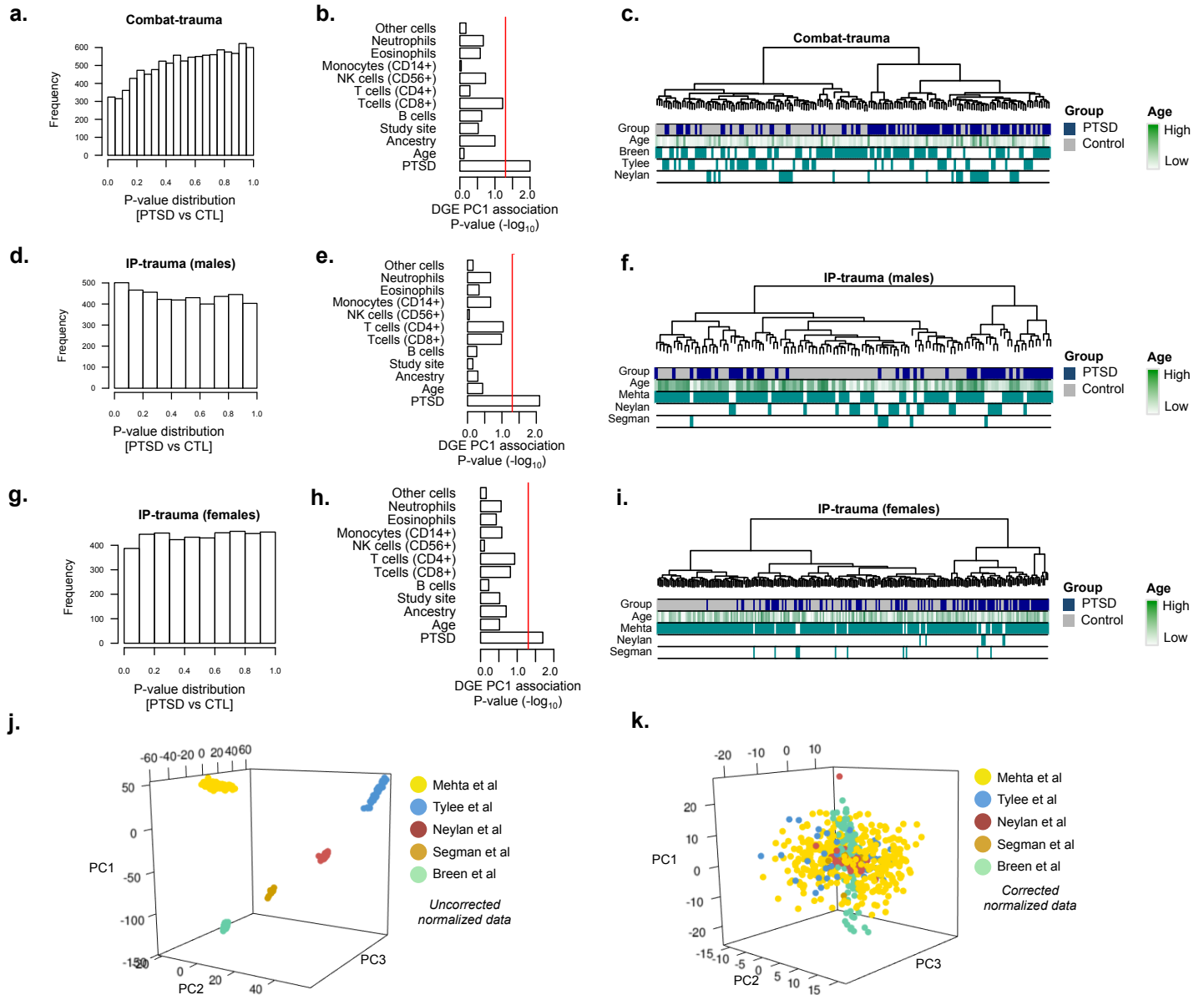
Supplementary Figure 5. WGCNA preservation analyses.

Supplementary Figure 6. Supervised multivariate classification results.

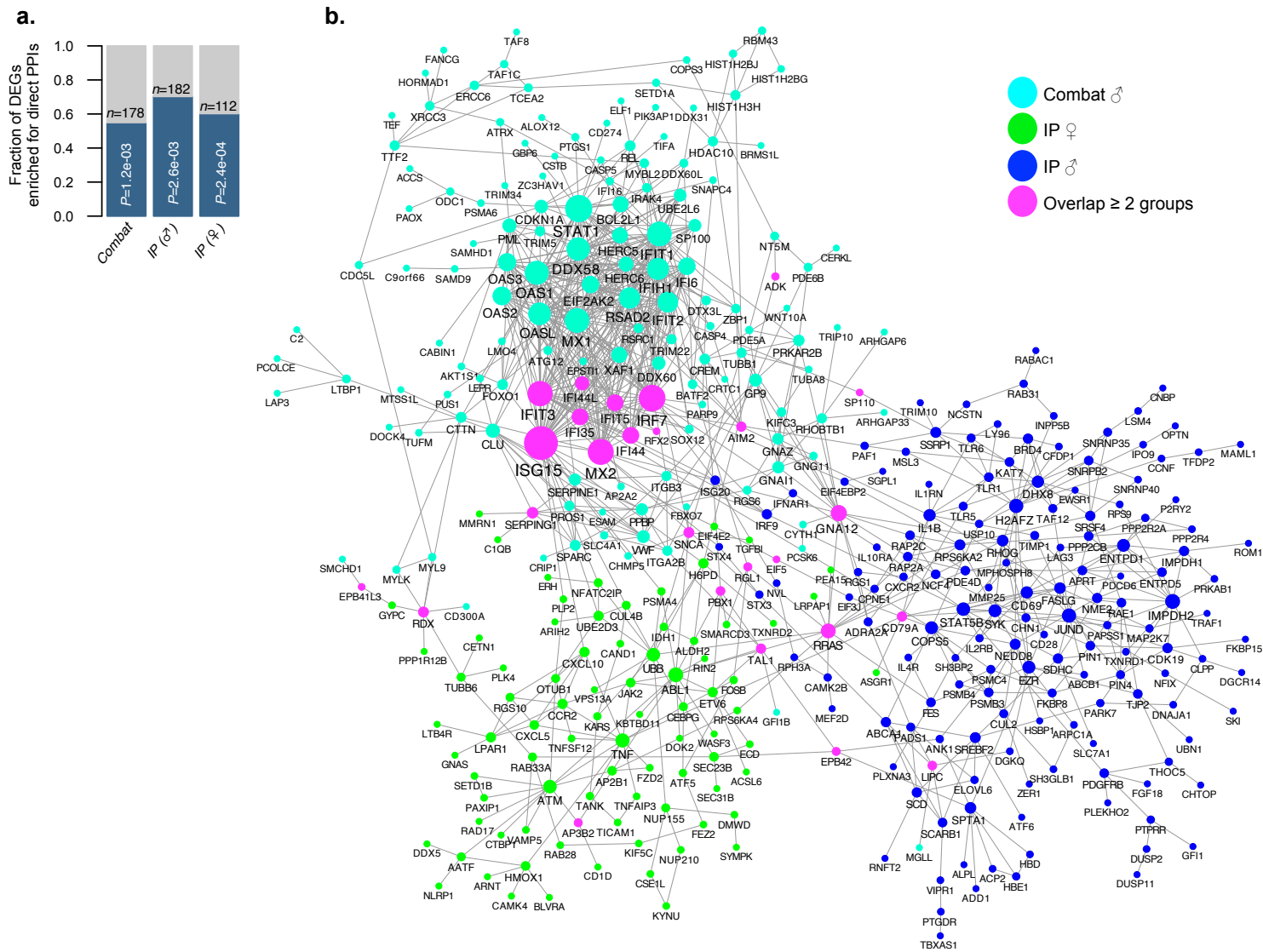
For all table files, please find supplementary table file.



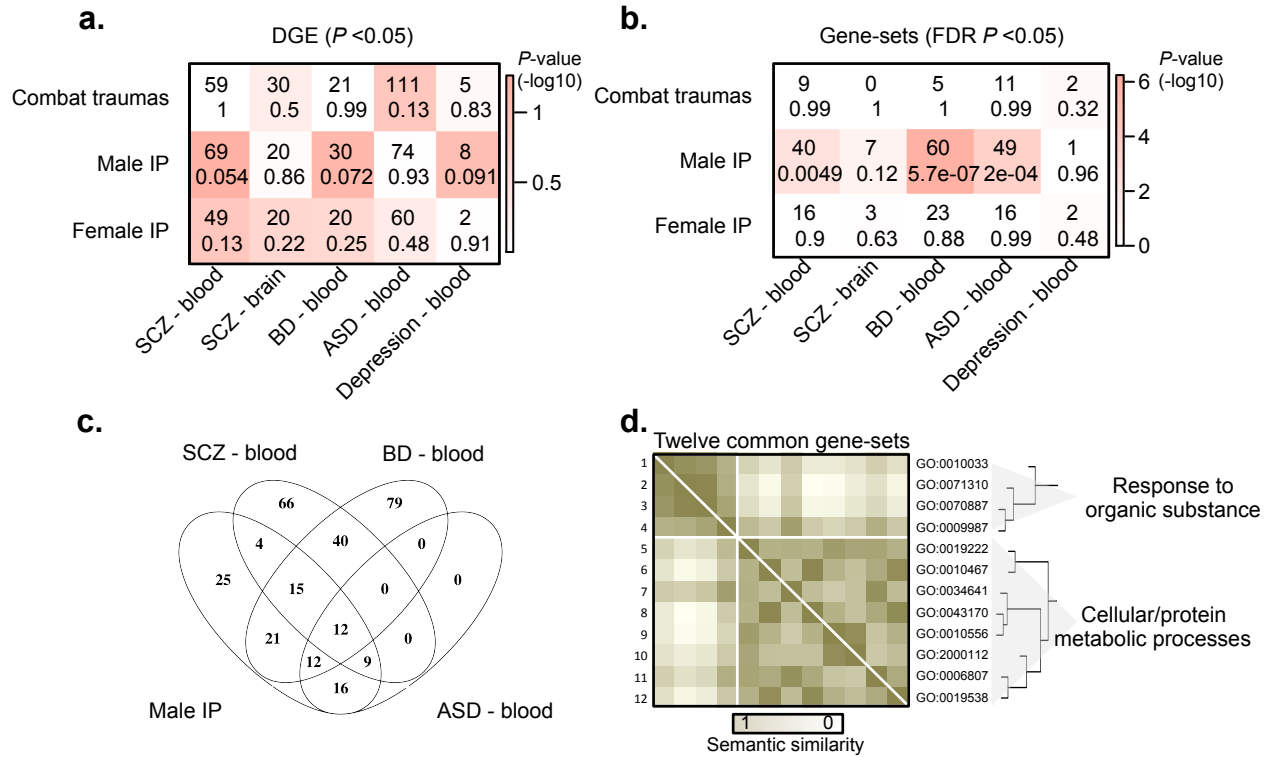
Supplementary Figure 1. Multi-step analytic workflow used in the current study.



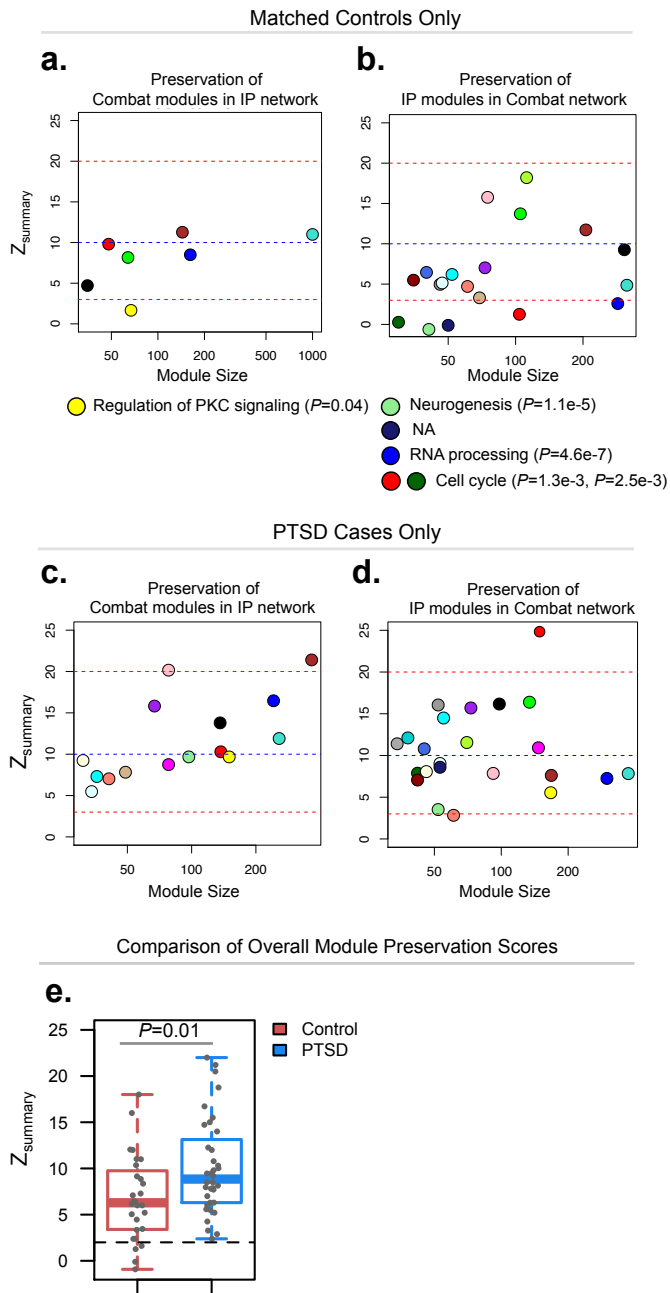
Supplementary Figure 2. Mega-analytic differential expression and unsupervised clustering. **(A, D, G)** Histograms of P -value distributions for each of the three mega-analytic case-control comparisons. **(B,E,H)** The first principal component of DGE signatures for each mega-analytic group are primarily associated with PTSD diagnosis, and not with age, ancestry, study site or estimated cell-type frequencies; red line indicates $P < 0.05$. Average linkage hierarchical clustering of the samples in each mega-analytic case-control group based on DGE signatures ($P < 0.05$) identified in **(C)** PTSD males exposed to combat trauma, **(F)** PTSD males exposed to IP trauma and **(I)** PTSD females exposed to IP trauma. Principal component analysis of all available gene expression studies **(J)** prior to and **(K)** following global standardization.



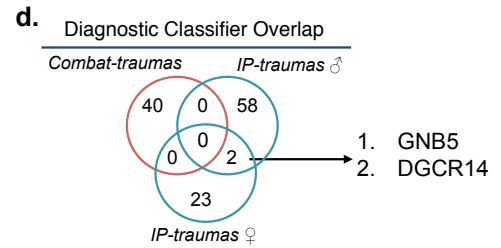
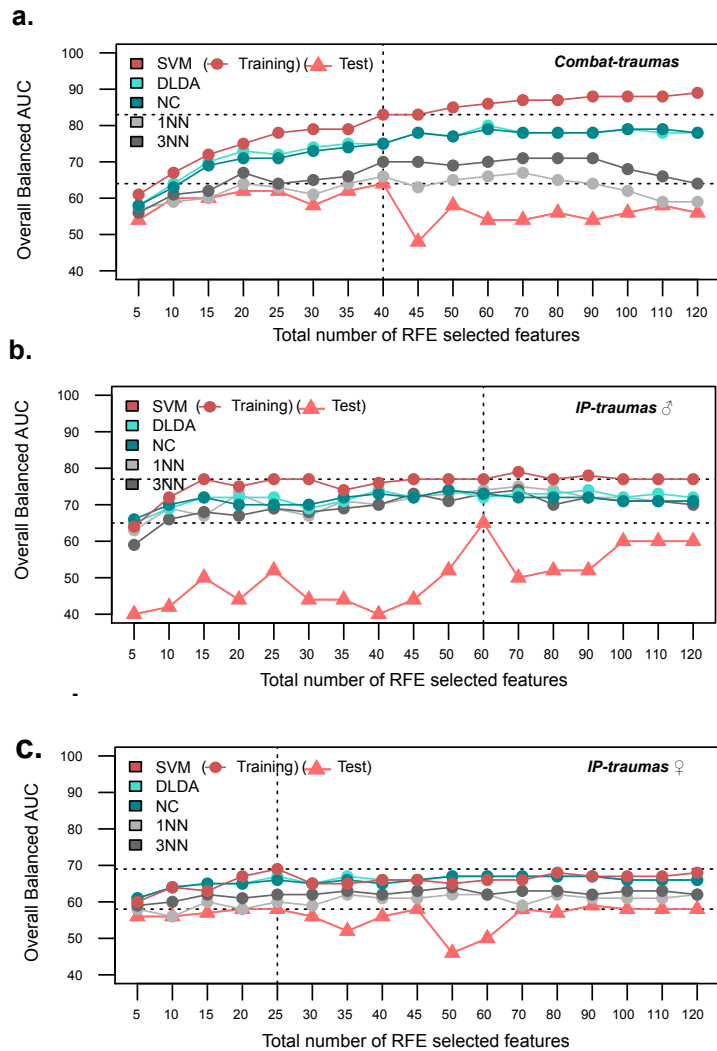
Supplementary Figure 3. STRING protein-protein interaction (PPI) network. **(A)** DGE lists were significantly enriched for observed direct protein-protein interactions (PPI) ($n=178$, 182 , 112), beyond what is expected by chance using a nontrivial random background model. **(B)** Three separate networks were constructed in Cytoscape and the union of all networks was analyzed. Larger node size indicates a larger set of degrees (connections) with protein neighbors. Colored nodes indicate: cyan, males exposed to combat-related traumas; green, males exposed to IP-related traumas; blue, females exposed to IP-related traumas; pink, nodes shared between 2 or more networks.



Supplementary Figure 4. Cross-disorder overlap of mega-analysis derived differential gene expression signatures. **(A)** PTSD gene expression signatures ($P < 0.05$) in this study are overlapped with those found in other large-scale mega-analyses, including schizophrenia (SCZ) peripheral blood and postmortem brain tissue, bipolar (BD) peripheral blood and autism spectrum disorder (ASD) peripheral blood. **(B)** Overlap at the gene-ontology (GO) level, for significantly altered biological processes (FDR $P < 0.05$). The top number in each cell represents the number of overlapping genes while the bottom number represents the corresponding P -value significance determined by a one-sided Fisher's exact test. Cells are shaded by level of significance. **(C)** Venn diagram of gene-sets between PTSD males exposed to IP, SCZ, BD and ASD. **(D)** Semantic similarity scores for twelve common GO pairs clustered by hierarchical clustering method (left) and common ancestor terms (right).



Supplementary Figure 5. WGCNA preservation analyses. Module preservation assessed co-regulatory patterns in control individuals exposed to **(A)** combat-trauma relative to IP-related traumas and **(B)** IP-related traumas relative to combat-related trauma. Similarly, co-regulatory patterns were assessed in PTSD cases exposed to **(C)** combat-trauma relative to IP-related traumas and **(D)** IP-related traumas relative to combat-related traumas. **(E)** We then assessed whether different traumatic backgrounds (that is, combat or IP traumas) effect co-regulatory patterns in control individuals differently than in PTSD cases. Collectively, when comparing the Z_{summary} scores computed within control individuals **(A-B)** to Z_{summary} scores computed within PTSD cases **(C-D)** significantly less preservation was observed in controls relative to PTSD cases ($P=0.01$; Wilcox Test) indicating a more heterogenous response to trauma within control individuals compared to PTSD cases. The top ranked FDR corrected $P < 0.05$ GO-term is reported for each module with $Z_{\text{summary}} < 2$. For complete lists of genes within each module see **Supplementary Table 4**.



Supplementary Figure 6. Supervised multivariate classification results. Five best-in-class machine-learning algorithms were applied to distinguish (A) PTSD males from controls exposed to combat trauma, (B) PTSD males from control individuals composed to IP-related traumas and (C) PTSD females from controls exposed to IP traumas using a nested leave-one-tenth out cross-validation approach. Recursive feature elimination (RFE) was used to select the top most informative features on the training data (70%) and iteratively worked down to the 5 most informative features, while testing prediction accuracies on completely withheld test data (30%). RFE selected features (x-axis) is plotted against overall balanced accuracy (AUC; y-axis) for each mega-analytical comparison. (D) Overlap of final gene classifiers identified 2 genes in common across males and females exposed to IP-traumas (GNB5 and DGCR14).