BMC
Systems Biology

RESEARCH ARTICLE

Core modular blood and brain biomarkers in social defeat mouse model for post traumatic stress disorder

Ruoting Yang[1,5,6†], Bernie J Daigle Jr[2†], Seid Y Muhie[5], Rasha Hammamieh[5], Marti Jett[5], Linda Petzold[1,2,3]
and Francis J Doyle III[1,4*]

Abstract

Background: Post-traumatic stress disorder (PTSD) is a severe anxiety disorder that affects a substantial portion of combat veterans and poses serious consequences to long-term health. Consequently, the identification of diagnostic and prognostic blood biomarkers for PTSD is of great interest. Previously, we assessed genome-wide gene expression of seven brain regions and whole blood in a social defeat mouse model subjected to various stress conditions.

Results: To extract biological insights from these data, we have applied a new computational framework for identifying gene modules that are activated in common across blood and various brain regions. Our results, in the form of modular gene networks that highlight spatial and temporal biological functions, provide a systems-level molecular description of response to social stress. Specifically, the common modules discovered between the brain and blood emphasizes molecular transporters in the blood-brain barrier, and the associated genes have significant overlaps with known blood signatures for PTSD, major depression, and bipolar disease. Similarly, the common modules specific to the brain highlight the components of the social defeat stress response (e.g., fear conditioning pathways) in each brain sub-region.

Conclusions: Many of the brain-specific genes discovered are consistent with previous independent studies of PTSD or other mental illnesses. The results from this study further our understanding of the mechanism of stress response and contribute to a growing list of diagnostic biomarkers for PTSD.

Background

Post-traumatic stress disorder (PTSD) is an anxiety disorder that is triggered after exposure to traumatic events. Individuals with PTSD have persistent fear memory and often feel emotionally numb. If left untreated, PTSD can be life-threatening, as it is often linked with substance abuse and severe depression. A study of 289,328 Iraq and Afghanistan veterans who were first-time users of Veterans Affairs (VA) health care between 2002 and 2008 showed that 22% of veterans were diagnosed with PTSD and 17% were diagnosed with depression [1]. Given the

predominance of PTSD and its negative consequences to long term health, it is very important to identify measurable and quantifiable biological parameters, i.e., biomarkers, which can serve as prognostic and diagnostic indicators for PTSD. Recent studies have proposed several candidate brain gene biomarkers that are associated with PTSD [2,3]. Even though PTSD is an illness of the brain, taking brain biopsy or spinal fluid is not a viable option for diagnosis. Instead, blood can be used as a surrogate for brain tissue for the purpose of identifying biomarkers [4-8]. Specifically, Rollins et al. recently found over 4,100 brain transcripts co-expressed in the blood of healthy human subjects [9]. Furthermore, it was shown that the mRNA levels of certain transcripts in PTSD patients remain changed with respect to controls even 16 years after the traumatic event [8,10]. Thus, blood gene expression assays are of particular interest for both short-term and

* Correspondence: doyle@engineering.ucsb.edu
†Equal contributors
[1]Institute for Collaborative Biotechnologies, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA
[4]Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA
Full list of author information is available at the end of the article

long-term diagnosis, prognosis, and treatment of PTSD. However, the identification of predictive blood markers requires the accurate separation of biologically relevant core markers from unrelated downstream signals. This task is particularly challenging when using surrogate tissues, since biological noise from the surrogate is confounded with noise from the primary tissue. Fortunately, studies performed with model organisms allow the direct assay of both surrogate and primary tissues. By characterizing the molecular changes present in both tissues simultaneously, we can more effectively filter out spurious signals in the surrogate.

We recently used repeated exposures of mice to a trained aggressor mouse as a "social defeat" model for evaluating PTSD symptoms [11]. This social defeat model has often been used to induce anxiety, depression-like and avoidance symptoms, which are the most prominent psychiatric features of PTSD and common co-morbidities. Using a "cage-within-cage resident-intruder" protocol (designed to model unpredictable threats of daily trauma), we exposed individual subject male C57BL/6 J mice to single aggressors for six hours daily for 5 or 10 days, and we placed individual control subject mice in the same cages but in the absence of any other mice. After allowing the subject animals to recover for either 1 or 10 days (5 day exposure) or 1 or 42 days (10 day exposure), we then collected tissue samples of blood and seven brain regions of mice under the different stress conditions and measured gene expression levels of these tissues using DNA microarrays. As described in [11], the durations of aggressor exposure were chosen to simulate shorter term (5-day) and longer term (10-day) stress. The shortest recovery phase duration (1 day) was chosen to study the immediate effects of stress. The longer of the two recovery phase durations for each exposure time were selected based on behavioral tests conducted throughout the study. These tests demonstrated 5-day exposure defeated mice showed signs of recovery around 10 days post-exposure, and 10-day exposure defeated mice showed signs of recovery at much longer times (up to 42 days post-exposure). Because PTSD represents a persistent stress response, it is important to identify differentially expressed genes (DEGs) active both immediately after the exposure and after a long recovery period. Thus, in the current work we focus on genes consistently over-/under-expressed across all experimental conditions, rather than on DEGs from individual conditions (we will address the latter in future work). The seven brain regions analyzed in this study were chosen due to their known roles in fear memory formation, emotion regulation, and decision-making—all processes important to the development and pathology of PTSD [3]. In particular, the amygdala regulates fear memory and emotional aspects; the

hippocampus is the center for short term memory, and the prefrontal cortex controls decision-making. In addition, the ventral striatum is strongly associated with emotional and motivational aspects of behavior, the stria terminalis serves as a major output pathway of the amygdala, and the septal area plays a role in reward and reinforcement along with the ventral striatum. We note that a similar protocol has also recently been used to profile social defeat-induced gene expression changes in the nucleus accumbens, ventral tegmental area, and blood plasma [12-14].

The field of systems biology has demonstrated that complex diseases such as PTSD are not caused by changes in a single gene or pathway. Rather, changes occur in a hierarchy of gene modules which collectively contribute to disruption of essential cellular functions [15-17]. To characterize this module hierarchy, many researchers have adopted an unsupervised approach [17-25] that constructs a network based on gene expression data and identifies functional modules based on network topology or "guilt-by-association". However, these methods usually face the problem of under-determination, where the number of interactions to be inferred far exceeds the number of independent measurements [22]. Other studies have adopted a supervised network identification approach that begins with a list of "seed" genes and gradually expands the list by adding interacting genes, ultimately resulting in a compact gene module network [26-28]. These supervised approaches have shown good performance for classification tasks, and we expand upon one of them in this work.

Previous computational and experimental work suggests that functional gene modules are highly conserved across conditions, tissues, and species [17,29,30]. Direct comparisons have been made between multiple mouse tissues [21], between human and mouse brains, and between human blood and brain tissue [4]. However, modules inferred separately from different conditions yield partial overlaps at best, which makes drawing comprehensive biological conclusions difficult. Recently, we developed a new module identification tool entitled COMBINER (COre Module Biomarker Identification with Network Exploration) that identifies distinct conserved expression modules across various conditions. The fundamental idea behind COMBINER is to infer candidate modules from data of one condition and validate the inferred modules in other conditions using supervised classification. Those candidate modules that perform well in classifying samples from multiple conditions are then defined as "core modules". There are three advantages to this approach: (1) The resulting modules are compact and thus exclude unrelated downstream signals; (2) The modules are distinct and well-defined with respect to which conditions/tissues/species

invoke them; (3) This method provides multiple robust discriminative biomarkers co-validated in at least two experimental conditions. Given these advantages, we have applied a customized version of COMBINER to mouse social defeat gene expression data deriving from seven brain regions along with blood to identify common expression modules.

In this work, we have attempted to answer two biological questions: *1. Which expression modules act in common between blood and brain tissue of the social defeat mouse model? 2. Which modules act in common between different brain regions?* To do so, we first performed a pair wise comparison of differential gene expression, biological pathways, and GO terms between tissues. We then applied a new version of COMBINER which we modified in two ways (discussed below). First, we used linear models to deconvolve time-dependent effects on gene expression from effects due to social defeat. Second, we developed an improved consensus feature elimination method to identify robust modules from data with a relatively small sample size. Our results, in the form of blood-brain and brain-brain social defeat core module networks, provide a concise biological description of social defeat and generate many candidate PTSD biomarkers for future study.

Results and discussion
Overlaps of DEGs/DEGOs/DEPaths

First, we identified differentially expressed genes (DEGs) in each individual tissue across all time points using a limma moderated t-test [31]. The numbers of significant DEGs ($p \leq 0.05$) for each tissue are listed in blue on the diagonal in Figure 1a. We then established the significance of DEG overlaps by computing a hypergeometric p-value for each pairwise tissue combination (listed in the off-diagonal). Hypergeometric p-values ≤ 0.05 are considered significant (cells highlighted in red in Figure 1a).

Next, we identified differentially expressed Biological Process GO terms (DEGOs) in each tissue by first ranking all genes in descending order of limma significance and then performing Iterative Group Analysis (iGA) [32] for each GO term with ≤ 100 constituent genes. We computed p-values for each term's iGA score using a null distribution obtained via 1000 random permutations of the original gene order. The numbers of significant DEGOs ($p \leq 0.05$) for each tissue are listed in blue on the diagonal in Figure 1b. We established the significance of DEGO overlaps in the same manner as in Figure 1a.

Finally, we identified differentially expressed MSigDB [33] (www.broadinstitute.org/gsea/msigdb/) canonical

(a) Tissue

Tissue	AY	BL	HC	HB	ST	MPFC	SE	VS
Amygdala	1789	0.767	6E-69	1.000	0.518	0.316	0.471	0.007
Blood		1612	0.506	0.424	0.211	0.262	0.705	0.064
Hippocampus			979	0.513	4E-04	0.446	0.095	0.055
Hemibrain				6619	0.164	1.000	0.999	5E-07
Stria Term.					710	0.003	0.822	0.324
Medial PFC						700	0.677	0.551
Septal Region							202	0.124
Ventral Striatum								1113

(b) Tissue

Tissue	AY	BL	HC	HB	ST	MPFC	SE	VS
Amygdala	258	0.514	0.182	0.779	0.992	0.723	0.957	0.785
Blood		324	0.974	0.969	0.926	0.746	0.035	0.609
Hippocampus			277	0.049	0.970	0.058	0.689	0.616
Hemibrain				360	0.968	0.870	0.912	0.935
Stria Term.					211	0.796	0.841	0.807
Medial PFC						249	0.996	0.993
Septal Region							262	0.238
Ventral Striatum								302

(c) Tissue

Tissue	AY	BL	HC	HB	ST	MPFC	SE	VS
Amygdala	17	0.433	0.033	0.622	1.000	1.000	1.000	1.000
Blood		75	0.835	1E-06	0.862	0.875	0.139	0.978
Hippocampus			66	0.998	0.064	0.526	0.388	0.331
Hemibrain				105	0.866	1.000	0.226	0.645
Stria Term.					39	1.000	0.567	0.948
Medial PFC						23	1.000	0.500
Septal Region							42	0.592
Ventral Striatum								63

Figure 1 Overlap analysis between blood and brain regions. Numbers of significant DEGs **(a)**, DEGOs **(b)**, and DEPATHs **(c)** are listed in blue on the diagonal, while hyper geometric p-values are listed in the off-diagonal. P-values ≤ 0.05 are considered significant (cells highlighted in red). We consider the DEPATH overlap between hippocampus and stria terminalis to be marginally significant (red font), as it has a p-value ≤ 0.1 and is supported by a highly significant DEG overlap between the same tissues. (HB: hemibrain (hemisphere), AY: amygdala, HS: hippocampus, MPFC: medial prefrontal cortex, VS: ventral striatum, SE: septal region and ST: stria terminalis).

sub-pathways (DEPATHs) in the same manner as DEGOs with the following modification. For each pathway, we performed iGA separately for all ordered sub-pathways ranging in size from three to 10 genes (when genes are ordered in terms of limma significance). We selected the highest scoring sub-pathway and established significance as before by repeating the procedure on 1000 random gene order permutations. The numbers of significant DEPATHs and significant DEPATH overlaps are denoted in the same manner as above.

The overlaps of particular interest include amygdala-hippocampus (AY-HC) and hippocampus-stria terminalis (HC-ST), as these two scored significantly in the DEG comparison and significantly or nearly significantly, respectively, in the DEPATH comparison. These DEPATHs describe processes such as inflammation, diabetes, apoptosis, and immune response. Tables 1 and 2 show the significantly overlapping DEPATHs of AY-HC and HC-ST, respectively. We list the original name of each sub-pathway along with the following information from the iGA sub-pathway analysis conducted on the hippocampus data: number of genes in the highest scoring sub-pathway (Sig. Genes), sub-pathway permutation p-value, and Benjamini-Hochberg corrected sub-pathway false discovery rate (FDR). We note that none of these pathways would have been identified as significant from the hippocampus data alone when using a FDR ≤ 0.05 cut-off. We also note significant overlaps in the blood-septal region and blood-Hemibrain comparisons, where DEGOs related to apoptosis and DEPATHs related to insulin/diabetes, respectively, were identified. Additional file 1: Table S1 and Additional file 2: Table S2 contains detailed lists of these DEGOs and DEPATHs, respectively.

Core module network

Although the differential expression overlap analysis provided some biological insight into the pairwise molecular similarities between mouse tissues during social defeat, overlap results between DEGs, DEGOs, and DEPATHs were not always consistent. Overlap analysis between multiple tissues is more desired, while these overlaps are very limited due to the high noise-to-signal ratio of microarray. In addition, it was not obvious how best to combine the results into an overall biological

Table 1 Significantly overlapping DEPATHs between amygdala and hippocampus

Name	Number of significant genes	p-value	FDR
BioCarta cytokine pathway	9	<0.002	<0.220
BioCarta inflam pathway	8	0.002	0.220
KEGG type I diabetes mellitus	7	0.012	0.440
KEGG JAK STAT signaling pathway	10	0.014	0.456

FDR denotes false discovery rate.

**Table 2 Significantly overlapping DEPATHs between hippocampus and stria terminalis**

Name	Number of significant genes	p-value	FDR
SA caspase cascade	6	0.005	0.400
BioCarta IL1R pathway	10	0.010	0.440
KEGG cytosolic DNA sensing pathway	9	0.021	0.499
KEGG graft versus host disease	6	0.029	0.562
KEGG prostate cancer	9	0.030	0.562
BioCarta keratinocyte pathway	9	0.046	0.636

description of mouse social defeat. Thus, we turned to a network-level analysis to provide deeper insight. Because the desired diagnostic biomarkers should be generally over-expressed in both the stress treatment and recovery period, we extended the COMBINER method [28] to accommodate all four conditions, which resulted in multiple-time-segment data. However, we would expect an age effect in the control mice. For example, the gene expression patterns of control mice in the 10-day treatment 1-day recovery and 10-day treatment 42-day recovery groups were significantly different due to mouse age. Thus, we used the limma software [31] to deconvolve the undesired effects of differing mouse ages as explanatory variables in a linear model, and we subtracted these variables from the original gene expression values. We then applied COMBINER to the "time standardized" data to construct a blood-brain network (common modules co-expressed in blood and seven brain regions, Figure 2) and a brain-brain network (common modules co-expressed in six brain regions, Figure 3).

Blood-brain network

We first investigated the expression modules active in both blood and multiple brain regions. Starting with the top 100 candidate modules (when ranked by pathway activity absolute t-score—see Methods) inferred from blood sample data, we identified modules that were also active in each brain region. To do so, we removed features using Consensus Feature Elimination until the average classification Area Under the ROC Curve (AUC) evaluated on each brain region exceeded 0.75 (see [28] for additional details). After repeating this procedure separately for all brain regions, a total of nine core modules remained. Figure 2a presents each module's brain region-specific expression patterns. We used average time curves (see Methods) to show the time-specific expression pattern of the modules as heat maps in Figure 2a. Figure 2b further shows the expression of the core modules and the protein-protein interactions (PPIs) between their gene products. The color of each gene denotes its expression level in the blood. Blue lines denote known PPIs within modules, while gray lines
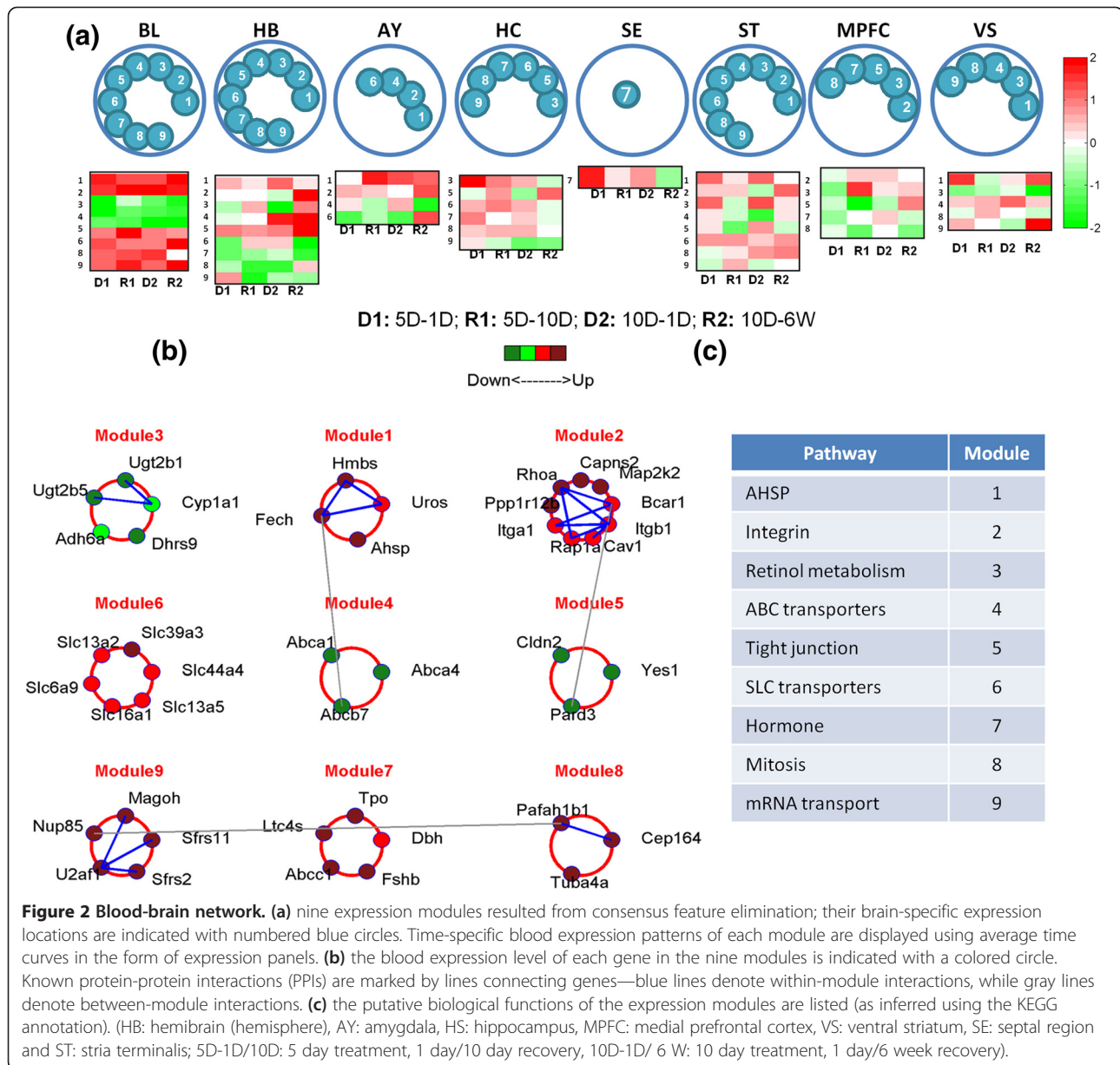
Figure 2 Blood-brain network. (a) nine expression modules resulted from consensus feature elimination; their brain-specific expression locations are indicated with numbered blue circles. Time-specific blood expression patterns of each module are displayed using average time curves in the form of expression panels. **(b)** the blood expression level of each gene in the nine modules is indicated with a colored circle. Known protein-protein interactions (PPIs) are marked by lines connecting genes—blue lines denote within-module interactions, while gray lines denote known PPIs between modules. **(c)** the putative biological functions of the expression modules are listed (as inferred using the KEGG annotation). (HB: hemibrain (hemisphere), AY: amygdala, HS: hippocampus, MPFC: medial prefrontal cortex, VS: ventral striatum, SE: septal region and ST: stria terminalis; 5D-1D/10D: 5 day treatment, 1 day/10 day recovery, 10D-1D/ 6 W: 10 day treatment, 1 day/6 week recovery).

denote known PPIs between modules. Figure 2c lists the putative biological functions of the core modules; detailed module information is summarized in Additional file 3: Table S3. We note that use of COMBINER resulted in seven discriminative blood biomarker sets (average 0.81 mean AUC and 0.26 mean error rate) which have each been validated using data from one of the brain regions. Table 3 lists the final number of modules identified from each blood-brain region pair with the associated mean AUC and mean error rate.

The resulting nine core modules represent biological functions related to molecular transport, integrin and tight junction function, retinol metabolism, cell cycle,

and mRNA transcription. Although initially inferred from blood tissue, most of these processes have been previously implicated in normal and pathological brain function. For example, tight junctions and ABC efflux transporters are present in the blood-brain barrier (BBB) and the blood-cerebrospinal fluid barrier (BCSFB) [34,35], and SLC transporters encode facilitated transporters and ion-coupled secondary active transporters such as neurotransmitters. The latter also represent the major class of transporters used in the delivery of drugs to the brain [36]. In addition, overexpressed integrin genes lead to vascular remodeling, which is believed to be highly correlated with mild Traumatic Brain Injury

Figure 3 (See legend on next page).
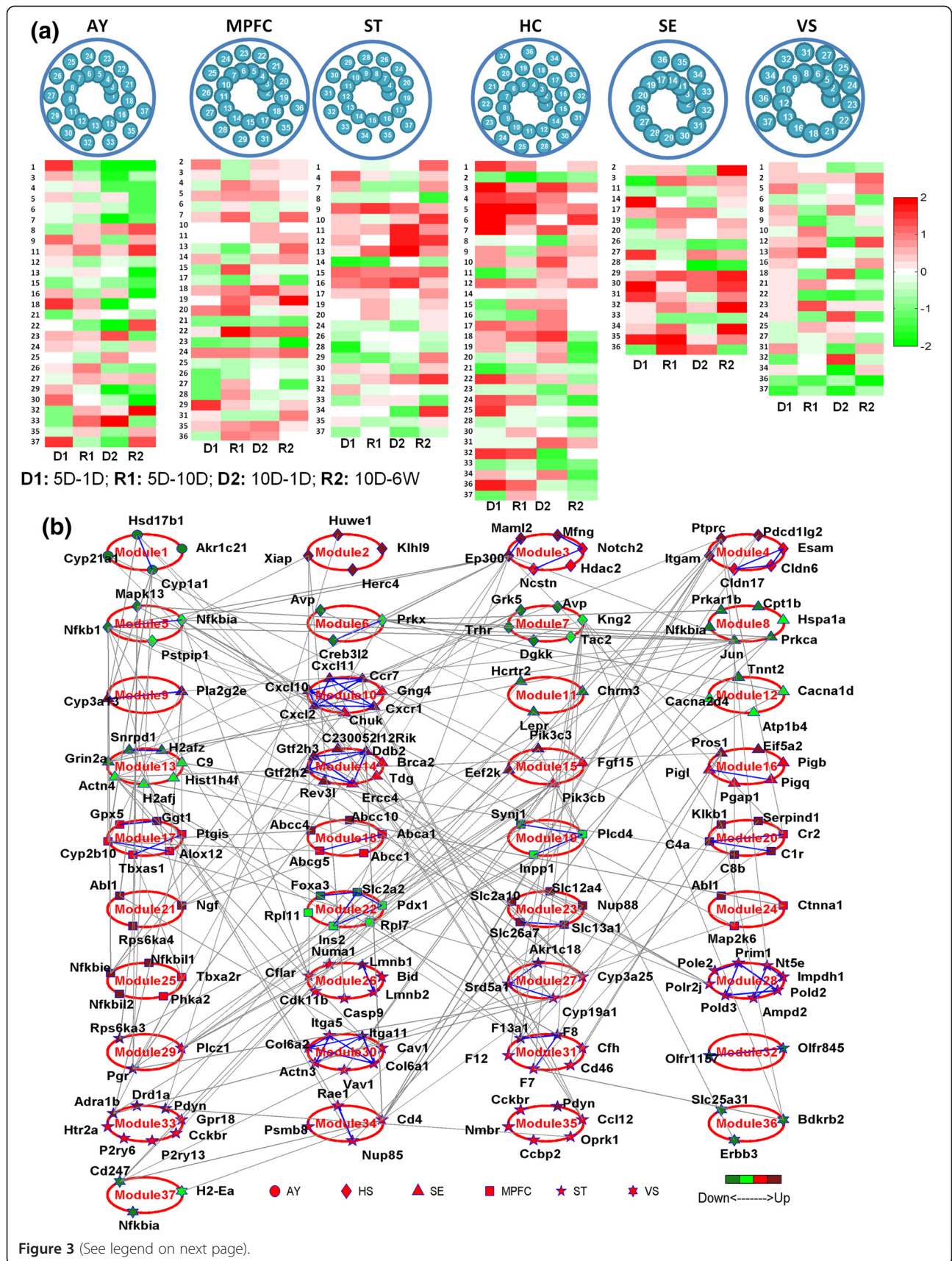
(See figure on previous page).
Figure 3 Brain-brain network. (a) application of COMBINER to brain data yields thirty-seven core modules. The tissue- and time-specific expression patterns of each module are presented in the same manner as before. **(b)** the expression levels and known PPIs of the core module genes are displayed. The shape of a gene represents its inference region, and the color denotes its expression level in that region. Blue lines denote known within-module protein-protein interactions (PPIs), while gray lines denote between-module PPIs. (HB: hemibrain (hemisphere), AY: amygdala, HS: hippocampus, MPFC: medial prefrontal cortex, VS: ventral striatum, SE: septal region and ST: stria terminalis; 5D-1D/10D: 5 day treatment, 1 day/10 day recovery, 10D-1D/ 6 W: 10 day treatment, 1 day/6 week recovery).

(mTBI) [37], a disease related to PTSD. Finally, retinoids are important for the maintenance of the nervous system and may play a role in Alzheimer's disease [38].

The resulting 43 core genes also exhibit ample evidence for association with brain function and/or PTSD. In particular, the genes *Abca4, Fech, Magoh, Ppp1r12b,* and *Uros* were previously shown to be differentially expressed in a human PTSD signature discovered by Segman et al. [8]. Seven of the 43 genes closely resemble genes from a blood signature for depression (*Ahsp, Dhrs9, Map2k2, Slc13a2, Slc16a1, Slc39a3, U2af1*) [39,40], while *Hmbs, Pafah1b1, Sfrs2,* and *Yes1* were previously identified as bipolar disorder blood markers [41]. In addition, *Ugt2b5* and *Slc6a9* are also present in a blood signature for brain injury [42], while *Dbh, Itgb1, Ltc4s,* and *Rhoa* were reported to be relevant to mTBI [43]. Many of the other genes have been associated with various mental illnesses and neurodegenerative diseases, including Schizophrenia, Alzheimer's disease, and sleep disorder. Detailed associations and references are listed in Additional file 4: Table S5.

Brain-brain network

In a similar manner as before, we first used COMBINER to infer the top 100 candidate modules for each brain region. We then identified common modules for each remaining brain region separately, removing features using Consensus Feature Elimination until the average AUC of the second region exceeded 0.75. Table 4 lists the final number of modules identified from each brain region pair, as well as the number of "core" modules and "core" genes for each brain tissue (i.e. those present in the majority of pair wise comparisons). In total, 37 core modules with 177 genes were identified in the brain-brain network.

We list the final number of modules identified from each brain region pair, as well as the overall numbers of core modules and core genes for each region. Figure 3a displays the tissue- and time-specific expression patterns of each brain-brain core module. Figure 3b shows the expression levels of the genes in each module, as well as the known PPIs occurring between genes. Unlike the blood-brain network, the shape of a gene represents the brain region in which it was inferred. Table 5 provides the putative biological functions of the core modules as inferred, while detailed module information is summarized in Additional file 5: Table S4.

In the brain-brain core module network, Modules 6, 8, 33, and 15 are of particular interest. An active Module 6 (*Creb3l2, Prkx, Avp*) in the hippocampus indicates a down regulated PKA-CREB long term potentiation pathway, which has been shown to impair memory [44]. In addition, the activity of Module 8 (*Prka1b, Hspa1a, Nfkbia, Jun, Cpt1b*) in the septal region shows down regulation of a heat shock protein (HSPA1A). Such activity has previously been found in other PTSD studies [45]. Module 33 depicts an up regulated dopamine pathway in the ventral striatum. This activity could potentially send excessive dopamine to the amygdala and other brain regions, which has been shown to lead to increased anxiety [46,47]. Finally, Module 15 implies an active pro-inflammatory response in the medial prefrontal cortex (MPFC) that agrees with the study in [48]. Other validated findings include olfactory impairment in the stria terminalis (ST) (module 32) [49]; alteration of complement pathways in the MPFC (module 20) [50] and activated coagulation function in the ST (module 31) [51].

The above findings highlight that while the putative biological functions of the brain-brain core modules

Table 3 Identified final modules between blood and brain regions with the associated mean AUC and mean error rate of both tissues

Validation	HB	AY	HC	ST	MPFC	SE	VS	Core module	Core gene
Inference									
Blood	92	20	23	27	22	2	19	9	43
Final gene	171	52	51	70	72	7	44		
Mean AUC	0.73	0.85	0.79	0.86	0.76	0.82	0.84		
Mean error rate	0.24	0.24	0.29	0.22	0.31	0.27	0.24		

The mean AUC and error rate were calculated by 500 LDA classifiers with random sampling on both tissues. (*HB* hemibrain (hemisphere), *AY* amygdala, *HS* hippocampus, *MPFC* medial prefrontal cortex, *VS* ventral striatum, *SE* septal region and *ST* stria terminalis).

Table 4 Numbers of COMBINER modules identified using data from six brain regions

Val Inf	AY	HC	ST	MPFC	SE	VS	Core module	Core gene
AY	/	17	25	23	0	16	1	4
HC	17	/	7	25	7	18	6	28
ST	22	29	/	23	7	22	9	45
MPFC	22	31	22	/	5	22	9	41
SE	18	26	25	17	/	28	10	53
VS	22	30	13	5	5	/	2	6

(*HB* Hemibrain (Hemisphere), *AY* amygdala, *HS* hippocampus, *MPFC* medial prefrontal cortex, *VS* ventral striatum, *SE* septal region and *ST* stria terminalis).

largely encompass the DEPATHs identified in the statistical overlap analysis (Tables 1 and 2), the COMBINER network-based analysis provides a much richer molecular description of mouse responses to social defeat. With additional validation in human studies, we expect these findings to yield robust prognostic and diagnostic biomarkers for PTSD.

Conclusions

The identification of diagnostic and prognostic blood biomarkers for PTSD currently is of great interest. In this work, we have improved the COMBINER method—a computational framework for identifying gene expres-sion modules that are activated in common across ex-perimental conditions—and applied it to blood and brain data from a mouse social defeat model. The resulting gene networks highlight stress-related biological pro-cesses active in both brain and blood and provide a com-prehensive molecular description of social defeat. In total, our approach identified seven blood biomarker sets that have each been validated for classification perform-ance in one brain sub-region. Some of the genes and processes discovered are consistent with previous inde-pendent studies of PTSD or other mental illnesses, while others represent novel candidate PTSD biomarkers. We note that the same approach can be readily applied to other disease models to construct gene networks that are activated in common across tissues; future work will focus on this task.

Methods

Blood, organ and tissue collection

Terminal organs, brain regions, and blood samples from subject and control C57BL/6 mice were collected after 24 hours, or 6 weeks (42 days) post 10-day social stress, and 24 hours or 1.5 weeks post 5-day social stress. Brains of C57BL/6 mice were carefully removed from the skulls, and left or right hemi-brain from each defeated or control mouse was dissected into different anatomical and functional regions: Hemibrain (Hemi-sphere) (HB), amygdala (AY), hippocampus (HS), medial

Table 5 The putative biological functions of the core modules in brain-brain network

Pathway	Module	Pathway	Module
Steroid	1	Complement and coagulation	20
Proteolysis	2	Neurotrophin signaling	21
Notch signaling	3	Regulation of beta cells	22
Cell adhesion	4	Transmembrane transport	23
NOD like receptor	5	Myogenesis	24
Vasopressin	6	G alpha 13 protein	25
G alpha Q protein	7	HIVNEF pathway	26
PRARA	8	Steroid	27
Linoleic acid	9	Purine	28
Chemokine signaling	10	Oocyte meiosis	29
Neuroactive ligand receptor	11	Focal adhesion	30
Muscle contraction	12	Complement and Coagulation	31
Systemic lupus erythematosus	13	Olfactory transductoino	32
DNA Repair	14	Class A1 rhodopsin like receptors	33
IRS Related events	15	Host interaction of HIV factors	34
Post translational protein modification	16	Peptide ligand biding receptors	35
Arachidonic acid	17	Calcium signaling	36
ABC Transporters	18	Downstream TCR signaling	37
Phosphadylinositol signaling	19		

prefrontal cortex (MPFC), ventral striatum (VS), septal region (SE) and stria terminalis (ST). The number of defeated and control mice in different regions and conditions are summarized in Table 6.

RNA isolation and quality assessment

Total RNA was isolated according to the Trizol® method (Invitrogen Inc., Grand Island, NY) from homogenized whole blood and brain regions. RNA from blood was isolated using the PreAnalytiX PAXgene® blood RNA kit (Qiagen Inc., Valencia, CA). We collected the seven organ tissues from 5-6 control and defeat mice, respectively. We evaluated RNA integrity using the Agilent Bioanalyzer and excluded samples of low quality, which appears to either have low total RNA, or low ribosomal RNA (rRNA) mass ratio between 28S and 18S rRNA and high amount of non-ribosomal RNAs in the electropherograms.

Microarray hybridization

Microarray assays were performed using Agilent's genome wide mouse expression array (GE 4x44K v2 two color microarray) slides and kits (Agilent Technologies Inc., Santa Clara, CA) following the manufacturer's protocol. To minimize batch effects, each sample was hybridized with a universal common reference that was used for all experiments. Hybridized microarray slides were scanned using Agilent Technologies Scanner G2505C US09493743.

Microarray data processing

Genespring with feature extraction 10.x (Agilent, CA) was used to process all two-color chips. Log2 transformation, Lowess normalization, and quantile normalization were applied to normalize within and between microarrays. For the latter, we applied quantile normalization separately on data from each tissue. Outlier spots were converted to missing values. If more than half of the expression values of a probe were missing, we removed the probe from consideration. We then imputed missing values using the k-nearest neighbor imputation method. To avoid incurring a bias in favor of genes represented by a greater number of probes, we aggregated multiple probes from the same Entrez Gene together by computing the mean of the "sibling" probes. We have deposited all microarray data for this study at the Gene Expression Omnibus (GEO):

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45035.

Linear model

We used a linear model-based approach to deconvolve the experimental time effects from the social defeat expression data. Assuming log-additive effects, our method estimates the contributions of each of the four experimental time points and subtracts them away from the remaining effect of social defeat. The linear model we used to deconvolve the experimental time effect is defined as follows:

$$
\begin{bmatrix} D_1 \\ C_1 \\ D_2 \\ C_2 \\ D_3 \\ C_3 \\ D_4 \\ C_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{defeat} \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \tag{1}
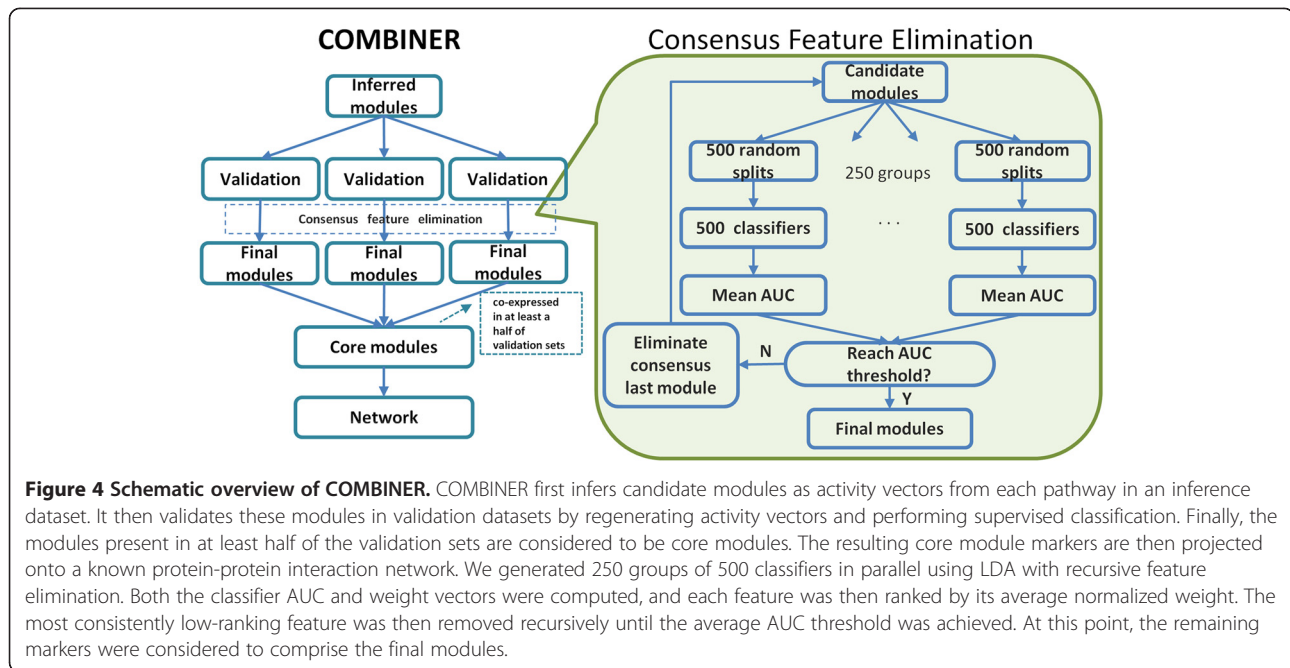$$

where D_i and C_i $i = 1, ..., 4$ denote $\log_2$ gene expression values of defeated and control mice in condition i, α_{defeat} denotes the overall effect of social defeat and $\beta_1, ..., \beta_4$ are the undesired time effects. In practice, we solve the above over determined system for each gene separately using least squares (implemented in the limma package), carrying forward only the gene-specific defeat effect for subsequent analyses.

Differential expression analysis

As described above in Section 2, we used the R/Bioconductor limma package and iterative Group Analysis (iGA) method for differentially expressed gene and GO term/pathway identification, respectively.

COMBINER

As shown in Figure 4, the COMBINER method first infers the statistically discriminative modules from an inference dataset, then validates them in various validation sets using consensus feature elimination. If a validated final module is co-expressed in at least half of the validation sets, then it is defined as a core module. Finally, we project these core modules onto known PPI

Table 6 Defeated and control mice (in the form of (number of defeated) / (number of control)) in different regions and conditions

Condition Tissue	5D-1D	5D-10D	10D-1D	10D42D
Blood	(5) / (5)	(5) / (5)	(5) / (4)	(5) / (5)
HB (Hemibrain)	5) / (5)	(5) / (5)	(5) / (5)	(5) / (5)
AY (Amygdala)	(2) / (3)	(4) / (5)	(4) / (3)	(4) / (3)
HC (Hippocampus)	(4) / (3)	(6) / (4)	(5) / (6)	(5) / (5)
MPFC (Medial Prefrontal Cortex)	(5) / (4)	(5) / (5)	(4) / (3)	(4) / (4)
SE (Septal Region)	(2) / (3)	(3) / (2)	(3) / (3)	(3) / (3)
ST (Stria Terminalis)	(5) / (5)	(5) / (5)	(4) / (4)	(5) / (4)
VS (Ventral Striatum)	(5) / (5)	(5) / (5)	(4) / (4)	(4) / (5)

(5D-1D/10D: 5 day treatment, 1 day/10 day recovery, 10D - 1D/ 6 W: 10 day treatment, 1 day/6 week recovery).

Figure 4 Schematic overview of COMBINER. COMBINER first infers candidate modules as activity vectors from each pathway in an inference dataset. It then validates these modules in validation datasets by regenerating activity vectors and performing supervised classification. Finally, the modules present in at least half of the validation sets are considered to be core modules. The resulting core module markers are then projected onto a known protein-protein interaction network. We generated 250 groups of 500 classifiers in parallel using LDA with recursive feature elimination. Both the classifier AUC and weight vectors were computed, and each feature was then ranked by its average normalized weight. The most consistently low-ranking feature was then removed recursively until the average AUC threshold was achieved. At this point, the remaining markers were considered to comprise the final modules.

networks. To remove features, we generated 250 groups of 500 classifiers in parallel and applied Linear Discriminant Analysis (LDA) with recursive feature elimination [52] to each to compute AUCs as well as weight vectors. Each feature was then ranked by its average normalized weight. The most consistently low-ranking feature was then removed recursively until the average AUC threshold of 0.75 was achieved. At this point, the remaining features were considered to comprise the final modules.

In our previous work [28], we applied both the Condition Responsive Genes (CORG) [53] and Core Module Inference (CMI) [28] methods to infer candidate modules and express them as pathway activities (PAs). In the greedy search process, CORG picks up either up- or down-regulated genes, while CMI identifies genes of both directions together. However, because of the multiple-time-point nature of the social defeat data, the application of the CMI method is not straightforward. Thus, in this work we used only the CORG method with the procedure described as follows. For a given pathway, we first rank the standardized gene expressions by their limma moderated t-score. If up-regulated genes are dominant, we rank the t-score in descending order; otherwise, ascending order is chosen. Next, we aggregate the first two genes using the formula $y = (x_1 + x_2)/\sqrt{2}$; if the expression of this aggregate yields a larger absolute t-score than the first gene, this combination is retained as a module with the combined expression becoming the PA. Otherwise, the procedure further adds the third gene using $y = (x_1 + x_2 + x_3)/\sqrt{3}$, and so

on, until the module-size limit, 25 genes, is reached. Finally, we ranked all modules using the absolute value of the pathway activity t-score.

We faced two major challenges when modifying our COMBINER method. First, the multiple-time-point nature of the data initially decreased the binary classification performance of the static LDA classifier [52]. Second, the small data sample size leads to a large variability of feature ranks after recursive feature elimination. To cope with the first challenge, we used a linear model to deconvolve the time effects from the original expression values. We solved the second problem by improving our method for consensus feature elimination. We generated 250 groups of 500 classifiers in parallel, then removed the bottom feature using the voting principle. In general, using additional groups of classifiers will further improve the reproducibility of the final modules. In our experience 250 groups were sufficient to yield a reproducible result (results not shown). Finally, we used a fixed average AUC threshold to determine the final modules instead of the max average AUC threshold described in [28]. This was required since the inference and validation sets can be very dissimilar, which leads to low values of the max average AUC.

We obtained pathway information from the MSigDB v3.0 Canonical Pathways subset [54]. To decrease redundancy, we applied pathway filtering to remove bulky pathways. This resulted in a pathway dataset containing 791 pathways with 5,633 genes assayed in all regions. The protein-protein interaction information was obtained from String v9.0 [55].

Average time curve

Let y_{ij} be the relative expression level of gene i at the sampling time j. The time-point expression patterns were modelled as follows,

$$y_{ij} = \mu_i(t_j) + \varepsilon_{ij} \qquad (2)$$

where $\mu_i(t_j) = \log_2\left(\left|\bar{x}^D(t_j)/\bar{x}^C(t_j)\right|\right)$ is the population average time curve for gene i evaluated at time t_j and where ε_{ij} is the random deviation from this curve. $\bar{x}^D(t_j)$ and $\bar{x}^C(t_j)$ are average expressions of disease and control mice respectively for gene i at time t_j.

Software

COMBINER was implemented in Matlab R2010a with Bioinformatics toolbox v3.5 (Math Works Inc., Natick, MA), statconn (http://www.statconn.com/), LinkR (http://www.mathworks.com/matlabcentral/fileexchange/5051), and R [56]. The source code can be found in Additional files 6 and 7.

Availability of supporting data

The supporting data is provided in the supporting information.

Additional files

Additional file 1: Table S1. Significantly overlapping DEPATHs between Blood and Hemibrain.

Additional file 2: Table S2. Significantly overlapping DEGOs between Blood and Septal Region.

Additional file 3: Table S3. List of blood-brain core modules with associated genes.

Additional file 4: Table S5. Biological Significance of blood core gene markers from literature mining.

Additional file 5: Table S4. List of brain-brain core modules with associated genes.

Additional file 6: Processed microarray data used in the paper.

Additional file 7: Source code.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RY, BJD, SYM, RH, MJ, LP, and FJD conceived and designed the research. BJD performed the overlap analysis, RY developed the COMBINER computations. SYM, RH, and MJ provided the experimental data. RY and BJD wrote the paper. All authors read and approved the final manuscript.

Disclaimers

The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as official Department of the Army position, policy, or decision, unless so designated by other official documentation. Research was conducted in compliance with the Animal Welfare Act, and other Federal statutes and regulations relating to animals and experiments involving animals and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals (NRC 2011) in facilities that are fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International. Citations of commercial organizations or trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

Author details

[1]Institute for Collaborative Biotechnologies, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA. [2]Department of Computer Science, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA. [3]Department of Mechanical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA. [4]Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106-5080, USA. [5]US Army Center for Environmental Health Research, Fort Detrick, MD, 21702-5010, USA. [6]Advanced Biomedical Computing Center, Frederick National Laboratory for Cancer Research, Frederick, MD 21702-5010, USA.

References

1. Seal KH, Metzler TJ, Gima KS, Bertenthal D, Maguen S, Marmar CR: **Trends and risk factors for mental health diagnoses among Iraq and Afghanistan veterans using department of veterans' affairs health care, 2002–2008.** *Am J Public Health* 2009, **99**(9):1651–1658.
2. Skelton K, Ressler KJ, Norrholm SD, Jovanovic T, Bradley-Davino B: **PTSD and gene variants: New pathways and new thinking.** *Neuropharmacology* 2012, **62**(2):628–637.
3. Broekman BFP, Olff M, Boer F: **The genetic background to PTSD.** *Neurosci Biobehav Rev* 2007, **31**(3):348–362.
4. Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R, Horvath S: **Is human blood a good surrogate for brain tissue in transcriptional studies?** *BMC Genomics* 2010, **11**(1):589.
5. Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ: **Gene expression profiling in human neurodegenerative disease.** *Nat Rev Neurol* 2012, **8**:518–530.
6. Sharp FR, Xu H, Lit L, Walker W, Apperson M, Gilbert DL, Glauser TA, Wong B, Hershey AD, Liu D, *et al*: **The future of genomic profiling of neurological diseases using blood.** *Arch Neurol* 2006, **63**(11):1529–1536.
7. Tang Y, Lu A, Aronow BJ, Sharp FR: **Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: Blood genomic fingerprints of disease.** *Ann Neurol* 2001, **50**(6):699–707.
8. Segman RH, Shefi N, Goltser-Dubner T, Friedman N, Kaminski N, Shalev AY: **Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors.** *Mol Psychiatry* 2005, **10**(5):500–513.
9. Rollins B, Martin MV, Morgan L, Vawter MP: **Analysis of whole genome biomarker expression in blood and brain.** *Am J Med Genet B* 2010, **153B**(4):919–936.
10. Zieker J, Zieker D, Jatzko A, Dietzsch J, Nieselt K, Schmitt A, Bertsch T, Fassbender K, Spanagel R, Northoff H, *et al*: **Differential gene expression in peripheral blood of patients suffering from post-traumatic stress disorder.** *Mol Psychiatry* 2007, **12**(2):116–118.
11. Hammamieh R, Chakraborty N, De Lima TCM, Meyerhoff J, Gautam A, Muhie S, D'Arpa P, Lumley L, Carroll E, Jett M: **Murine model of repeated exposures to conspecific trained aggressors simulates features of post-traumatic stress disorder.** *Behav Brain Res* 2012, **235**(1):55–66.
12. Berton O, McClung CA, DiLeone RJ, Krishnan V, Renthal W, Russo SJ, Graham D, Tsankova NM, Bolanos CA, Rios M, *et al*: **Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress.** *Science* 2006, **311**(5762):864–868.
13. Krishnan V, Han MH, Graham DL, Berton O, Renthal W, Russo SJ, LaPlant Q, Graham A, Lutter M, Lagace DC, *et al*: **Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions.** *Cell* 2007, **131**(2):391–404.
14. Chuang JC, Cui H, Mason BL, Mahgoub M, Bookout AL, Yu HG, Perello M, Elmquist JK, Repa JJ, Zigman JM, *et al*: **Chronic social defeat stress disrupts regulation of lipid synthesis.** *J Lipid Res* 2010, **51**(6):1344–1353.

15. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** *Nat Rev Genet* 2011, **12**(1):56–68.
16. Meunier D, Lambiotte R, Fornito A, Ersche K, Bullmore ET: **Hierarchical modularity in human brain functional networks.** *Front Neuroinf* 2009, **3**:37.
17. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302**(5643):249–255.
18. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol* 2005, **4**:17.
19. Yamaguchi R, Yoshida R, Imoto S, Higuchi T, Miyano S: **Finding module-based gene networks with state-space models - Mining high-dimensional and short time-course gene expression data.** *IEEE Signal Proc Mag* 2007, **24**(1):37–46.
20. Sameith K, Antczak P, Marston E, Turan N, Maier D, Stankovic T, Falciani F: **Functional modules integrating essential cellular functions are predictive of the response of leukaemia cells to DNA damage.** *Bioinformatics* 2008, **24**(22):2602–2607.
21. Keller MP, Choi Y, Wang P, Belt Davis D, Rabaglia ME, Oler AT, Stapleton DS, Argmann C, Schueler KL, Edwards S, et al: **A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility.** *Genome Res* 2008, **18**(5):706–716.
22. De Smet R, Marchal K: **Advantages and limitations of current network inference methods.** *Nat Rev Micro* 2010, **8**(10):717–729.
23. Xia K, Xue H, Dong D, Zhu S, Wang J, Zhang Q, Hou L, Chen H, Tao R, Huang Z, et al: **Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms.** *PLoS Comput Biol* 2006, **2**(11):e145.
24. Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, Hou L, Zhang Q, Zhang Y, Han JDJ: **A modular network model of aging.** *Mol Syst Biol* 2007, **3**:147.
25. Langfelder P, Luo R, Oldham MC, Horvath S: **Is my network module preserved and reproducible?** *PLoS Comput Biol* 2011, **7**(1):e1001057.
26. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
27. Edelman EJ, Guinney J, Chi JT, Febbo PG, Mukherjee S: **Modeling cancer progression via pathway dependencies.** *PLoS Comput Biol* 2008, **4**(2):e28.
28. Yang R, Daigle B, Petzold L, Doyle F: **Core module biomarker identification with network exploration for breast cancer metastasis.** *BMC Bioinformatics* 2012, **13**(1):12.
29. Miller JA, Horvath S, Geschwind DH: **Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways.** *PNAS* 2010, **107**(28):12698–12703.
30. Wuchty S, Oltvai ZN, Barabasi AL: **Evolutionary conservation of motif constituents in the yeast protein interaction network.** *Nat Genet* 2003, **35**(2):176–179.
31. Smyth G: **limma: Linear models for microarray data.** In *Bioinformatics and computational biology solutions Using R and Bio conductor.* Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. New York: Springer; 2005:397–420.
32. Breitling R, Amtmann A, Herzyk P: **Iterative Group Analysis (iGA): A simple tool to enhance sensitivity and facilitate interpretation of microarray experiments.** *BMC Bioinforma* 2004, **5**(1):34.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *PNAS* 2005, **102**(43):15545–15550.
34. Begley DJ: **ABC transporters and the blood-brain barrier.** *Curr Pharm Design* 2004, **10**(12):1295–1312.
35. Förster C: **Tight junctions and the modulation of barrier function in disease.** *Histochem Cell Biol* 2008, **130**(1):55–70.
36. Giacomini KM, YS: **Membrane transporters and drug response.** In *Goodman & Gilman's the pharmacological basis of therapeutics.* Edited by Brunton LL LJ, Parker KL. New York: McGraw-Hill; 2006:41–70.
37. Alford PW, Dabiri BE, Goss JA, Hemphill MA, Brigham MD, Parker KK: **Blast-induced phenotypic switching in cerebral vasospasm.** *PNAS* 2011, **108**(31):12705–12710.
38. Corcoran JPT, So PL, Maden M: **Disruption of the retinoid signalling pathway causes a deposition of amyloid β in the adult rat brain.** *Euro J Neurosci* 2004, **20**(4):896–902.
39. Pajer K, Andrus BM, Gardner W, Lourie A, Strange B, Campo J, Bridge J, Blizinsky K, Dennis K, Vedell P, et al: **Discovery of blood transcriptomic markers for depression in animal models and pilot validation in subjects with early-onset major depression.** *Transl Psychiatry* 2012, **2**:e101.
40. Yi Z, Li Z, Yu S, Yuan C, Hong W, Wang Z, Cui J, Shi T, Fang Y: **Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder.** *PLoS ONE* 2012, **7**(2):e31283.
41. Beech RD, Lowthert L, Leffert JJ, Mason PN, Taylor MM, Umlauf S, Lin A, Lee JY, Maloney K, Muralidharan A, et al: **Increased peripheral blood expression of electron transport chain genes in bipolar depression.** *Bipolar Disord* 2010, **12**(8):813–824.
42. Pathan N, Hemingway CA, Alizadeh AA, Stephens AC, Boldrick JC, Oragui EE, McCabe C, Welch SB, Whitney A, O'Gara P, et al: **Role of interleukin 6 in myocardial dysfunction of meningococcal septic shock.** *Lancet* 2004, **363**(9404):203–209.
43. Stevens SL, Leung PY, Vartanian KB, Gopalan B, Yang T, Simon RP, Stenzel-Poore MP: **Multiple preconditioning paradigms converge on interferon regulatory factor-dependent signaling to promote tolerance to ischemic brain injury.** *J Neurosci* 2011, **31**(23):8456–8463.
44. Silva AJ, Kogan JH, Frankland PW, Kida S: **CREB and memory.** *Annu Rev Neurosci* 1998, **21**(1):127–148.
45. Sriram K, Rodriguez-Fernandez M, Doyle FJ III: **A detailed modular analysis of heat-shock protein dynamics under acute and chronic stress and its implication in anxiety disorders.** *PLoS ONE* 2012, **7**(8):e42958.
46. Pezze MA, Feldon J: **Mesolimbic dopaminergic pathways in fear conditioning.** *Prog Neurobiol* 2004, **74**(5):301–320.
47. Yang R, Sriram K, Doyle FJ: **Control circuitry for fear conditioning associated with post-traumatic stress disorder (PTSD).** In *IEEE Conference on Decision and Control (CDC): 15-17 Dec. 2010.* Atlanta, USA: Institute of Electrical and Electronics Engineers Inc.; 2010:2541–2546.
48. Kubera M, Obuchowicz E, Goehler L, Brzeszcz J, Maes M: **In animal models, psychosocial stress-induced (neuro) inflammation, apoptosis and reduced neurogenesis are associated to the onset of depression.** *Prog Neuropsychopharmacol Biol Psych* 2011, **35**(3):744–759.
49. Vasterling JJ, Brailey K, Sutker PB: **Olfactory identification in combat-related posttraumatic stress disorder.** *J Trauma Stress* 2000, **13**(2):241–253.
50. Hovhannisyan L, Mkrtchyan G, Sukiasian S, Boyajyan A: **Alterations in the complement cascade in post-traumatic stress disorder.** *Allergy Asthma Clin Immunol* 2010, **6**(1):3.
51. Robicsek O, Makhoul B, Klein E, Brenner B, Sarig G: **Hyper coagulation in chronic post-traumatic stress disorder.** *Isr Med Assoc J* 2011, **13**:548–552.
52. Friedman JH: **Regularized discriminant analysis.** *J Am Stat Assoc* 1989, **84**(405):165–175.
53. Lee E, Chuang HY, Kim JW, Ideker T, Lee D: **Inferring pathway activity toward precise disease classification.** *PLoS Comput Biol* 2008, **4**(11):e1000217.
54. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP: **Molecular signatures database (MSigDB) 3.0.** *Bioinformatics* 2011, **27**(12):1739–1740.
55. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, et al: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**(suppl 1):D561–D568.
56. R Development Core Team: *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2005.