Post-mortem molecular profiling of three psychiatric disorders reveals widespread dysregulation of cell-type associated transcripts and refined disease-related transcription changes

Kevin M. Bowling[1,*], Ph.D., Ryne C. Ramaker[1,2,*], B.S., Brittany N. Lasseigne[1*], Ph.D., Megan H. Hagenauer[3], Ph.D., Andrew A. Hardigan[1,2], Nick S. Davis[1,#], B.S., Jason Gertz[1,%], Ph.D., Preston M. Cartagena[4], Psy.D., David M. Walsh[4], Psy.D., Marquis P. Vawter[4], Ph.D., Edward G. Jones (deceased), M.D., Ph.D., Alan F. Schatzberg[5], M.D., Jack D. Barchas[6], M.D., Ph.D., Stan J. Watson[3], M.D., Ph.D., Blynn G. Bunney[4], Ph.D., Huda Akil[3], Ph.D., William E. Bunney[4], M.D., Jun Z. Li[7], Ph.D., Sara J. Cooper[1], Ph.D., and Richard M. Myers[1], Ph.D.

*These authors contributed equally to this study

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

[2]Department of Genetics, The University of Alabama at Birmingham, Birmingham, AL, USA

[3]Mental Health Research Institute, University of Michigan, Ann Arbor, MI, USA

[4]Department of Psychiatry and Human Behavior, College of Medicine, University of California, Irvine, Irvine, CA, USA

[5]Department of Psychiatry, Stanford University School of Medicine, Stanford, CA, USA

[6]Psychiatry, Weill Cornell Medical College, New York, NY, USA

[7]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

Present address:

[#]Duke University, Durham, NC, USA

[%]University of Utah School of Medicine, Salt Lake City, UT, USA

To whom correspondence should be addressed:

Richard M. Myers, Ph.D.

HudsonAlpha Institute for Biotechnology

601 Genome Way

Huntsville, AL 35806

Telephone: 256-327-0431

FAX: 256-327-0978

rmyers@hudsonalpha.org

Abstract

Background

Psychiatric disorders are multigenic diseases with complex etiology contributing significantly to human morbidity and mortality. Although clinically distinct, several disorders share many symptoms suggesting common underlying molecular changes exist that may implicate important regulators of pathogenesis and new therapeutic targets.

Results

We compared molecular signatures across brain regions and disorders in the transcriptomes of postmortem human brain samples. RNA sequencing was performed on tissue from the anterior cingulate cortex, dorsolateral prefrontal cortex, and nucleus accumbens from three groups of 24 patients each diagnosed with schizophrenia, bipolar disorder, or major depressive disorder, and from 24 control subjects and result were validated in the independent cohort. The most significant disease differences were observed in the anterior cingulate cortex of schizophrenia samples compared to controls and biochemical consequences of gene expression changes were assessed with untargeted metabolomic profiling. We detected significant transcriptional heterogeneity within schizophrenia and bipolar disorder samples and find no evidence this heterogeneity is linked to any known clinical phenotypes or technical variables, however it does appear to be highly correlated with transcripts previously identified as specific to brain cell types. We also present evidence that altered expression of the transcription factor *NPAS4* in schizophrenia patients is robust to the effects of cell-type heterogeneity

and that *NPAS4* regulates the expression of several genes implicated by previous schizophrenia GWAS analyses.

Conclusions

We provide a set of high confidence, independently validated genes differentially expressed between schizophrenia and control patients and propose a refined list of putative cell-type independent transcription changes, including *NPAS4*, which may particularly warrant further investigation.

Keywords

Schizophrenia, Bipolar Disorder, Major Depression Disorder, RNA sequencing, metabolomics, *NPAS4*

Background

Schizophrenia (SZ), bipolar disorder (BPD), and major depression disorder (MDD) are multigenic diseases with complex etiology and are large sources of morbidity and mortality in the population. All three disorders are associated with high rates of suicide, with ~90% of the ~41,000 people who commit suicide each year in the U.S. having a diagnosable psychiatric disorder [1]. Notably, while clinically distinct, these disorders also share many symptoms, including psychosis, suicidal ideation, sleep disturbances and cognitive deficits [2]. This phenotypic overlap suggests potential common genetic etiology, which is supported by recent large-scale genome-wide association studies [3–6]. However, this overlap has not been fully characterized with functional genomic approaches. Current therapies for these psychiatric disorders are ineffective in many patients and often only treat a subset of an individual patient's symptoms [7].

Approaches targeting the underlying molecular pathologies within and across these types of disorders are necessary to address the enormous psychiatric component of global disease burden and improve care for the millions of people diagnosed with these conditions.

Previous studies [8–10] analyzed brain tissue with RNA sequencing (RNA-seq) in SZ and BPD, and identified altered expression of GABA-related genes in the superior temporal gyrus and hippocampus, as well as differentially expressed transcripts related to neuroplasticity and mammalian circadian rhythms. Our study focused on the anterior cingulate cortex (AnCg), dorsolateral prefrontal cortex (DLPFC), and nucleus accumbens (nAcc) regions, which are often associated with mood alterations, cognition, impulse control, motivation, reward, and pleasure – all behaviors known to be altered in psychiatric disorders [11,12]. To assess gene expression changes associated with psychiatric disease in these brain regions, we performed RNA-seq on macro-dissected postmortem tissues in four well-documented cohorts of 24 patients each with SZ, BPD, MDD and 24 controls (CTL) (96 individuals total) across these three brain regions. Additionally, we conducted metabolomic profiling in AnCg tissue from the same subjects. RNA-seq analysis revealed common expression profiles in SZ and BPD patients supporting the notion that these disorders share a common molecular etiology. Interestingly, SZ and BPD samples contained a high degree of heterogeneity that appears to be correlated with differences in cell-type composition. We validated this result with RNA-seq data from an independent cohort of 35 cases each of SZ, BPD, and CTL postmortem cingulate cortex samples from the Stanley Neuropathology Consortium Integrative Database (SNCID; http://sncid.stanleyresearch.org) Array

Collection. We present a set of validated genes differentially expressed between SZ and CTL patients, propose a short list of priority gene targets identified by a novel computational approach, and investigate the downstream regulatory network of the most robust gene target, *NPAS4*.

Results

Region-specific gene expression in control and psychiatric brain tissue

We collected postmortem human brain tissue and associated clinical data, including age, sex, brain pH, and postmortem interval (PMI) (Table S1) for matched cohorts of 24 patients each diagnosed with SZ, BPD, or MDD, as well as 24 control individuals with no personal history of, or first-degree relatives diagnosed with, psychiatric disorders. Importantly, we included only patients with an agonal factor score of zero and a minimum brain pH of 6.5 to minimize expression differences due to those confounders [13]. Using RNA-seq [14], we profiled gene expression in three macro-dissected brain regions (AnCg, DLPFC, nAcc). After quality control, we analyzed 48,956 ENSEMBL transcripts in a total of 281 brain samples (Table S2).

To examine heterogeneity across brain regions and subjects, we performed a principal component analysis (PCA; Figure S1A) of all transcripts. The first principal component (PC1, 26% of the variation) separates cortical AnCg and DLPFC samples from subcortical nAcc samples. Examination of the second principal component (PC2) reveals a separation of some SZ and BPD samples from all other samples (Figure S1B). Subsequent principal components did not associate with clinical covariates (Table S3).

Disease-specific gene expression in control and psychiatric brains

We next sought to identify whether transcript profiles could differentiate patient tissues from controls. PCA within each brain region revealed some separation of SZ and BPD samples from CTL and MDD patients (Figure 1A-C). We applied DESeq2 [15], a method for differential analysis of sequence read count data, to identify genes differentially expressed across disorders within each brain region. The largest number of significant expression changes occurred in AnCg between SZ and CTL individuals (1,203 transcripts, FDR<0.05). The most striking transcriptional changes were also confirmed by microarray analysis performed in a subset of patients (r^2=0.71; Figure S2). Pathway enrichment analysis of differentially expressed genes between SZ and CTL patients revealed 610 GO terms with an FDR<0.05 (Table S4). Significant GO terms fall into the broad categories of: neuronal structure, neuronal function (e.g., neurotransmitter transport, synaptic transmission, behavior), immune response, and energy metabolism (Figure S3). Additionally, seven transcripts were differentially expressed between SZ and CTL individuals in DLPFC. Three of these were also identified in AnCg: *SST*, *KLHL14* and *SERPINA3*. No transcripts had an FDR<0.05 when comparing BPD or MDD samples to CTLs in any brain region, or comparing SZ and CTL tissues in nAcc (Table S5). To examine potential common gene expression patterns between the psychiatric disorders, we performed pair-wise correlation calculations of all transcript $\log_2$ fold changes for each disorder versus controls in each brain region. Of the nine case-control comparisons (for three regions and three diseases), a particularly strong correlation is observed between BPD and SZ compared to either SZ or BPD and MDD in each brain region (Figure 2A). In the AnCg, BPD and

SZ share 2,606 common genes differentially expressed at a relaxed threshold of uncorrected p-value<0.05 compared to only 248 and 85 genes shared between MDD and SZ or BPD respectively (Figure 2B). This strong overlap between BPD and SZ (Fisher's exact p-value<0.001) indicates that although expression changes are weaker in BPD they are similar to those involved in SZ.

Because previous postmortem analyses have been limited by, and are particularly vulnerable to, biases inherent to examining a single patient cohort, we sought to generate a robust set of SZ associated transcripts by validating our observed expression changes in an independent cohort. To accomplish this, we examined gene expression differences between SZ and CTL samples in the SNCID RNA-seq validation dataset, revealing 129 high confidence genes differentially regulated in the same direction in both datasets with an FDR<0.05 (Table S6). The magnitude and direction of change in significant transcripts in the Pritzker dataset were highly correlated with SNCID dataset (r^2=0.66), particularly in transcripts that met an FDR<0.05 cutoff (r^2=0.89; Figure 2C). We performed hierarchical clustering of SZ and CTL samples using the 129 transcripts differentially expressed between SZ and CTL in both datasets (Figures 2D, 2E), and two clusters emerged. In the Pritzker dataset, one consists of only SZ (100% SZ) and the other contains both CTL and SZ (36.8% SZ), suggesting that a subset of SZ individuals could be driving observed gene expression changes between SZ and CTL samples.

In both datasets, one of the largest observed gene expression changes between SZ and CTL was *SERPINA3* (serpin peptidase inhibitor, clade A, member 3), an inflammation marker previously shown to be elevated in SZ in four postmortem cohorts,

including this SNCID dataset [16–19]. GABA pathway genes have also been repeatedly implicated in SZ and BPD [20–22], and we also observed significant expression changes indicating reduction of GABA levels, including decreased expression of *GAD1* and *GAD2* (enzymes catalyzing the rate-limiting step in GABA synthesis) and *SLC32A1* (the vesicular inhibitory amino acid transporter). Interestingly, *SLC32A1* was identified in a recent GWAS meta-analysis [3]. *EGR1*, *MDK*, and *CLU*, also identified in this GWAS analysis, were significant in both the Pritzker and the SNCID dataset.

Altered metabolite levels recapitulate transcriptomic changes

To confirm the biochemical consequences of expression changes, we used 2D-GCMS to measure metabolite levels in 86 of the AnCg samples (sufficient tissue was unavailable for 10 samples). This included 20 SZ (6 tissues from clustered subset), 22 BPD (8 tissues from clustered subset), 22 CTL and 22 MDD individuals. We measured and identified 100 unique metabolites (Table S7). Similar to our transcript analysis, metabolite levels (Table S8) successfully differentiated SZ and BPD patients from CTLs (Figures 3A-B), while MDD metabolite profiles were very similar to CTLs. Several of the most significant metabolites, including GABA, were also implicated by RNA-seq and are known to be relevant to BPD and SZ [23]. PCA analysis of metabolite data revealed the second principle component (11% of variance explained) separated SZ and BPD from CTL patients (Figure 3C, p-value < 0.001) and was correlated with the first principle component of our RNA-seq analysis (Figure 3D, r^2=0.16, p-value < 0.001). Furthermore, GABA/glutamate metabolite ratios also correlate strongly with average *GAD1* and *GAD2* expression levels (r^2 = 0.19, p-value < 0.001). This metabolite-gene relationship

is consistent with previous multi-level phenomic analyses [24] and demonstrates realized biochemical consequences from altered gene expression.

Heterogeneity within schizophrenia and bipolar disorder

Examination of RNA-seq-based PCA plots reveals a subset of SZ and BPD patients are intermixed and cluster separately from the remaining patients and controls in both cohorts and across platforms (Figure 1 and 4A; Figure S4A). This suggests these disorders may be particularly heterogeneous and is further supported by widespread increases in transcript level variance across SZ and BPD patients in both cohorts (Figure 4B-C). Transcript heterogeneity was highly concordant between cohorts as transcript loading values of the first principle components were strongly correlated (rho=0.58, p-value < 0.001, Figure 4D). To explore further this relationship and bin samples in an unbiased manner, we performed consensus clustering (an unsupervised method for establishing group membership) on cingulate cortex samples from both cohorts [25]. This analysis defined a subset of primarily BPD and SZ patients that display expression profiles distinct from the remaining samples and controls (Figure S4B-C). GO term enrichment of genes driving consensus clusters revealed 730 GO terms with FDR<0.05, including a large network of up-regulated immune signaling and down-regulated neuronal structure and signaling genes (Figure 3E; Table S9).

To identify drivers of this observed expression pattern, we first sought to eliminate any technical artifacts or clinical differences among patients that might lead to this result. This consensus cluster grouping of patients in the Pritzker cohort is not significantly associated with any known clinical covariates (age, sex, race, number of hospitalizations, blood alcohol level, age of onset, PMI), and the agonal factor was 0 for

all subjects (Table S10). Fisher's exact testing for major drug classes identified in postmortem cytotoxicology analysis also failed to associate drug use with either patient cluster (Table S11). Examination of medical records, including patient prescription history, symptoms, and severity, was also not linked to observed expression patterns. Previous studies [13] have shown a key role for brain pH in gene expression patterns, and we found brain pH to be decreased when comparing between BPD individuals across the two consensus cluster groups (p=0.02), but not between SZ individuals across groups (p=0.66; Figure S5A). However, pH was only a small contributor to heterogeneity in SZ and BPD and after normalizing to pH, the original sample distribution largely remains (Figure S5B). We also considered whether dissection differences could contribute to this result. Our data indicate that this is not the source of variability, as a majority of sample heterogeneity was confined to BPD and SZ in both cohorts, which were dissected by independent groups, and the BPD/SZ-associated expression patterns were observed across multiple brain regions.

If the identified clusters are not a technical artifact or explained by any available clinical variables, then, given the large number of genes affected and GO term enrichments observed, one possible explanation for these changes is that cell-type differences are driving the observed disease heterogeneity. Virtual microdissection, particularly in the use of tissue-based transcriptomic datasets, is a rapidly emerging field of investigation that has proven useful for identifying cell-type related and independent patterns in case-control studies [26]. Darmanis et al. recently published a survey of human brain transcriptome diversity in single cells [27]. Their results included a panel of transcripts capable of classifying individual cells into the major neuronal, glial, and vascular cell-

types in the brain. We generated cell-type indices using the median of normalized counts for each cell-type specific transcript set. Examining cell-type indices in a previously published RNA-seq analysis of purified brain cells reveals high specificity of each index to the appropriate cell type and accurate deconvolution of transcriptomes mixed *in silico* [28,29] (Figure S6A-F). Moreover, median values from 10,000 randomly sampled transcript sets are not able to deconvolute mixed cell transcriptomes, demonstrating that predictive power is relatively unique to the Darmanis et al. transcript sets (Figure S6G-I). Examining AnCg cell-type composition in the context of the patient clustering patterns, we found that the subset of BPD and SZ patients that cluster separately from CTLs is enriched for astrocyte- (p-value < 0.001) and endothelial-specific expression (p-value < 0.001), but depleted of neuron-specific expression (p-value < 0.001) (Figure 5A). The correlation between PC1 in the cingulate cortex and the neuron/astrocyte ratio is strong in both the Pritzker dataset (rho=0.41 p-value < 0.001) and the SNCID dataset (rho=0.64, p-value < 0.001, Figures 5B-C). Interestingly, normalizing potential cell-type differences in our previous SZ/CTL differential expression analysis by adding the neuron/astrocyte ratio as an additional covariate in our DESeq2 regression model reduces the number of significant transcripts from 1,203 to 3 and dramatically shifts the p-value distribution of our SZ case-control comparison in both datasets (Figure 5D-E). The remaining, putatively cell-type independent, significant transcripts are *IGKC*, *IGHG1*, and the transcriptional regulator *NPAS4*, which has been previously linked to schizophrenia-like symptoms [30] (Table S12).

NPAS4 regulates previously implicated SZ genes

We further investigated the potential relevance of *NPAS4*, whose decreased expression in SZ patients is the most significant transcriptional alteration after normalizing to cell-type heterogeneity, by performing replicate ChIP-seq experiments in the neuroblastoma cell line, IMR-32. We observed strong concordance between replicate experiments and the 2,832 high-confidence binding sites common to both replicates were enriched for derivatives of the previously reported canonical *NPAS4* motif [31,32] (Figure S7). Genes previously implicated in SZ by GWAS analyses were enriched for close proximity to observed *NPAS4* binding sites compared to the background transcriptome (p-value < 0.001) with transcription start sites of 13 transcripts found within 1 kb of an *NPAS4* binding site (Figure 6A, Table S13). To confirm the regulatory role of *NPAS4* for previously associated SZ genes we performed shRNA-mediated knock down of *NPAS4* and observed a concordant decrease in expression of three of five genes tested that contain a neighboring *NPAS4* binding site (Figure 6B, Table S13).

Discussion

Here, we have described a large transcriptomic dataset across three brain regions (DLPFC, AnCg, and nAcc) in SZ, BPD, and MDD patients, as well as CTL samples matched for agonal state and brain pH. In MDD, we do not identify any transcripts that meet genome-wide significance for differential expression between cases and controls in any brain region. This finding agrees with previous postmortem RNA-seq studies [33], however sample size and the choice of brain regions examined likely contributed to our inability to replicate results from previous non-sequencing-based approaches comparing MDD to CTL in postmortem brain [34].

The most dramatic gene expression signals were brain region-specific. The majority of disease-associated expression differences were seen in the AnCg of SZ compared to CTL patients. The AnCg has been associated with multiple disease-relevant functions, including cognition, error detection, conflict resolution, motivation, and modulation of emotion [35–37]. We observed a striking overlap in SZ- and BPD-associated expression changes consistent with previous findings [23,38]. We also observed significant heterogeneity within SZ and BPD samples with a subset of patients exhibiting gene expression profiles that differed significantly from both the larger subset of SZ and BPD patients and from the MDD and CTL cohorts. We find no evidence that this disease heterogeneity is linked to any known clinical phenotypes or technical variables, however it does appear to be highly correlated with transcripts previously identified as specific to brain cell types. Specifically, we see evidence for depletion of neuron-specific transcripts and increased levels of endothelial and astrocyte-specific transcripts in a subset of SZ and BPD patients. Insufficient tissue remains from our patient cohort to validate computational cell-type predictions immunohistochemically, however our data strongly suggests that future postmortem studies should be cognizant of cell-type heterogeneity across patient samples. We recommend future studies use robust techniques for assessing tissue composition to examine potential cell-type proportion differences between disease cohorts and to identify which transcriptional changes occur in conjunction with, and independent of, those differences. After normalizing for potential cell-type associated differences between samples a refined list of transcriptional changes remain that may merit further investigation. Of particular interest is the transcription factor *NPAS4* (14-fold decreased expression in SZ compared to

CTL), which is highly expressed in the brain, has been previously characterized as an important regulator of neuronal inhibitory pathways, and whose dysregulation may be implicated in psychiatric disease [39]. Intriguingly, *NPAS4* knockout mice display several characteristics common to SZ including sensorimotor gating deficits, hyperactivity and social anxiety, thus it may play an important role in driving relevant disease pathophysiology [40]. Our characterization of *NPAS4*'s genome-wide binding events and regulation of previously implicated schizophrenia genes further supports *NPAS4*'s potential importance as a driver of disease, however a comprehensive analysis of *NPAS4*'s gene regulatory network and the mechanisms by which *NPAS4* targets may contribute to patient phenotypes remains of great interest and an active area of investigation. Other genes significant after cell-type correction include *IGKC* and *IGHG1* (6-fold and 4-fold increased expression in SZ compared to CTL, respectively), both immunoglobulin components. Immune response has been extensively implicated as playing a role in psychiatric disease [16,18,33], however, a specific role for either of these two genes in psychiatric disease pathology remains to be characterized.

Conclusions

Our study provides several meaningful and novel contributions to the understanding of psychiatric disease. First, we have conducted one of the most extensive characterizations of the molecular overlap between SZ and BPD at the transcript and metabolite level across multiple brain regions. Second, we provide a high confidence set of genes differentially expressed between SZ and CTL patients utilizing two independent cohorts. Third, we demonstrate that a disproportionate amount of molecular heterogeneity occurs within SZ and BPD patient cohorts, and that this may

drive a significant proportion of observed case-control differences. We also establish that a majority of variation within SZ and BPD cohorts is strongly correlated with expression changes in previously identified cell-type specific transcripts. Finally, we propose several putative cell-type independent transcriptional changes that may warrant future investigation and highlight *NPAS4*, a transcription factor that may regulate several genes important for the development of disease.

Methods

See supplemental methods for additional detail.

Patient Sample Collection and Preparation

Sample collection, including human subject recruitment and characterization, tissue dissection, and RNA extraction, was described previously [13,41] as part of the Brain Donor Program at the University of California, Irvine, Department of Psychiatry and Human Behavior (Pritzker Neuropsychiatric Disorders Research Consortium) under IRB approval. In brief, all samples were diagnosed by psychological autopsy, which included collection and analyses of medical and psychiatric records, toxicology, medical examiners' reports, and 141-item family interviews. Controls were selected based upon absence of severe psychiatric disturbance and mental illness within first-degree relatives. We obtained fastq files from RNA-seq experiments for our validation cohort from the Stanley Neuropathology Consortium Integrative Database (SNCID; http://sncid.stanleyresearch.org) Array Collection comprising 35 cases each of SZ, BPD, and CTL of postmortem cingulate cortex with permission on June 30, 2015. For our analysis, we included the 27 SZ, 26 CTL, and 25 BPD SNCID samples that were

successfully downloaded and represented unique samples. SNCID RNA-seq methodology and data processing is described in detail in a previous publication that makes use of the data [8].

RNA-seq and Data Processing

Total RNA was extracted from 20 mg of postmortem brain tissue (350 µL tissue homogenate; 281 total tissue samples) using the Norgen Animal Tissue RNA Purification Kit (Norgen Biotek Corporation). We made RNA-seq libraries from 250 ng total RNA using polyA selection (Dynabeads mRNA DIRECT kit, Life Technologies) and transposase-based non-stranded library construction (Tn-RNA-seq) as described previously [14]. The pooled libraries were sequenced on an Illumina HiSeq 2000 sequencing machine using paired-end 50 bp reads and a 6 bp index read, resulting in an average of 48.2 million reads per library. To quantify the expression of each gene in both Pritzker and SNCID datasets, we used TopHat v1.4.1 [42] to align RNA-seq read pairs with GENCODE version 9 as a transcript reference. Raw read counts were calculated for each GENCODE transcript with Cufflinks 1.3.0 [43] and BEDTools[44]. Transcripts expressed from the X and Y chromosomes were omitted from the study.

Sequencing Data Analysis

All data analysis in R was performed with version 3.1.2.

Differential Expression Analysis and Normalization

To examine gene expression changes, we employed the R package DESeq2 [15] (version 1.6.3), using default settings, but employing likelihood ratio test (LRT) hypothesis testing, and removing non-convergent transcripts from subsequent analysis.

Genes differentially expressed between each disorder and CTL samples, by brain region, were identified with DESeq2 (adjusted p-value<0.05), including age, brain pH, and PMI as covariates. For downstream heatmap visualization, PCA, consensus clustering, and cell-type analysis, transcripts were normalized using DESeq2's varianceStabilizingTransformation function unless otherwise specified.

PCA, Hierarchical, and Consensus Clustering

PCA analysis was performed in R on normalized data using the prcomp() command. Hierarchical clustering of normalized transcript data was done in R with the hclust command (method="ward", distance="euclidean"). Consensus clustering was performed on normalized transcript data using the ConsensusCluster Plus [25] package in R (version 1.20.0) with *k*-means clustering from *k*=2 to *k*=4 (Euclidean distance) for 1000 repetitions of 80% of the data. Examination of the cumulative distribution function (CDF) revealed group optimization at *k*=2 for each analysis.

Pathway Enrichment Analysis

Pathway analysis was conducted using the web-based tool LRPath [45] using all GO term annotations, adjusting to transcript read count with RNA-Enrich, including directionality and limiting maximum GO term size to 500 genes. GO term visualization was performed using the Cytoscape Enrichment Map plug-in [46]. The Genesetfile (.gmt) GO annotations from July 24, 2015 were downloaded from http://download.baderlab.org/EM_Genesets/. Mapping parameters were; p-value cutoff = 0.005, FDR cutoff = 0.05 and overlap coefficient = 0.5. Resulting networks were exported as PDFs.

Cell-Specific Enrichment Analysis

Sets of transcripts uniquely expressed by several brain cell-types were obtained from figure 1B in Darmanis et. al [27]. An index for each cell-type was created by finding the median normalized expression value for each cell-type associated transcript set. Index values were compared across patient clusters by non-parametric rank sum tests and spearman correlation with top principle components. Ultimately, the neuron/astrocyte index ratio for each patient was included as a covariate in our DESeq2 regression model to normalize putative differences due to changes in these cell populations. To validate our method, we calculated cell-type specific indices from an independent cohort of previously published purified brain cells [28,29]. FPKM-normalized transcript data was obtained from supplemental table 4 of Zhang et. al. (2014) and cell-type indexes were calculated as described above. To examine index performance in mixed cell populations, we obtained fastq files for neuron and astrocyte-purified brain samples from GEO accession GSE73721 and generated raw count files as described above. We next mixed expression profiles *in silico* by performing random down-sampling of neuron and astrocyte count levels and summing the results such that mixed populations containing specific proportions of counts from neuron- and astrocyte-purified tissue were generated. For example, to generate an 80/20 neuron to astrocyte mixture, neuron and astrocyte count columns (which started at an equivalent number of 5,759,178 aligned reads) were randomly down-sampled to 4,607,342 and 1,151,836 counts respectively and summed across each gene to result in a proportionately mixed population of aligned count data simulating heterogeneous tissue. Then we calculated a neuron/astrocyte index ratio capable of predicting the *in silico* mixing weights. Briefly,

we assumed index values for mixed cell populations were directly proportional to mixing weights of their respective purified tissue, thus the predicted cell proportion for a given cell type was simply calculated as:

predicted cell proportion = observed index value/purified tissue index value

To insure cell-type predictive power was unique to indices derived from Darmanis et. al transcripts, we generated indices from 10,000 randomly sampled transcript sets of equivalent size and examined their performance in predicting *in silico* mixing weights. Mean squared prediction errors (MSE) were calculated for each of the 10,000 null indices and compared to the MSE of Darmanis et. al.-derived indices.

Metabolomics

Sample preparation

Sections of approximately 100mg of frozen tissue were weighed and homogenized for 45 seconds at 6.5M/s with ceramic beads in 1mL of 50% methanol using the MP FastPrep-24 homogenizer (MP Biomedicals). A sample volume equivalent to 10mg of initial tissue weight was dried down at 55°C for 60 minutes using a vacuum concentrator system (Labconco). Derivatization by methoximation and trimethylsilylation was done as previously described [47].

We analyzed technical replicates of each tissue sample, in randomized order.

GCxGC-TOFMS analysis

All derivatized samples were analyzed on a Leco Pegasus 4D system (GCxGC-TOFMS), controlled by the ChromaTof software (Leco, St. Joseph, MI). Samples were analyzed as described previously [47] with minor modifications in temperature ramp.

Data analysis and metabolite identification

Peak calling, deconvolution and library spectral matching were done using ChromaTOF 4.5 software. Peaks were identified by spectral match using the NIST, GOLM [48], and Fiehn libraries (Leco), and confirmed by running derivatized standards (Sigma). We used Guineu for multiple sample alignment [49].

NPAS4 Analysis

IM-32 cells were obtained from ATCC and grown under recommended cell culture conditions using EMEM media containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. ChIP-seq was performed as previously described [50] and the antibody for *NPAS4* was obtained from Abnova (H00266743-M09). ChIP-seq peaks were identified using MACS 2.1.0 and motif analysis was performed on all called peaks using MEME [51,52]. Only reproducible binding sites identified by replicate ChIP-seq experiments were used for downstream analysis. The SZ-implicated GWAS gene list was obtained from Ripke et. al [53]. Stable IM-32 knockdown cell lines were made with three independent shRNAs targeting *NPAS4* (SCHLNG-NM_178864) and a GFP (SHC002) targeting control shRNA (Sigma). Lentiviral packaging of shRNAS was performed as previously described [54,55]. One million IM-32 cells were transduced with packaged shRNAs and cells were selected with puromycin for one week. RNA was extracted the using a total RNA purification kit (Norgen) according to manufacturer protocol. Reverse transcription was performed using a high capacity cDNA reverse transcription kit (Applied Biosystems) and qPCR peformed with taqman probsets (NPAS4:Hs00698866_m1, WBPL1:Hs00740771_s1, ABCB9:Hs00608640_m1,

CHRNA3:Hs01088199_m1, EDC4:Hs01089874_g1, NDUFA6:Hs00899690_m1) to genes of interest (ThermoFisher).

List of abbreviations

RNA-seq – RNA sequencing

GABA – gamma-Aminobutyric acid

GWAS – genome-wide association study

SZ – schizophrenia

BPD – bipolar disorder

MDD – major depression disorder

CTL – control

AnCg – anterior cingulate gyrus

DLPFC – dorsolateral prefrontal cortex

nAcc – nucleus accumbens

GO – gene ontology

ChIP-seq – chromatin immunoprecipitation with DNA sequencing

shRNA – short hairpin RNA

PCA – principle component analysis

Declarations

Ethics approval and consent to participate

Sample collection, including human subject recruitment and characterization, was conducted as part of the Brain Donor Program at the University of California, Irvine, Department of Psychiatry and Human Behavior (Pritzker Neuropsychiatric Disorders Research Consortium) under IRB approval (UCI 88-041, UCI 97-74).

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are available in the GEO repository, GSE80655.

Competing interests

The authors declare that they have no competing interests.

Funding

The Pritzker Neuropsychiatric Disorders Research Fund L.L.C. and the NIH-National Institute of General Medical Sciences Medical Scientist Training Program (5T32GM008361-21) supported this work.

Author's Contributions

HA, SJW, AFS, WEB, JDB, HK, SJC and RMM conceived of study

KMB, RCR, BNL, SJC, and RMM designed the experiments

EGJ performed brain dissections

PMC procured the brain tissue samples

MPV analyzed pH on all cases and matched the 4 cohorts

DWM obtained demographic and clinical data on all subjects through analyses of medical records and next-of-kin interviews

NSD, JG, and KMB collected RNAs and performed Tn-RNA-seq library construction

RCR and BNL analyzed the RNA-seq data

MH and JZL conducted array experiments and analysis

RCR and SJC performed and analyzed metabolomics experiments

RCR and AAH performed and analyzed *NPAS4* ChIP-seq experiments

RCR performed and analyzed qPCR experiments

KMB, RCR, and BNL wrote the first draft of the paper

JZL, BGB, WEB, SJW, SJC, HA and RMM contributed to the writing of the paper

All authors read and approved the final manuscript.

Acknowledgements

We thank Marie Kirby, Brian Roberts, Mark Mackiewicz, and Greg Cooper for many helpful discussions and comments on the manuscript, and all the members of the Pritzker Neuropsychiatric Disorders Consortium for their support and advice.

References

1. CDC Burden of Mental Illness. 2015.

2. DSM-5. Diagnostic and Statistical Manual. 2013.

3. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans P a., et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature [Internet].

2014;511:421–7. Available from: http://www.nature.com/doifinder/10.1038/nature13595

4. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460:748–52.

5. Group C, Consortium PG. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet [Internet]. 2013 [cited 2014 Jan 20];381:1371–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23453885

6. Lee SH, Ripke S, Neale BM, Faraone S V, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. [Internet]. 2013 [cited 2014 Jan 21];45:984–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23933821

7. Jääskeläinen E, Juola P, Hirvonen N, McGrath JJ, Saha S, Isohanni M, et al. A systematic review and meta-analysis of recovery in schizophrenia. Schizophr. Bull. 2013;39:1296–306.

8. Hwang Y, Kim J, Shin JY, Kim JI, Seo JS, Webster MJ, et al. Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia. Transl. Psychiatry. 2013;3:e321.

9. Kohen R, Dobra A, Tracy JH, Haugen E. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. Transl. Psychiatry. 2014;4:e366.

10. Akula N, Barb J, Jiang X, Wendland JR, Choi KH, Sen SK, et al. RNA-sequencing of

the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Mol. Psychiatry [Internet]. 2014 [cited 2015 May 27];19:1179–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24393808

11. Olsen CM. Natural rewards, neuroplasticity, and non-drug addictions. Neuropharmacology [Internet]. 2011 [cited 2015 Jun 17];61:1109–22. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3139704&tool=pmcentrez&rendertype=abstract

12. Wenzel JM, Rauscher NA, Cheer JF, Oleson EB. A role for phasic dopamine release within the nucleus accumbens in encoding aversion: a review of the neurochemical literature. ACS Chem. Neurosci. [Internet]. 2015 [cited 2015 Aug 20];6:16–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25491156

13. Li JZ, Vawter MP, Walsh DM, Tomita H, Evans SJ, Choudary P V, et al. Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. Hum. Mol. Genet. [Internet]. 2004 [cited 2015 Apr 18];13:609–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14734628

14. Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, et al. Transposase mediated construction of RNA-seq libraries. Genome Res. 2012;22:134–41.

15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. [Internet]. 2014 [cited 2014 Dec 23];15:550. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4302049&tool=pmcentrez&rendertype=abstract

16. Fillman SG, Cloonan N, Catts VS, Miller LC, Wong J, McCrossin T, et al. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. Mol. Psychiatry [Internet]. Nature Publishing Group; 2012;18:206–14. Available from: http://dx.doi.org/10.1038/mp.2012.110

17. Fillman SG, Sinclair D, Fung SJ, Webster MJ, Shannon Weickert C. Markers of inflammation and stress distinguish subsets of individuals with schizophrenia and bipolar disorder. Transl. Psychiatry [Internet]. 2014 [cited 2015 Jul 28];4:e365. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3944638&tool=pmcentrez&rendertype=abstract

18. Saetre P, Emilsson L, Axelsson E, Kreuger J, Lindholm E, Jazin E. Inflammation-related genes up-regulated in schizophrenia brains. BMC Psychiatry [Internet]. 2007 [cited 2015 Jul 31];7:46. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2080573&tool=pmcentrez&rendertype=abstract

19. Arion D, Unger T, Lewis DA, Levitt P, Mirnics K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. Biol. Psychiatry [Internet]. 2007 [cited 2015 Jul 31];62:711–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2080683&tool=pmcentrez&rendertype=abstract

20. Akbarian S, Huntsman MM, Kim JJ, Tafazzoli A, Potkin SG, Bunney WE, et al. GABAA receptor subunit gene expression in human prefrontal cortex: comparison of

schizophrenics and controls. Cereb. Cortex [Internet]. [cited 2015 Apr 17];5:550–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8590827

21. Akbarian S, Huang HS. Molecular and cellular mechanisms of altered GAD1/GAD67 expression in schizophrenia and related disorders. Brain Res. Rev. 2006;52:293–304.

22. Duncan CE, Webster MJ, Rothmond D a., Bahn S, Elashoff M, Shannon Weickert C. Prefrontal GABAA receptor□??-subunit expression in normal postnatal human development and schizophrenia. J. Psychiatr. Res. [Internet]. Elsevier Ltd; 2010;44:673–81. Available from: http://dx.doi.org/10.1016/j.jpsychires.2009.12.007

23. Thompson M, Weickert CS, Wyatt E, Webster MJ. Decreased glutamic acid decarboxylase(67) mRNA expression in multiple brain areas of patients with schizophrenia and mood disorders. J. Psychiatr. Res. [Internet]. 2009 [cited 2015 May 14];43:970–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19321177

24. Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO, Johansson M, et al. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res. [Internet]. 2013 [cited 2015 Aug 20];23:1496–504. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3759725&tool=pmcentrez&rendertype=abstract

25. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.

26. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic

ductal adenocarcinoma. Nat. Genet. [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2015 [cited 2015 Sep 10];47:1168–78. Available from: http://dx.doi.org/10.1038/ng.3398

27. Darmanis S, Sloan S a., Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. 2015;201507125.

28. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. J. Neurosci. [Internet]. 2014 [cited 2014 Sep 5];34:11929–47. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4152602&tool=pmcentrez&rendertype=abstract

29. Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, et al. Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron [Internet]. 2015 [cited 2015 Dec 16];89:37–53. Available from: http://www.sciencedirect.com/science/article/pii/S0896627315010193

30. Coutellier L, Beraki S, Ardestani PM, Saw NL, Shamloo M. Npas4: a neuronal transcription factor with a key role in social and cognitive functions relevant to developmental disorders. PLoS One [Internet]. 2012 [cited 2015 Aug 20];7:e46604. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3460929&tool=pmcentrez&rendertype=abstract

31. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature [Internet]. Nature Publishing Group; 2010;465:182–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3020079&tool=pmcentrez&rendertype=abstract

32. Ooe N, Saito K, Mikami N, Nakatuka I, Kaneko H. Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression. Mol. Cell. Biol. 2004;24:608–16.

33. Kim S, Hwang Y, Webster MJ, Lee D. Differential activation of immune/inflammatory response-related co-expression modules in the hippocampus across the major psychiatric disorders. Mol. Psychiatry [Internet]. Nature Publishing Group; 2015;1–10. Available from: http://www.nature.com/doifinder/10.1038/mp.2015.79

34. Sequeira A, Morgan L, Walsh DM, Cartagena PM, Choudary P, Li J, et al. Gene expression changes in the prefrontal cortex, anterior cingulate cortex and nucleus accumbens of mood disorders subjects that committed suicide. PLoS One [Internet]. 2012 [cited 2014 Jul 15];7:e35367. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3340369&tool=pmcentrez&rendertype=abstract

35. Weissman DH, Gopalakrishnan a., Hazlett CJ, Woldorff MG. Dorsal anterior cingulate cortex resolves conflict from distracting stimuli by boosting attention toward relevant events. Cereb. Cortex. 2005;15:229–37.

36. Paus T. Primate anterior cingulate cortex: where motor control, drive and cognition

interface. Nat. Rev. Neurosci. 2001;2:417–24.

37. Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science. 1998;280:747–9.

38. Woo T-UW, Kim AM, Viscidi E. Disease-specific alterations in glutamatergic neurotransmission on inhibitory interneurons in the prefrontal cortex in schizophrenia. Brain Res. [Internet]. 2008 [cited 2015 May 27];1218:267–77. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2665281&tool=pmcentrez&rendertype=abstract

39. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 2011;43:977–83.

40. Jaehne EJ, Klarić TS, Koblar SA, Baune BT, Lewis MD. Effects of Npas4 deficiency on anxiety, depression-like, cognition and sociability behaviour. Behav. Brain Res. 2015;281:276–82.

41. Evans SJ, Choudary P V, Vawter MP, Li J, Meador-Woodruff JH, Lopez JF, et al. DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. Neurobiol. Dis. [Internet]. 2003 [cited 2015 Apr 18];14:240–50. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3098567&tool=pmcentrez&rendertype=abstract

42. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11.

43. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010;28:511–5.

44. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics [Internet]. 2014 [cited 2014 Dec 4];47:11.12.1–11.12.34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25199790

45. Kim JH, Karnovsky A, Mahavisno V, Weymouth T, Pande M, Dolinoy DC, et al. LRpath analysis reveals common pathways dysregulated via DNA methylation across cancer types. BMC Genomics [Internet]. 2012 [cited 2015 Aug 20];13:526. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3505188&tool=pmcentrez&rendertype=abstract

46. Isserlin R, Merico D, Voisin V, Bader GD. Enrichment Map - a Cytoscape app to visualize and explore OMICs pathway enrichment results. F1000Research [Internet]. 2014 [cited 2015 Aug 20];3:141. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4103489&tool=pmcentrez&rendertype=abstract

47. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat. Protoc. [Internet]. 2011 [cited 2015 Dec 28];6:1060–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21720319

48. Hummel J, Strehmel N, Selbig J, Walther D, Kopka J. Decision tree supported

substructure prediction of metabolites from GC-MS profiles. Metabolomics [Internet]. 2010 [cited 2016 Feb 3];6:322–33. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2874469&tool=pmcentrez&rendertype=abstract

49. Castillo S, Mattila I, Miettinen J, Orešič M, Hyötyläinen T. Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. Anal. Chem. [Internet]. 2011 [cited 2015 Aug 24];83:3058–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21434611

50. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol. Cell [Internet]. Elsevier Inc.; 2013;52:25–36. Available from: http://dx.doi.org/10.1016/j.molcel.2013.08.037

51. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. [Internet]. 2008;9:R137. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2008-9-9-r137

52. Bailey TL, Williams N, Misleh C, Li WW. MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006;34:369–73.

53. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature [Internet]. 2014 [cited 2014 Jul 22];511:421–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4112379&tool=pmcentrez&rendertype=abstract

54. Kutner RH, Zhang X-Y, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nat. Protoc. 2009;4:495–505.

55. Kirby MK, Ramaker RC, Gertz J, Davis NS, Johnston BE, Oliver PG, et al. RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4. Mol. Oncol. [Internet]. Elsevier B.V; 2016;1–14. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1574789116300400

Figure Legends

Figure 1. Brain region-specific gene expression in psychiatric disorder and control postmortem brains. PCA of all 48,956 gene transcripts across psychiatric disorders and controls in (A) AnCg, (B) DLPFC, and (C) nACC. CTL (gray squares), BPD (blue triangles), MDD (green circles), SZ (red triangles).

Figure 2. (A) Pairwise spearman correlations of $\log_2$ fold gene expression changes between each disorder and CTL in each brain region. Circle sizes are scaled to reflect Spearman correlations. (B) Venn diagram showing overlap of genes differentially expressed between SZ (red), BPD (blue), MDD (green) vs. CTL at a p-value<0.05. (C) Log_2 fold change of 1,203 genes with FDR<0.05 comparing SZ and CTL (AnCg) in the Pritzker dataset (open circles), with transcripts also significant in the Stanley Cohort (n=129, Spearman coefficient=0.89, p-value<0.0001) identified with red circles. (D) Hierarchical clustering of 24 SZ and 24 CTL tissues in the Pritzker dataset and (E) 27 SZ and 26 CTL tissues in the SCNID dataset using variance-stabilized expression of 129 significant genes between SZ and CTL in the AnCg identified by DESeq2

(FDR<0.05) in both datasets. CTL (dark gray), SZ (red), low gene expression (blue pixels), high gene expression (yellow pixels).

Figure 3. Metabolomic analysis of SZ, BPD, MDD, and CTL in AnCg Hierarchical clustering of (A) the 25 most different metabolite levels between SZ and CTL, (B) the 25 most different metabolite levels between BPD and CTL. (C) Metabolite PC2 distinguishes BPD and SZ patients from CTL and is correlated with RNA-seq PC1 (D) demonstrating a subset of SZ and BPD patients differ from CTL, MDD and a majority of SZ and BPD patients.

Figure 4. Heterogeneity within psychiatric disorders. (A) PCA of SNCID cingulate cortex gene expression data. CTL (gray squares), BPD (blue triangles), SZ (red triangles). (B-C) Smooth loess fit of coefficient of variation (CV) for each transcript with 95% confidence intervals in each disorder in Pritzker (B) and SCNID (C) datasets. (D) Correlation of PC1 loadings in the Pritzker and SCNID datasets. (E) GO term analysis of differentially expressed genes between the two consensus clusters in the Pritzker dataset. Up-regulation (red circles), down-regulation (blue circles).

Figure 5. Cell-type gene expression changes in AnCg of clustered subset of patients (consensus cluster 2) compared to the remaining patients (consensus cluster 1) (A) Glial cells (oligodendrocytes, astrocytes, and microglia) and endothelial cells are increased in cluster 2, whereas neurons are decreased. (B) The ratio of astrocytes to neurons correlates strongly with PC1 in the Pritzker dataset and (C) the SNCID dataset. Histogram of p-value distribution in SZ vs. CTL (AnCg) before (red) and after (purple) neuron/astrocyte correction in both (D) the Pritzker dataset and (E) the SNCID dataset.

Figure 6. (A) Cumulative distribution function (CDF) plot indicating *NPAS4* binds near previously implicated SZ genes (red) compared to a background containing all genes (black). (B) Relative expression of *NPAS4* and previously implicated SZ genes within 1 Kb of an *NPAS4* binding site in triplicate transduction of non-targeting shRNAs (blue) and *NPAS4* targeting shRNAs (red).

Supplementary Figure Legends

Figure S1. Brain region-specific gene expression in psychiatric disorder and control postmortem brains. A) Principal components analysis of all 281 brain tissues and 48,956 gene transcripts. AnCg (red squares), DLPFC (blue triangles), nAcc (green circles). B) Principal components analysis of all 281 brain tissues and 48,956 gene transcripts. CTL (gray squares), BPD (blue triangles), MDD (green circles), SZ (red triangles).

Figure S2. Correlation between RNA-seq and microarray analysis in AnCg, SZ vs. CTL (R^2=0.71, p-value<0.0001). FDR<0.1 in both datasets (red circles).

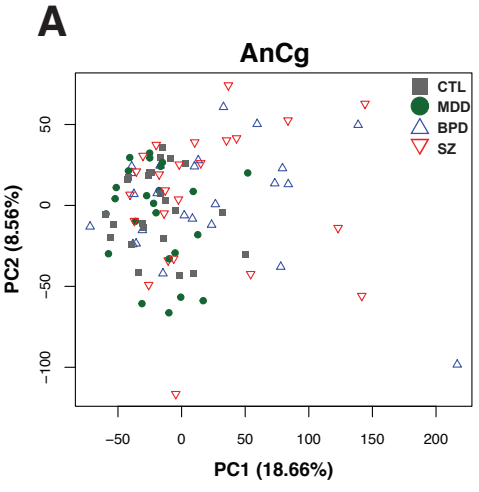
Figure S3. GO-term analysis for transcripts differentially expressed in SZ vs. CTL in AnCg (FDR<0.05). Up-regulation (red circles), down-regulation (blue circles).

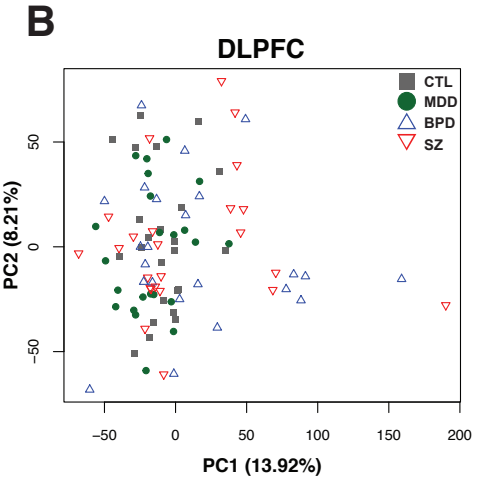
Figure S4. SZ and BPD expression signature overlap in a subset of patients. (A) PC1 in RNA-seq vs. microarray data. CTL (gray squares), BPD (blue triangles), MDD (green circles), SZ (red triangles). Heatmap of consensus matrix on all transcripts for (B) Pritzker (Cluster 1: 16 BPD, 16 SZ, 23 MDDs, 23 CTL; Cluster 2: 8 BPD, 8 SZ, 1 MDD, 1 CTL) and (C) SCNID (Cluster 1: 19 BPD, 25 CTL, 18 SZ; Cluster 2: 6 BPD, 1 CTL, 9 SZ).

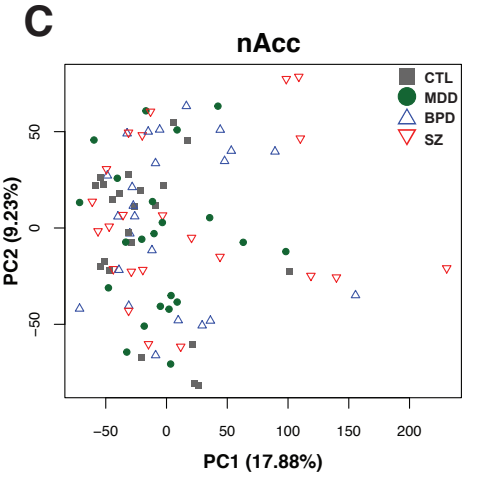
Figure S5. (A) Boxplots of brain pH by cluster for SZ and BPD cohorts. Outlier patients are represented as cluster 2 in pink. (B) PCA of all patient transcript residuals after pH normalization in all brain regions colored by cluster (pink outlier subset, gray – the rest of patients).

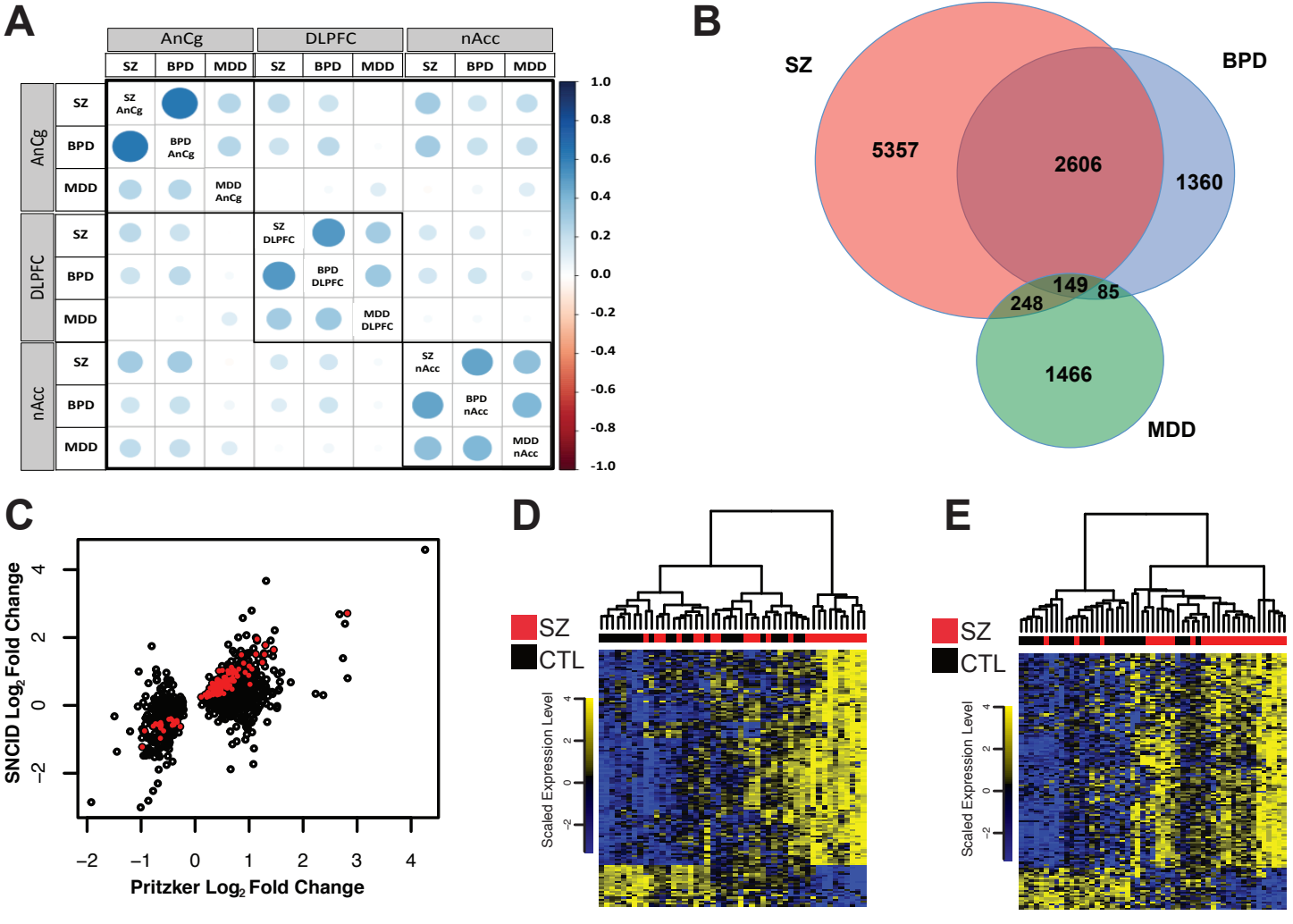
Figure S6. Examination of cell-type specific index in purified (A) neuron, (B) astrocytes, (C) oligodendrocytes, (D) microglia, and (E) endothelial cells from brain tissue. (F) Neuron and astrocyte indices are capable of predicting *in silico* mixed cell-type proportions. (G) Mean values with standard deviation for predictions of indices generated on 10,000 randomly sampled, null transcript sets. (H, I) Histogram of mean squared error of null index cell-type proportion predictions for mixed neuron and astrocyte transcriptomes with Darmanis et al. transcript performance indicated in red.

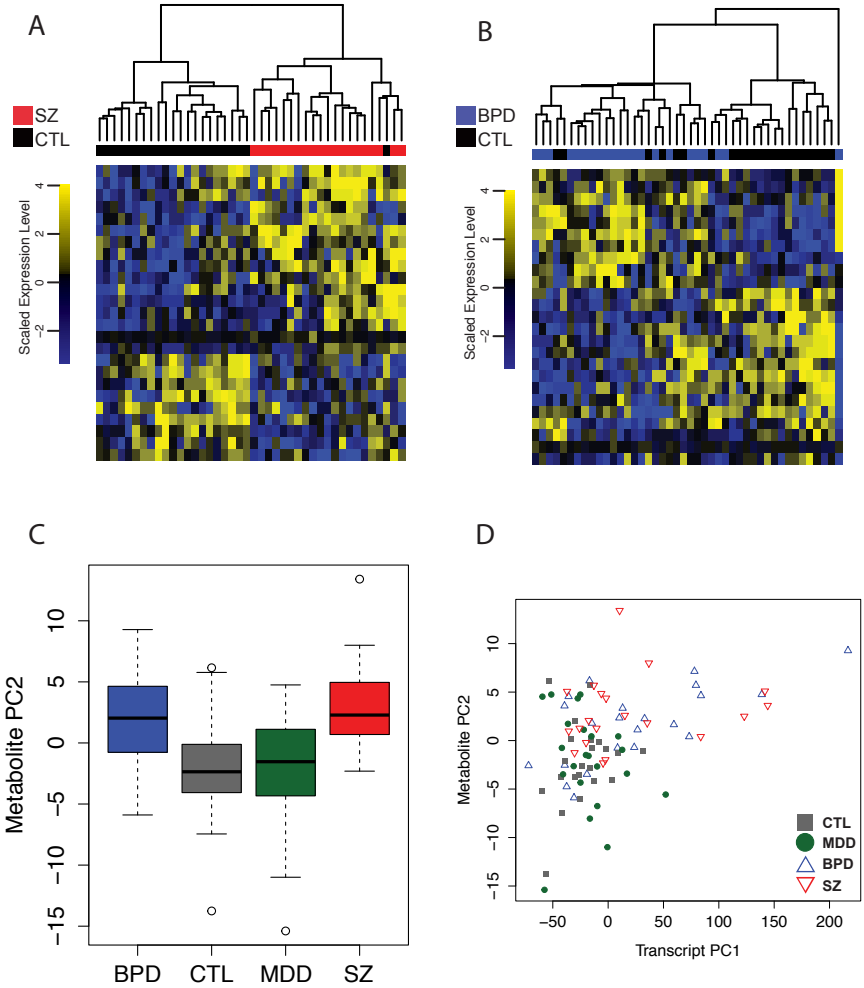
Figure S7. (A) Venn diagram of binding sites between *NPAS4* ChIP-seq replicates. (B) Correlation of fold enrichment over reverse cross-link controls of each *NPAS4* ChIP-seq replicate. (C) Significant motifs close to *NPAS4* canonical motifs identified from MEME analysis of high confidence ChIP-seq binding sites.

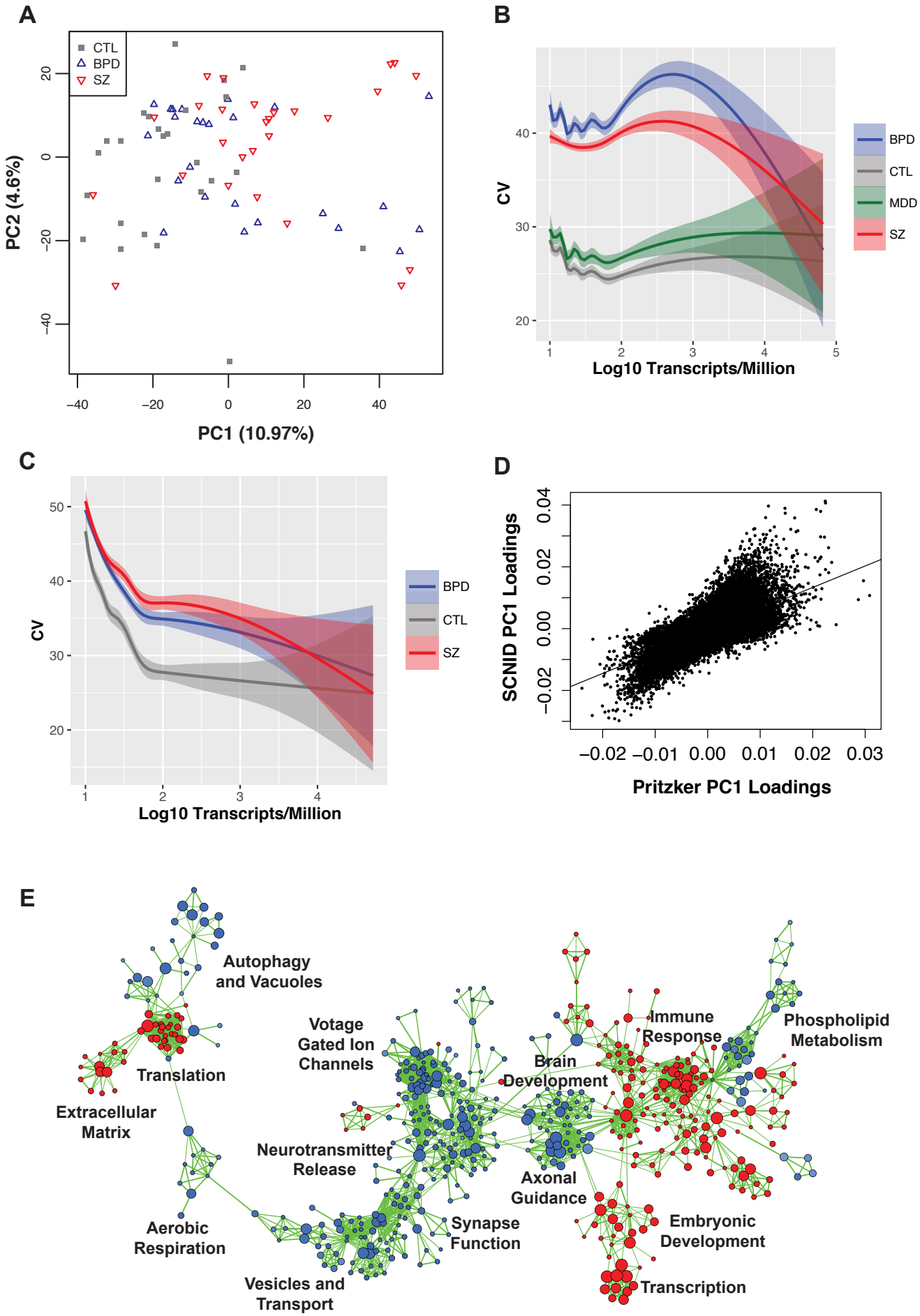

A
AnCg
CTL
MDD
BPD
SZ
PC2 (8.56%)
PC1 (18.66%)
50
0
−50
−100
−50
0
50
100
150
200

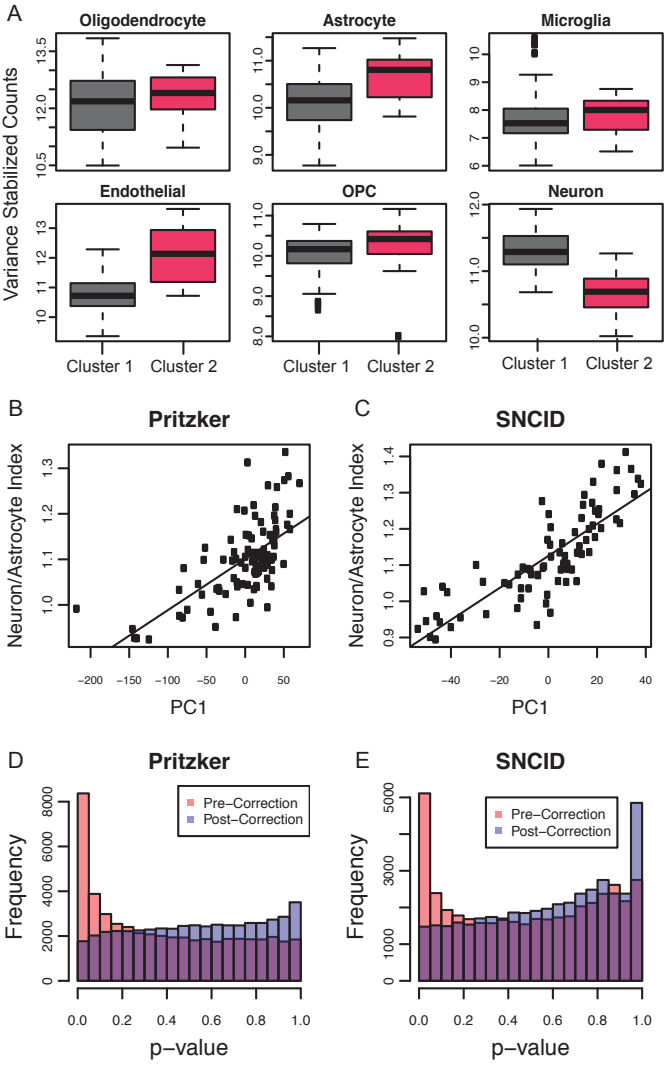

B
DLPFC
CTL
MDD
BPD
SZ
PC2 (8.21%)
PC1 (13.92%)
50
0
−50
−50
0
50
100
150
200

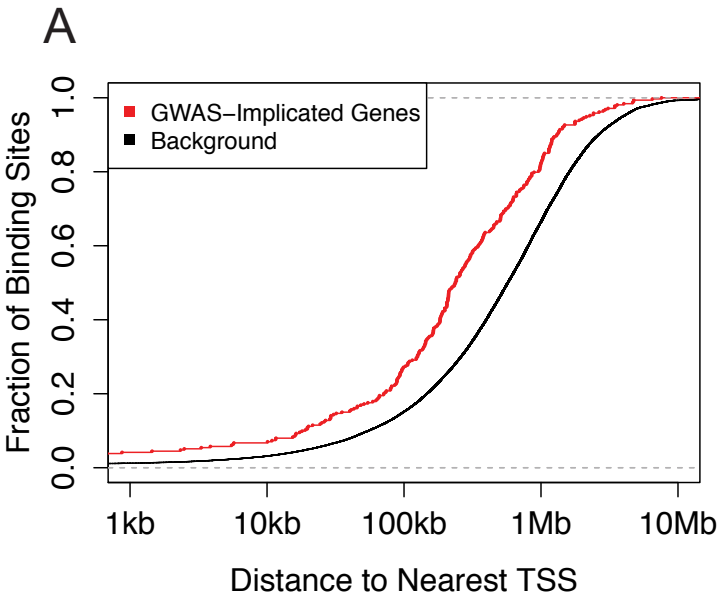

C
nAcc
CTL
MDD
BPD
SZ
PC2 (9.23%)
PC1 (17.88%)
50
0
−50
−50
0
50
100
150
200

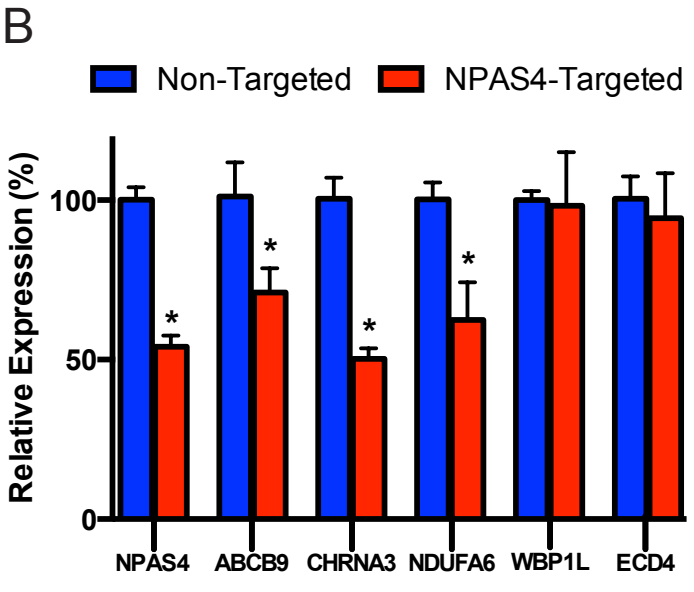

A
AnCg
DLPFC
nAcc
SZ
BPD
MDD
SZ AnCg
BPD AnCg
MDD AnCg
SZ DLPFC
BPD DLPFC
MDD DLPFC
SZ nAcc
BPD nAcc
MDD nAcc
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
B
SZ
BPD
MDD
5357
2606
1360
149
248
85
1466
C
SNCID Log2 Fold Change
Pritzker Log2 Fold Change
D
SZ
CTL
Scaled Expression Level
E
SZ
CTL
Scaled Expression Level

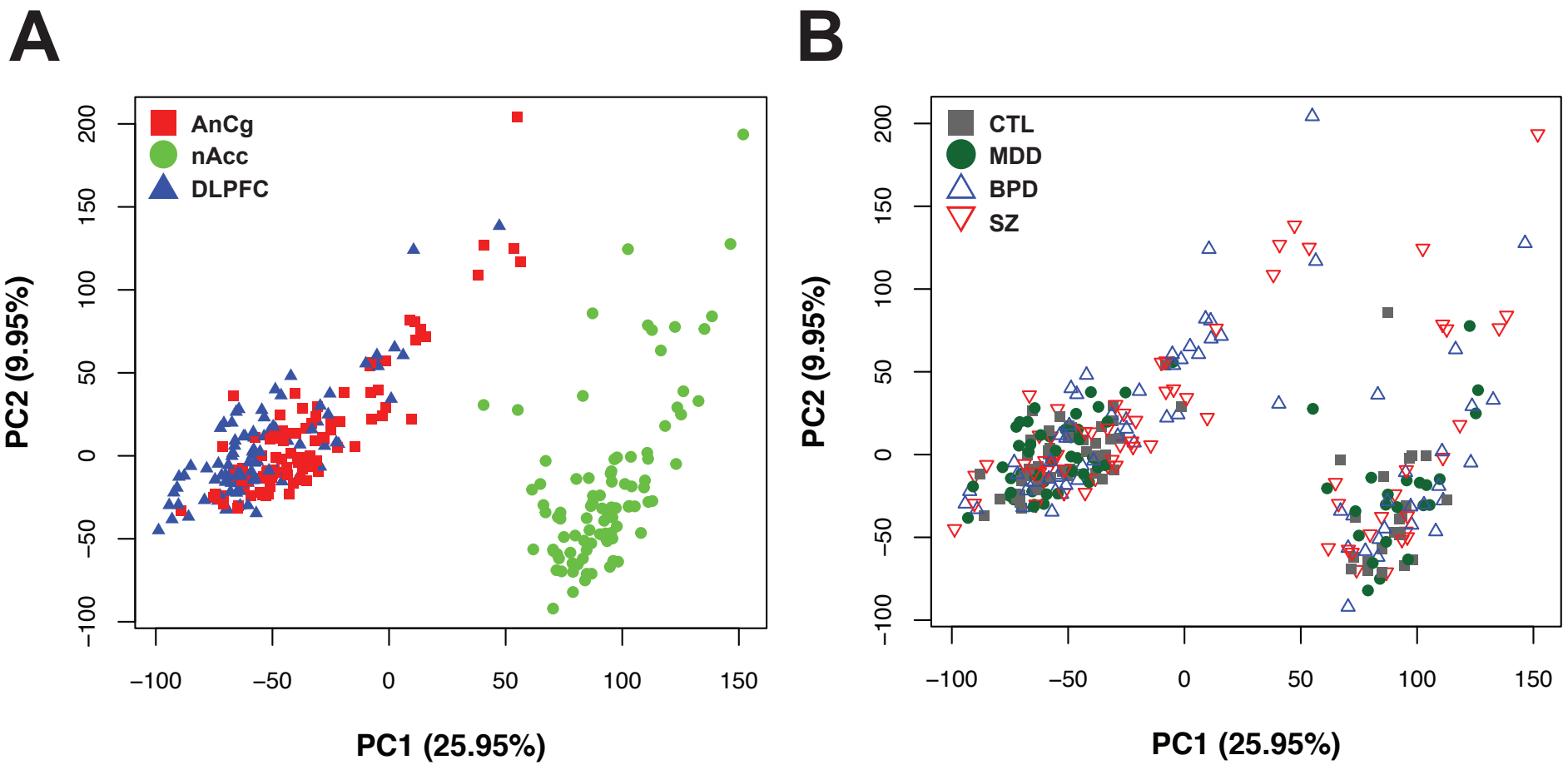

A
SZ
CTL
Scaled Expression Level
4
2
0
−2
B
BPD
CTL
Scaled Expression Level
4
2
0
−2
C
Metabolite PC2
10
5
0
−5
−10
−15
BPD
CTL
MDD
SZ
D
Metabolite PC2
10
5
0
−5
−10
−15
−50
0
50
100
150
200
Transcript PC1
CTL
MDD
BPD
SZ

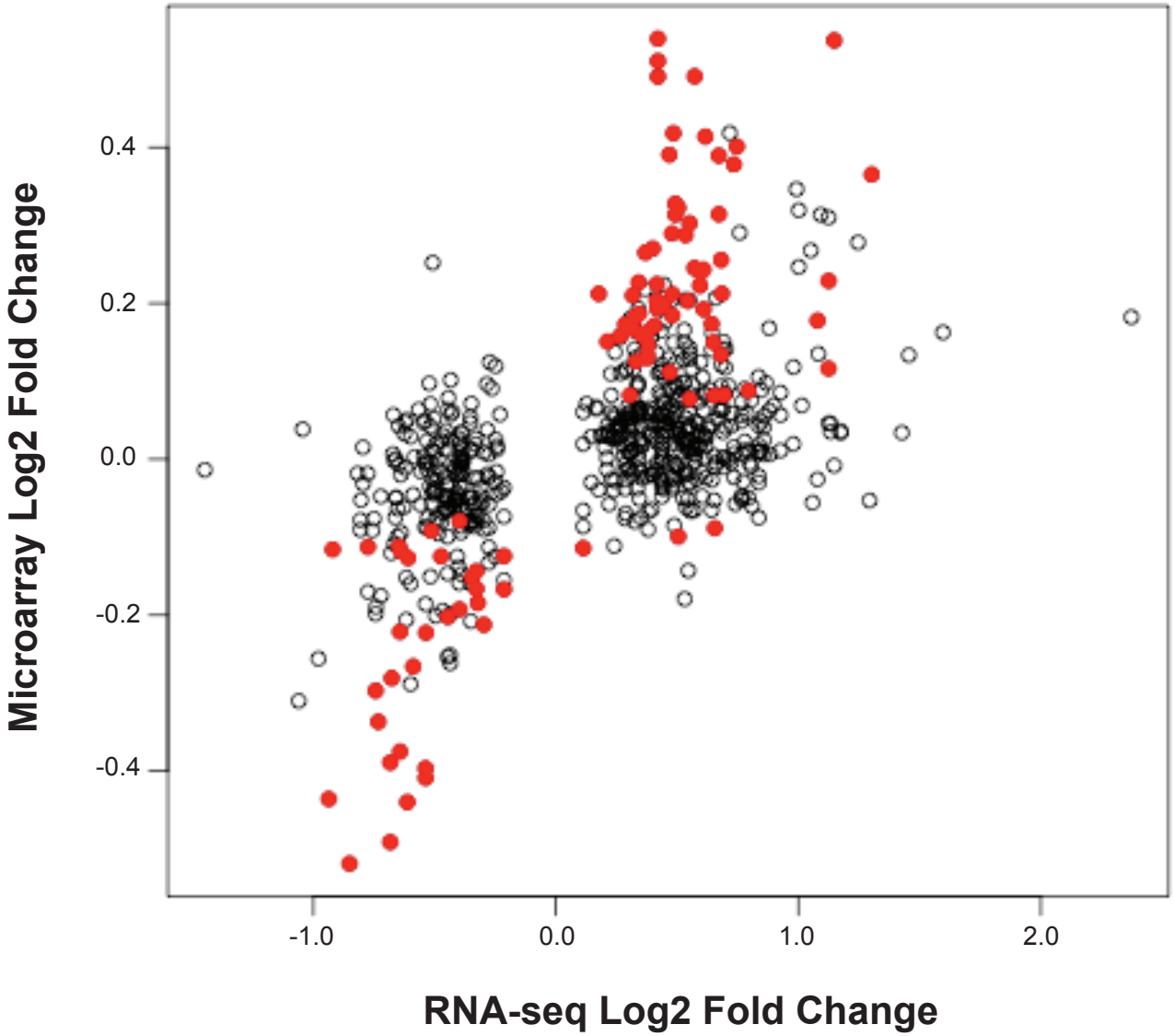

A
CTL
BPD
SZ
PC2 (4.6%)
PC1 (10.97%)
B
CV
Log10 Transcripts/Million
BPD
CTL
MDD
SZ
C
CV
Log10 Transcripts/Million
BPD
CTL
SZ
D
SCNID PC1 Loadings
Pritzker PC1 Loadings
E
Autophagy and Vacuoles
Translation
Extracellular Matrix
Aerobic Respiration
Votage Gated Ion Channels
Neurotransmitter Release
Vesicles and Transport
Synapse Function
Brain Development
Axonal Guidance
Immune Response
Phospholipid Metabolism
Embryonic Development
Transcription

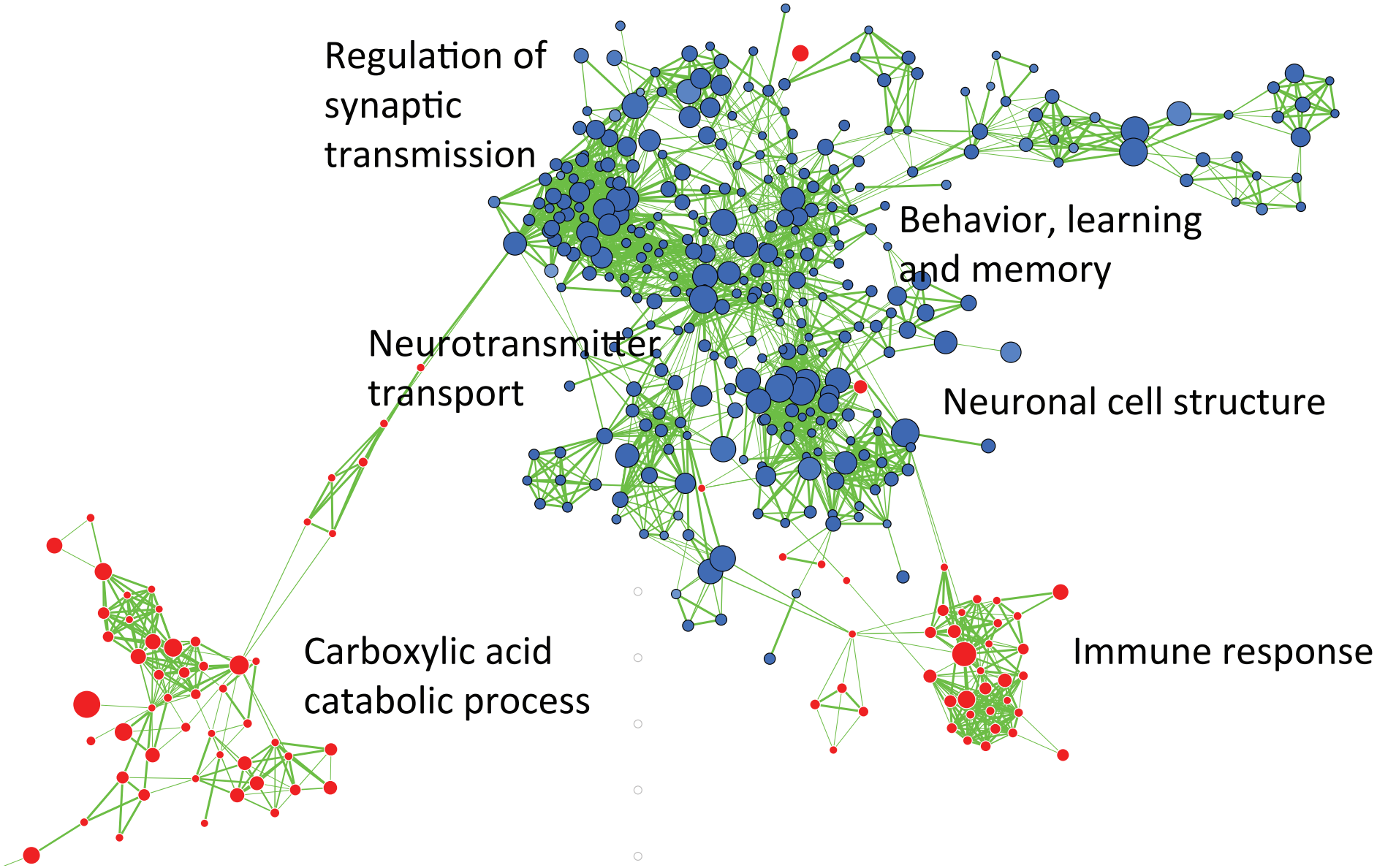

A
Oligodendrocyte
Astrocyte
Microglia
Endothelial
OPC
Neuron
Variance Stabilized Counts
Cluster 1
Cluster 2
B
Pritzker
Neuron/Astrocyte Index
PC1
C
SNCID
Neuron/Astrocyte Index
PC1
D
Pritzker
Pre−Correction
Post−Correction
Frequency
p−value
E
SNCID
Pre−Correction
Post−Correction
Frequency
p−value

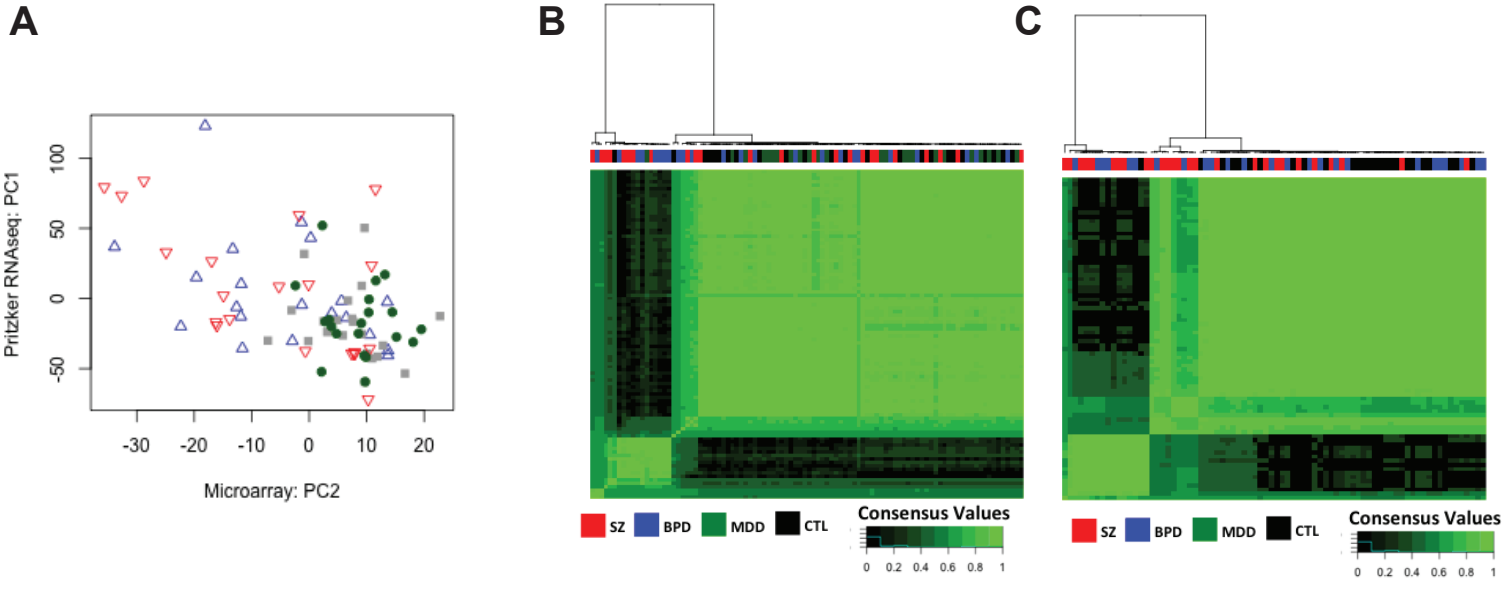

A
GWAS–Implicated Genes
Background
Fraction of Binding Sites
0.0
0.2
0.4
0.6
0.8
1.0
1kb
10kb
100kb
1Mb
10Mb
Distance to Nearest TSS

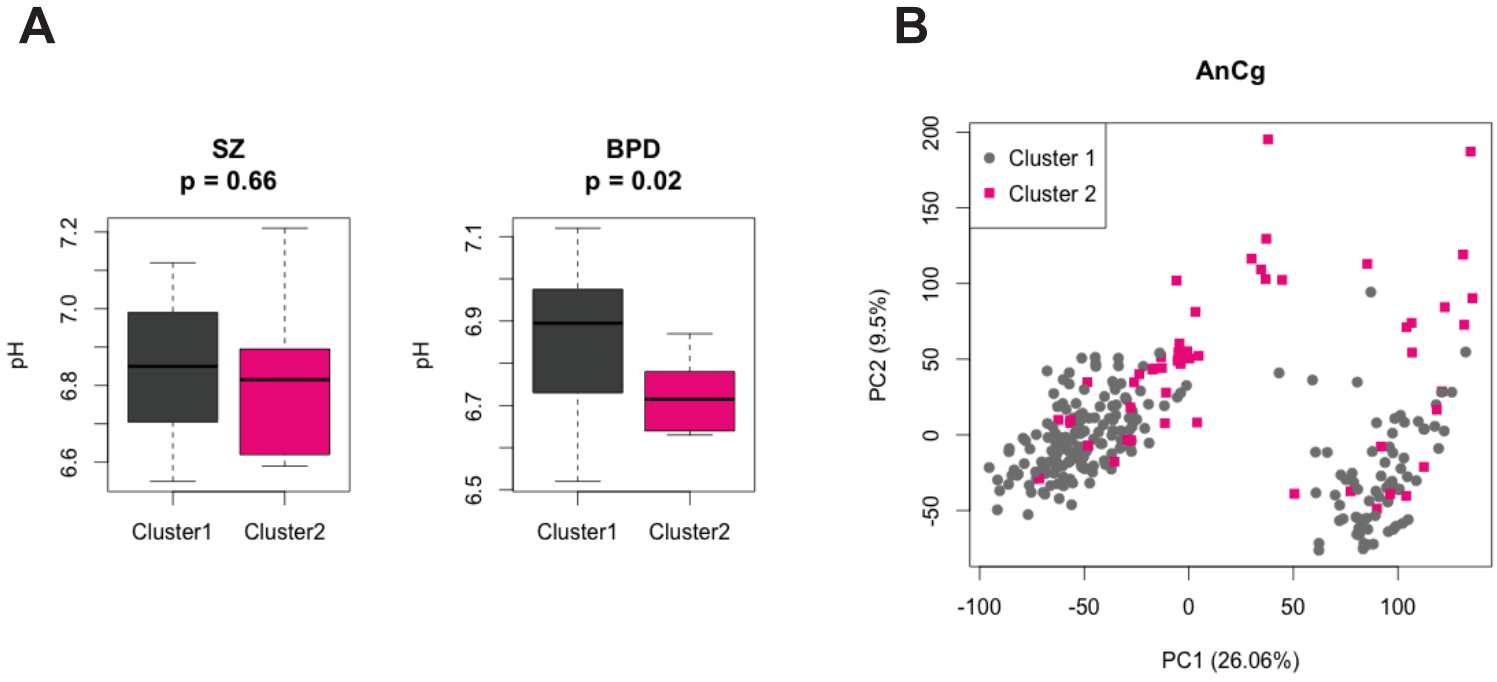

B
Non-Targeted
NPAS4-Targeted
Relative Expression (%)
100
50
0
*
NPAS4
ABCB9
CHRNA3
NDUFA6
WBP1L
ECD4

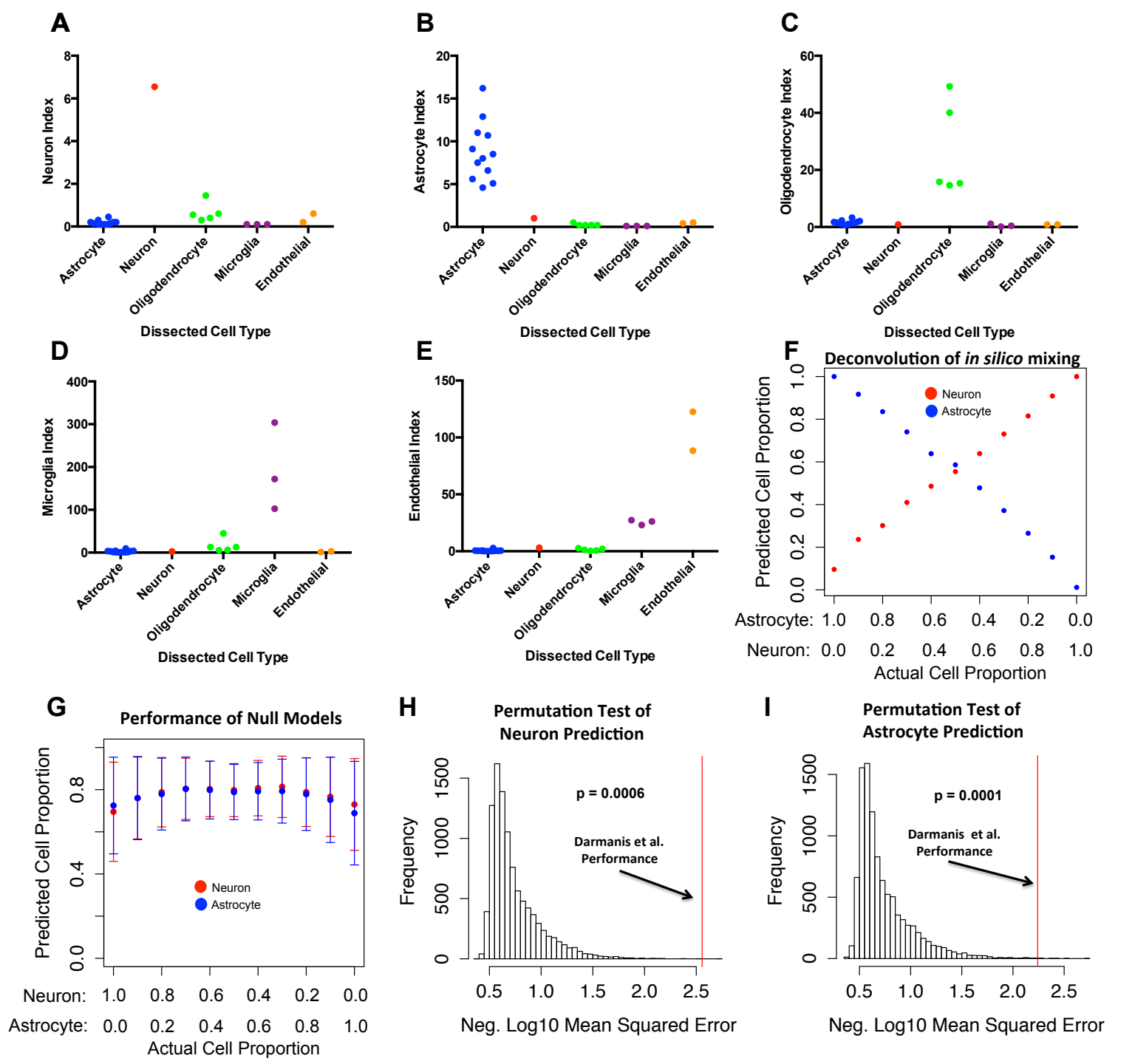

A
AnCg
nAcc
DLPFC
PC2 (9.95%)
PC1 (25.95%)
200
150
100
50
0
−50
−100
−100
−50
0
50
100
150
B
CTL
MDD
BPD
SZ
PC2 (9.95%)
PC1 (25.95%)
200
150
100
50
0
−50
−100
−100
−50
0
50
100
150

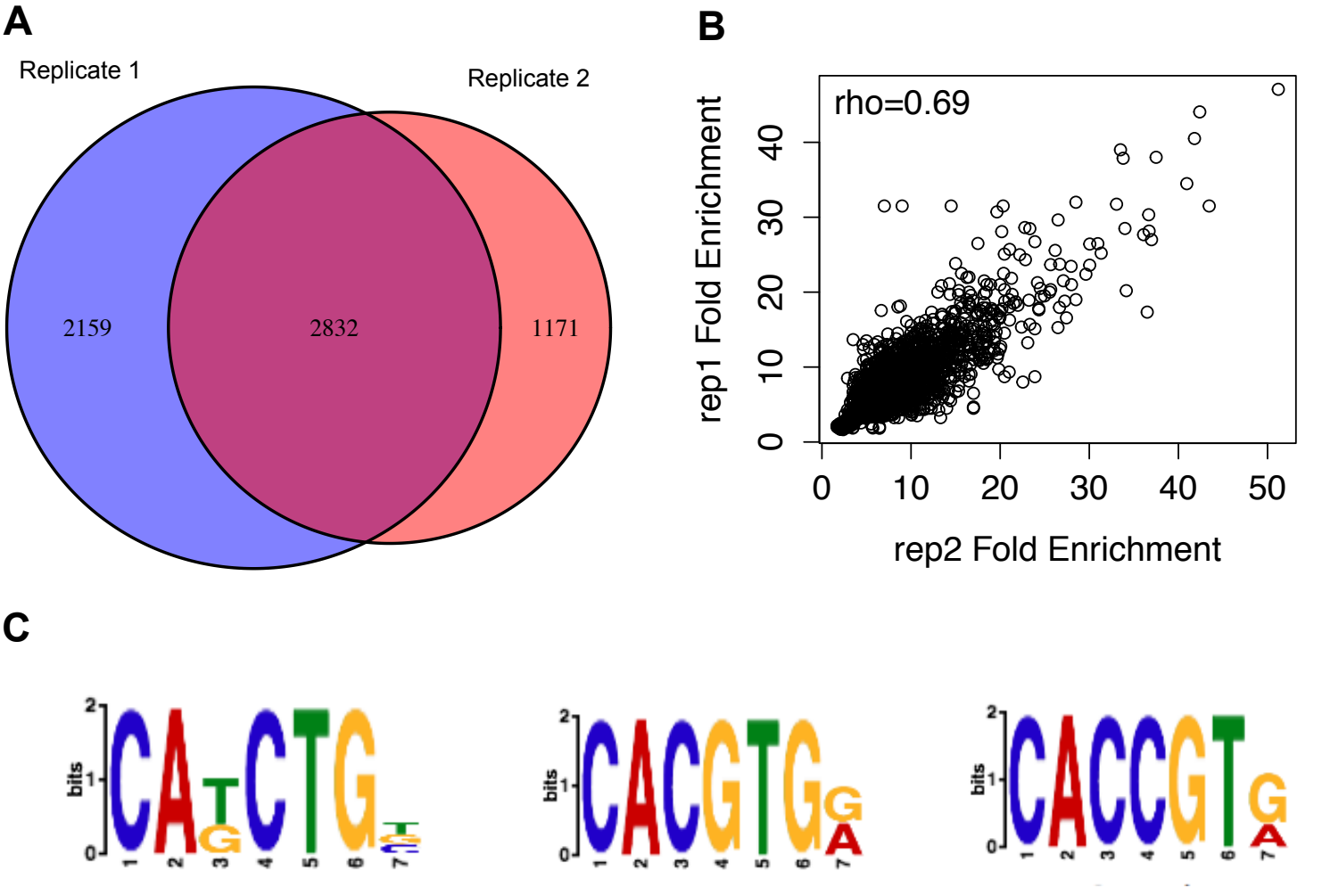

Microarray Log2 Fold Change
0.4
0.2
0.0
-0.2
-0.4
-1.0
0.0
1.0
2.0
RNA-seq Log2 Fold Change


Regulation of synaptic transmission
Behavior, learning and memory
Neurotransmitter transport
Neuronal cell structure
Carboxylic acid catabolic process
Immune response


A
Pritzker RNAseq: PC1
100
50
0
-50
-30
-20
-10
0
10
20
Microarray: PC2
B
SZ
BPD
MDD
CTL
Consensus Values
0
0.2
0.4
0.6
0.8
1
C
SZ
BPD
MDD
CTL
Consensus Values
0
0.2
0.4
0.6
0.8
1


A
SZ
p = 0.66
pH
7.2
7.0
6.8
6.6
Cluster1
Cluster2
BPD
p = 0.02
pH
7.1
6.9
6.7
6.5
Cluster1
Cluster2
B
AnCg
Cluster 1
Cluster 2
PC2 (9.5%)
200
150
100
50
0
-50
-100
-50
0
50
100
PC1 (26.06%)


A
Neuron Index
Astrocyte
Neuron
Oligodendrocyte
Microglia
Endothelial
Dissected Cell Type
B
Astrocyte Index
Dissected Cell Type
C
Oligodendrocyte Index
Dissected Cell Type
D
Microglia Index
Dissected Cell Type
E
Endothelial Index
Dissected Cell Type
F
Deconvolution of in silico mixing
Predicted Cell Proportion
Neuron
Astrocyte
Astrocyte: 1.0 0.8 0.6 0.4 0.2 0.0
Neuron: 0.0 0.2 0.4 0.6 0.8 1.0
Actual Cell Proportion
G
Performance of Null Models
Predicted Cell Proportion
Neuron
Astrocyte
Neuron: 1.0 0.8 0.6 0.4 0.2 0.0
Astrocyte: 0.0 0.2 0.4 0.6 0.8 1.0
Actual Cell Proportion
H
Permutation Test of Neuron Prediction
Frequency
p = 0.0006
Darmanis et al. Performance
Neg. Log10 Mean Squared Error
I
Permutation Test of Astrocyte Prediction
Frequency
p = 0.0001
Darmanis et al. Performance
Neg. Log10 Mean Squared Error


A
Replicate 1
Replicate 2
2159
2832
1171
B
rho=0.69
rep1 Fold Enrichment
rep2 Fold Enrichment
0
10
20
30
40
50
C
bits
1
2
3
4
5
6
7