Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder

Michael J. Gandal [1-4*], Pan Zhang [5], Evi Hadjimichael [6-8, 19], Rebecca Walker [2-4], Chao Chen [9], Shuang Liu [10], Hyejung Won [2-4], Harm van Bakel [7], Merina Varghese [19, 20], Yongjun Wang [11], Annie W. Shieh [12], Jillian Haney [1-3], Sepideh Parhami [1-3], Judson Belmont [6-8, 19], Minsoo Kim [1, 2], Patricia Moran Losada [5], Zenab Khan [7], Justyna Mleczko [14], Yan Xia [9], Rujia Dai [9], Daifeng Wang [13], Yucheng T. Yang [10], Min Xu [10], Kenneth Fish [14], Patrick R. Hof [19, 20], Jonathan Warrell [10], Dominic Fitzgerald [15], Andrew E. Jaffe [16, 17], PsychENCODE Consortium[‡], Kevin White [15], Mette A. Peters [18], Mark Gerstein [10], Chunyu Liu [9, 12*], Lilia M. Iakoucheva [5*], Dalila Pinto [6-8, 19*], Daniel H. Geschwind [1-4*]

Affiliations

[1] Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA.

[2] Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[3] Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA.

[4] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[5] Department of Psychiatry, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA

[6] Department of Psychiatry, and Seaver Autism Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[7] Department of Genetics and Genomic Sciences, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[8] The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[9] The School of Life Science, Central South University, Changsha, Hunan 410078, China

[10] Program in Computational Biology and Bioinformatics, Departments of Molecular Biophysics and Biochemistry and Computer Science, Yale University, New Haven, CT, USA

[11] The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China

[12] Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA

[13] Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA

[14] Departments of Medicine and Cardiology, Cardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[15] Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA.

[16] Lieber Institute for Brain Development, Baltimore, MD, USA

[17] Departments of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

[18] CNS Data Coordination group, Sage Bionetworks, Seattle, WA 98109, USA

[19] Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[20] Fishberg Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[‡] Full list of authors and affiliations are listed in the supplementary materials

Print Summary

Introduction and Rationale:

Understanding of the pathophysiology of psychiatric disorders, including autism spectrum disorder (ASD), schizophrenia (SCZ), and bipolar disorder (BD), lags behind most other fields of medicine. In the absence of clearly defined pathology, the diagnosis and study of psychiatric disorders are dependent on behavioral, symptomatic characterization. Defining genetic contributions to disease risk provides a substantial foothold for biological understanding. But, leveraging genetic risk to infer disease mechanisms is challenged by substantial genetic complexity and polygenicity, and the lack of a cohesive neurobiological model through which to interpret genetic findings. Recent work demonstrates that the transcriptome represents a quantitative phenotype that provides biological context for understanding the molecular pathways disrupted in major psychiatric disorders (*1*, *2*). We reasoned that RNA sequencing (RNA-Seq) in a large cohort of cases and controls would substantially advance knowledge of the biology disrupted in each disorder and provide a foundational resource for integration with other genomic and genetic data.

Results:
Analysis of multiple levels of transcriptomic organization – gene expression, local splicing, transcript isoform expression, and co-expression networks for both protein-coding and non-coding genes – provides an in depth view of ASD, BD, and SCZ molecular pathology. Over 25% of the transcriptome exhibits differential splicing (DS) or expression (DE) in at least one disorder, including 916 non-coding RNAs (ncRNAs), most of which have unexplored functions, and as a group are under increased constraint in humans relative to ncRNA genome-wide. Local splicing analysis permits identification of genes exhibiting isoform switching across disorders and cell types. Changes at the isoform-level, rather than gene-level, show the largest effect sizes, genetic enrichment, and greatest disease specificity. We identify 61 co-expression modules associated with at least one disorder, the majority of which show enrichment for cell type-specific marker genes, with 5 modules significantly dysregulated across all three disorders. These modules allow parsing of previously shared downregulated neuronal and synaptic components into a variety of cell type- and disease-specific signals, including multiple excitatory neuron and distinct interneuron modules with differential patterns of disease association. We also refine the glial-immune signal, demonstrating shared disruption of the blood-brain-barrier and upregulation of NFkB-associated genes, as well as disease-specific alterations in microglial, astrocyte and interferon-response modules. To identify candidate causal drivers, we integrate polygenic risk scores (PRS) and perform a transcriptome-wide association study (TWAS). Dozens of genes are significantly associated with PRS, and TWAS prioritizes novel candidate risk genes likely mediated by *cis*-effects on brain expression, including 12 in BD, 5 in ASD, and 107 in SCZ.

Conclusion: By integrating RNA-sequencing and genetic data in an unprecedented cohort to refine the shared and distinct molecular pathology of ASD, BD, and SCZ, we provide a quantitative, genome-wide resource for mechanistic insight and therapeutic development and an interactive browser to permit further biological exploration of gene co-expression networks. These data inform the molecular pathways and cell types involved, emphasizing the importance of local splicing and isoform-level gene regulatory mechanisms in defining cell type and disease specificity and when integrated with GWAS, permit the discovery of new candidate risk genes.

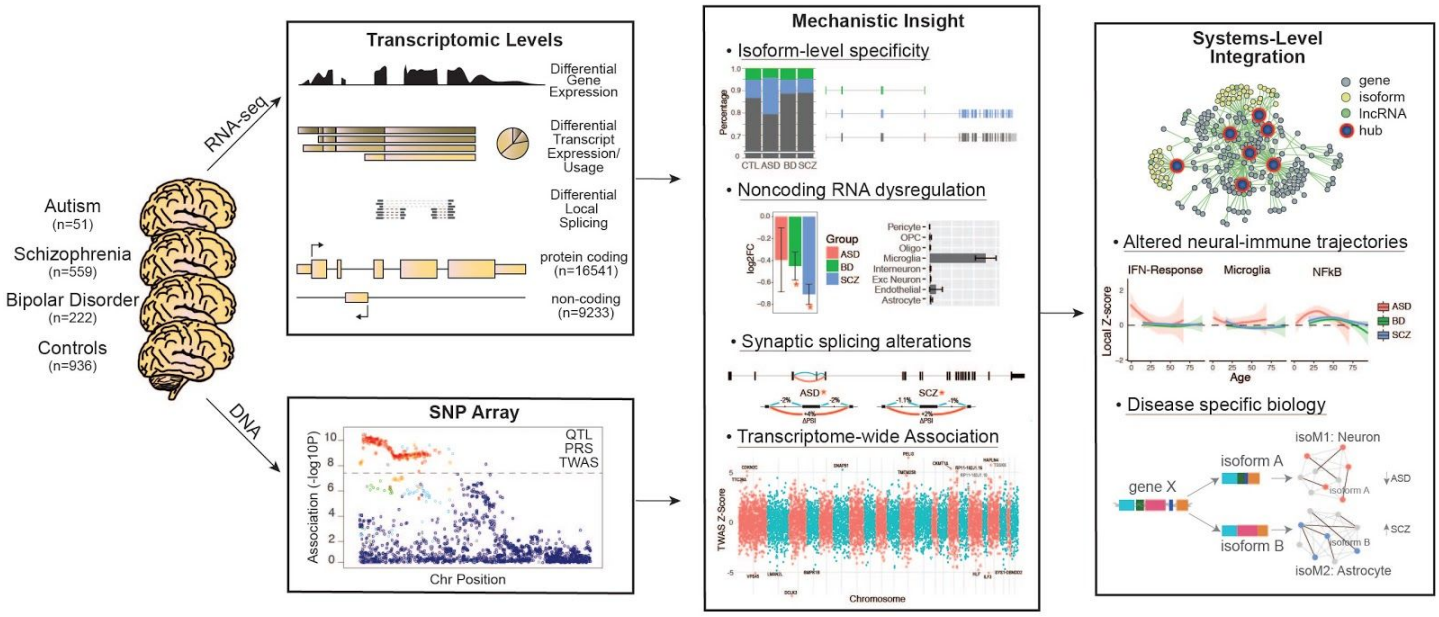


Fig 0. The PsychENCODE cross-disorder transcriptomic resource

Human brain RNA-Seq was integrated with genotypes across 1695 unique subjects with ASD, SCZ, BD, and controls. A comprehensive analysis identified dysregulation across all levels of transcriptomic organization – including protein-coding, non-coding, alternative splicing and isoform-level changes. Systems-level and integrative functional genomic analyses prioritize novel neurogenetic mechanisms and identify shared and distinct aspects of the molecular neuropathology of three major psychiatric disorders.

Abstract

Most genetic risk for psychiatric disease lies in regulatory regions, implicating pathogenic dysregulation of gene expression and splicing. However, comprehensive assessments of transcriptomic organization in disease brain are limited. Here, we integrate genotype and RNA-sequencing in brain samples from 1695 subjects with autism, schizophrenia, bipolar disorder and controls. Over 25% of the transcriptome exhibits differential splicing or expression, with isoform-level changes capturing the largest disease effects and genetic enrichments. Co-expression networks isolate disease-specific neuronal alterations, as well as microglial, astrocyte, and interferon response modules defining novel neural-immune mechanisms. We prioritize disease loci likely mediated by *cis*-effects on brain expression via transcriptome-wide association analysis. This transcriptome-wide characterization of the molecular pathology across three major psychiatric disorders provides a comprehensive resource for mechanistic insight and therapeutic development.

Main Text

Introduction

Developing more effective treatments for autism (ASD), schizophrenia (SCZ), and bipolar disorder (BD), three common psychiatric disorders that confer lifelong disability, is a major international public health priority (*3*). Studies have identified hundreds of causal genetic variants robustly associated with these disorders, and thousands more that likely contribute to their pathogenesis (*4*). However, the neurobiological mechanisms through which genetic variation imparts risk, both individually and in aggregate, are still largely unknown (*4*–*6*).

The majority of disease-associated genetic variation lies in non-coding regions (*7*) enriched for non-coding RNAs and *cis* regulatory elements that regulate gene expression and splicing of their cognate coding gene targets (*8*, *9*). Such regulatory relationships show substantial heterogeneity across human cell types, tissues, and developmental stages (*10*), and are often highly species-specific (*11*). Recognizing the importance of understanding transcriptional regulation and non-coding genome function, several consortia (*10*, *12*–*14*) have undertaken large-scale efforts to provide maps of the transcriptome and its genetic and epigenetic regulation across human tissues. Although some have included CNS tissues, a more comprehensive analysis focusing on the brain in both healthy and disease states is necessary to accelerate our understanding of the molecular mechanisms of these disorders (*1*, *15*–*17*).

We present results of the analysis of RNA-sequencing (RNA-Seq) data from the PsychENCODE Consortium (*17*), integrating genetic and genomic data from over 2000 well-curated, high-quality post-mortem brain samples from individuals with SCZ, BD, ASD, and controls (*18*). We provide a comprehensive resource of disease-relevant gene expression changes and transcriptional networks in the postnatal human brain (see Resource.PsychENCODE.org for raw data and annotations). Data was generated across eight studies (*2*, *19*, *20*), uniformly processed, and combined through a consolidated genomic data processing pipeline ((*21*); **Fig S1**), yielding a total of 2188 samples passing quality control (QC) for this analysis, representing frontal and temporal cerebral cortex from 1695 unique subjects across the lifespan, including 279 technical replicates (**Fig S2**). Extensive quality control steps were taken within and across individual studies resulting in the detection of 16,541 protein-coding and 9,233 non-coding genes with the Gencode v19 annotations ((*21*); **Fig S3**). There was substantial heterogeneity in RNA-Seq methodologies across cohorts, which we accounted for by including 28 surrogate variables and aggregate sequencing metrics as covariates in downstream analyses of differential expression (DE) at gene, isoform, and local splicing levels (*21*). Differential expression did not overlap with experimentally defined RNA degradation metrics in brain, indicating that results were not driven by RNA-quality confounds (**Fig S4**) (*22*).

To provide a comprehensive view of the transcriptomic architecture of these disorders, we characterize several levels of transcriptomic organization – gene-level, transcript isoform, local splicing, and co-expression networks – for both protein-coding and non-coding gene biotypes. We integrate results with common genetic variation and disease GWAS to identify putative regulatory targets of genetic risk variants. Although each level provides important disease-specific and shared molecular pathology that we highlight below, we find that isoform-level changes show the largest effects in disease brain, are most reflective of genetic risk, and provide the greatest disease specificity when assembled into co-expression networks.

We recognize that these analyses involve a variety of steps and data types and are necessarily multifaceted and complex. We have therefore organized the work into two major sections. The first is at the level of individual genes and gene products, starting with gene level transcriptomic analyses, isoform and splicing analyses, followed by identification of potential genetic drivers. The second section is anchored in gene network analysis, where we identify coexpression modules at both gene and isoform levels and assess

their relationship to genetic risk. As these networks reveal many layers of biology, we provide an interactive web-browser to permit their in depth exploration (Resource.PsychENCODE.org).

Gene and Isoform Expression Alterations in Disease

RNA-Seq based quantifications enabled assessment of coding and non-coding genes and transcript isoforms, imputed using RSEM guided by Gencode v19 annotations (*21*, *23*). In accordance with previous results (*1*), we observed pervasive differential gene expression (DGE) in ASD, SCZ, and BD (n=1611, 4821, and 1119 genes at FDR<0.05, respectively; **Fig 1A**; **Table S1**). There was substantial cross-disorder sharing of this DE signal and a gradient of transcriptomic severity with the largest changes in ASD compared with SCZ or BD (ASD vs SCZ, mean $|\log_2 FC|$ 0.26 *vs* 0.10, $P<2x10^{-16}$, Kolmogorov-Smirnov (K-S) test; ASD *vs* BD, mean $|\log_2 FC|$ 0.26 *vs* 0.15, $P<2x10^{-16}$, K-S test), as observed previously (*1*). Altogether, over a quarter of the brain transcriptome was affected in at least one disorder (**Fig 1A**-**C**; complete gene list, **Table S1**).

DGE results were highly concordant with previously published datasets for all three disorders, although some had overlapping samples (**Fig S4**). We observed significant concordance of DGE effect sizes with those from a microarray meta-analysis of each disorder (ASD: ρ=0.8, SCZ: ρ=0.78, BD: ρ=0.64, Spearman ρ of $\log_2 FC$, all P's$<10^{-16}$, (*1*)) and with previous RNA-Seq studies of individual disorders (ASD: ρ=0.96, ref (*19*); SCZ ρ=0.78, ref (*2*); SCZ ρ=0.80, ref (*24*); BD ρ=0.85, ref (*1*); Spearman ρ of $\log_2 FC$, all P's$<10^{-16}$). These DE genes exhibited substantial enrichment for known pathways and cell type specific markers derived from single nucleus RNA-Seq in human brain (**Fig 1D**-**E**) (*21*), consistent with previously observed patterns (*1*, *19*).

Expanding these analyses to the transcript isoform-level, we observe widespread differential transcript expression (DTE) across ASD, SCZ, and BD (n=767, 3803, and 248 isoforms at FDR<0.05, respectively; **Table S1**). Notably, at the DTE level, the cross-disorder overlap was significantly attenuated (**Fig 1C**), suggesting that alternative transcript usage and/or splicing confers a substantial portion of disease specificity. In addition to greater disease specificity, isoform-level alterations in disease exhibited substantially larger effect sizes compared with gene-level changes (mean $|\log_2 FC|$ 0.25 *vs* 0.14, $P<2x10^{-16}$, K-S test), particularly for protein coding biotypes (**Fig 1A**), consistent with recent work demonstrating the importance of splicing dysregulation in disease pathogenesis (*25*). Furthermore, although isoform and gene-level changes were overall similar in terms of pathways and cell types affected (e.g. **Fig 1D**-**E**), isoform-level analysis identified DE transcripts that did not show DGE ('isoform-only DE'), including 811 in SCZ, 294 in ASD, and 60 in BD. These isoform-only DE genes were more likely to be downregulated than upregulated in disease (one sample t-test, $P<10^{-16}$), were most significantly enriched in excitatory neuron clusters (OR's > 4, Fisher's exact test, FDR's$<10^{-10}$), and showed significant enrichment for neuron projection development, mRNA metabolism, and synaptic pathways ($FDR<3x10^{-3}$; **Table S1**). To validate DTE results, we performed PCR on several selected transcripts in a subset of ASD, SCZ and control samples (*21*), and find significant concordance in fold-changes compared with those from RNA-Seq data (**Fig S5A-B**). Together, these results suggest that isoform-level changes are most reflective of neuronal and synaptic dysfunction characteristic of each disorder.

Though there are multiple shared pathways at the levels of DGE or DTE across disorders, there are also several distinctive features (**Fig 1D**-**E**). Disorder-specific pathway enrichments include decreased transmembrane transport, synapse and synaptic components, with increases in innate immune response genes in ASD; decreased chemokine signaling, regulation of lymphocyte regulated immunity and natural killer cell chemotaxis in BD; and decreased signaling receptor and transmembrane receptor activity with increases in genes involved in the inflammatory response in SCZ. With regards to cell type enrichments (**Fig 1E**), although there was substantial downregulation of neuronal synaptic and signaling genes, only SCZ and BD also showed increases in the expression of a distinct subset of excitatory and inhibitory neuronal genes, whereas SCZ and ASD showed upregulation of genes expressed in astrocytes. ASD was the only disorder with enrichment of microglia among upregulated features.

Differential Expression of the Non-coding Transcriptome

Non-coding RNAs (ncRNAs) represent the largest class of transcripts in the human genome and are associated with complex phenotypes (*26*). However, most have limited functional annotation, particularly in human brain and have not been studied in psychiatric disease. Based on Gencode annotations, we identify 944 ncRNAs exhibiting gene- or isoform-level DE in at least one disorder (herein referred to as 'neuropsychiatric (NP) ncRNAs' (*21*)), 693 of which were DE in SCZ, 178 in ASD, and 174 in BD, of which 208, 60, and 52 are annotated as intergenic long non-coding RNAs (lincRNAs), in each disorder, respectively. To place these NPncRNAs within a functional context, we examined expression patterns across human tissues, cell types, and developmental time periods, as well as sequence characteristics including evolutionary conservation, selection, and constraint. We highlight several noncoding genes exhibiting DE across multiple disorders (**Fig S6**) and provide comprehensive annotations for each NPncRNA (**Table S2**)**,** including cell type specificity, developmental trajectory, and constraint, to permit placement of these NPncRNAs within a functional context in human brain.

As a class, NPncRNAs were under greater selective constraint compared to all Gencode annotated ncRNAs (**Fig 1F**), consistent with the observed increased purifying selection in brain-expressed genes (*27*). We identify 74 NPncRNAs (~8%) under purifying selection in humans, with average exon-level context-dependent tolerance scores (CDTS) below the 10th percentile (*21*). Of the 944 NP ncRNAs, 212 exhibited broad and non-specific expression patterns across cell types, whereas 66 showed specific expression within a single cell type class (**Table S2**). These data provide a foundation for understanding cell type specific, circuit level aspects of lncRNA function in neuropsychiatric disease. Two notable examples are, *LINC00996,* which is downregulated in SCZ (log_2FC -0.71, FDR<$5x10^{-11}$) and BD (log_2FC -0.45, FDR=0.02) and restricted to microglia in brain (**Fig S6**), and *LINC00343,* expressed in excitatory neurons, and downregulated in BD (log_2FC -0.33, FDR=0.012) with a trend in SCZ (log_2FC -0.15, FDR 0.065).

Local Splicing Dysregulation in Disease

Isoform-level diversity is achieved by combinatorial use of alternative transcription start sites, polyadenylation, and splicing (*28*). We next used LeafCutter (*29*) to assess local differential splicing (DS) differences in ASD, SCZ and BD using *de novo* aligned RNA-seq reads, controlling for the same covariates as DGE/DTE (**Fig S7**). This approach complements DTE by considering aggregate changes in intron usage affecting exons that may be shared by multiple transcripts and thus, is not restricted to the specified genome annotation (*21*). Previous studies have highlighted the contribution of local DS events in ASD (*19*, *30*) and in smaller cohorts in SCZ (*2*, *24*) and BD (*31*).

We identified 515 DS intron clusters in 472 genes across all disorders (FDR<0.1), 117 of which (25%) contained one or more novel exons (**Table S3**; **Fig 2A**). Validation of DS changes for 9 genes in a subset of cases and controls (n= 5-10 in each group) by semiquantitative RT-PCR showed percent spliced-in (PSI) changes consistent with those reported by LeafCutter (**Fig S5C-E**). The most commonly observed local splicing change was exon skipping (41-60%), followed by alternative 5' exon inclusion (e.g. due to alternative promoter usage; 11-21%) and alternative 3' splice site usage (5-18%) (**Table S3**; **Fig S8A**). DS genes overlapped significantly with DTE results for ASD and SCZ (**Fig S8B**), but not BD, which likely still remains underpowered. There was significant cross-disorder correlation in PSI changes (Spearman's ρ=0.59 SCZ-BD, ρ=0.52 SCZ-ASD, all P<10^{-4}) and subsequently, overlap among DS genes (**Fig 2A**-**B**), although the majority of splicing changes still are disorder specific. Only two genes, *DTNA* and *AHCYL1*, were significantly DS in all three disorders (**Fig S9**). DS genes showed significant (FDR<0.05) enrichment for signaling, cell communication, actin cytoskeleton, synapse, and neuronal development pathways across disorders (**Figs 2C**, **S8C**), and were predominantly expressed in neuronal cell types, astrocytes (in ASD, SCZ), microglia and oligodendrocytes (in SCZ) (**Fig 2D**). Disorder specific pathways implicated by splicing dysfunction include

plasma membrane receptor complex, endocytic vesicle, regulation of cell growth and cytoskeletal protein binding in ASD; angiotensin receptor signaling in BD; and GTPase receptor activity, neuron development and actin cytoskeleton in SCZ. We also find significant enrichment of splicing changes in targets of two RNA binding proteins that regulate synaptic transmission and whose targets are implicated in both ASD and SCZ, the neuronal splicing regulator *RBFOX1* (FDR=5.16×10^{-11}) (*32*) and the fragile X mental retardation protein (FMRP) (FDR=3.10×10^{-21}) (*33*). Notably, 48 DS genes (10%; FDR=8.8×10^{-4}) encode RNA binding proteins or splicing factors (*34*), with at least six splicing factors also showing DTE in ASD (*MATR3*), SCZ (*QKI*, *RBM3*, *SRRM2*, *U2AF1*) or both (*SRSF11*).

Many differential splicing events show predictable functional consequences on protein isoforms. Notable examples include *GRIN1* and *NRXN1*, which are known risk loci for neurodevelopmental disorders (*35*, *36*). *GRIN1* encodes the obligatory subunit of the NMDA-type glutamate ionotropic receptors, is upregulated in SCZ and BD and shows increased skipping of exon 4 in both ASD and SCZ that impacts its extracellular ligand-binding domain (**Fig 2E**-**G**). *NRXN1* is a heterotypic, presynaptic cell adhesion molecule that undergoes extensive alternative splicing and plays a key role in the maturation and function of synapses (*35*, *37*). We observed various DS and/or differential transcript usage (DTU) changes in *NRXN1* in ASD, SCZ and/or BD (**Fig 2H**-**K**). An exon skipping event in ASD disrupts a laminin domain in *NRXN1* (**Fig 2I-J**), while the isoform expression switch affects the expression of laminin, neurexin-like and EGF-like domains; changes which are predicted to have major effects on its function (**Fig 2H**). Another example is *CADPS*, which is located within an ASD GWAS risk locus and supported by Hi-C defined chromatin interactions as a putative target gene (*38*) and manifests multiple isoform and splice alterations in ASD (**Fig S9**; **Tables S1** and **S3**).

We found significant overlap (42%, P=3.42×10^{-27}; Fisher's exact test) of the ASD DS intron clusters and splicing changes identified in a previous study (*19*) that used a different method and only a subset of the samples in our ASD and control cohorts (**Table S3**). Overall, this examination of local splicing across three major neuropsychiatric disorders, coupled with the analysis of isoform-level regulation, emphasizes the need to understand the regulation and function of transcript isoforms at a cell type specific level in the human nervous system.

Identifying Drivers of Transcriptome Dysregulation

We next set to determine whether changes observed across multiple levels of transcriptomic organization in psychiatric disease brain are reflective of the same, or distinct, underlying biological processes. Furthermore, such transcriptomic changes may represent a causal pathophysiology or may be a consequence of disease. To begin to address this, we assessed the relationships among transcriptomic features and with polygenic risk scores (PRS) for disease, which provide a causal genetic anchor (**Fig 3A**). Across all three disorders, there was strong concordance among differential gene, isoform, and ncRNA signals, as summarized by their first principal component (**Fig 3A**). Notably, differential splicing shows greatest overlap with the ncRNA signal, suggesting a role for non-coding genes in regulating local splicing events.

Significant associations with PRS were observed for DGE and DTE signal in SCZ, with greater polygenic association at the isoform level in accordance with the larger transcript isoform effect sizes observed. Concordantly, transcript-level differential expression also showed the most significant enrichment for SCZ SNP-heritability, as measured by stratified LD score regression (*21*, *39*) (**Fig 3B**). The overall magnitude of genetic enrichment was modest, however, suggesting that most observed transcriptomic alterations are less a proximal effect of genetic variation and more likely the consequence of a downstream cascade of biological events following earlier acting genetic risk factors.

We were also interested to determine the degree to which genes showed increases in the magnitude of DE over the duration of illness, as a positive relationship would be expected if age-related cumulative exposures (e.g. drugs, smoking) were driving these changes.To assess this, we fit local regression models to case and control sample-level expression measurements as a function of age and computed age-specific DE

effect-sizes (**Fig S10**). Of 4821 DE genes in SCZ, only 143 showed even nominal association between effect size magnitude and age. Similar associations were seen in 29 of 1119 DE genes in BD and 85 of 1611 DE genes in ASD. Consequently, this would not support substantial age-related environmental exposures as the mechanism for the vast majority of differentially expressed genes.

Using gene expression data from animal models, we investigated whether exposure to commonly used psychiatric medications could recapitulate observed gene expression changes in disease (**Fig S11**). Overall, with the exception of lithium, chronic exposure to medications including antipsychotics (clozapine, haloperidol), mood stabilizers (lamotrigine), and antidepressants had a minimal effect on the transcriptome, in most cases with no differentially expressed genes at traditional FDR thresholds (*21*). Even at more liberal thresholds, the overlap between medication-driven and disease signal remains sparse. One notable exception was a module that reflects major components of a well-described (*40*) neural activity-dependent gene expression program, whose disease relationships are refined in the network analysis section below. Finally, we do note that there are other factors that were not measured that can potentially contribute to gene expression variation in post-mortem tissue, including agonal events and pH (*22*, *41*, *42*) in addition to those measured and used as covariates, such as RNA integrity and post mortem interval (PMI). We used surrogate variable correction in our analyses to account for such unmeasured confounders (*43*), which is a standard approach (*44*).

Transcriptome-wide Association

We next sought to leverage this transcriptome-wide dataset to prioritize potential risk genes whose brain expression may be altered by disease-associated genetic variation. We first assessed whether gene or isoform level expression measures were significantly associated with PRS for each disorder (*21*), identifying 45 genes or isoforms whose expression was significantly associated with PRS (FDR < 0.05), including 32 in ASD, 2 in BD, and 11 in SCZ (**Fig 3C**; **Table S4**). In ASD, the majority of associations map to 17q21.31, which harbors a known common inversion polymorphism and rare deleterious structural variants associated with intellectual disability (*45*). PRS for BD was associated with isoforms of the neuronal calcium sensor *NCALD* and *SNF8*, an endosomal sorting protein. In SCZ, we identify upregulation of the established risk gene *C4A* as the most significant PRS association (*5*). Concordantly, we find a strong positive correlation between *C4A* expression and genetically imputed *C4A* copy number ($R=0.25$, $P=7x10^{-14}$), as well as with imputed number of *C4-HERV* elements ($R=0.24$, $P=1.2x10^{-12}$), but not *C4B* copy number ($R=0.006$, $P=0.85$) (*21*). Additional associations with PRS were observed in the MHC region in SCZ, which harbors the largest GWAS association comprised of multiple independent signals (*5*), but is difficult to parse due to complex patterns of LD. These included two lncRNAs, *HCG17* and *HCG23,* as well as the MHC class I heavy chain receptor *HLA-C*. However, expression of all three of these genes were also significantly ($P<0.05$) correlated with *C4A* copy number, indicating a need for further validation given the complex LD structure in this region.

Taking an orthogonal approach to integrate genetic information, we performed a formal transcriptome-wide association study (TWAS; (*46*)) to directly identify those genes whose *cis*-regulated expression is associated with disease (*21*). Compared with the PRS-based approach, TWAS is restricted to *cis* effects on expression and genes with evidence of heritable expression patterns in our dataset. TWAS and related methods have the advantage of collapsing signals onto specific genes, reducing multiple comparisons and increasing power for association testing (*46*, *47*). Further, by imputing the *cis*-regulated heritable component of brain gene expression into the association cohort, TWAS enables direct prediction of the transcriptomic effects of disease-associated genetic variation, identifying potential mechanisms through which variants may impart risk. However, the limited size of brain eQTL datasets to date has necessitated use of non-CNS tissues to define TWAS weights (*46*), limiting identification of brain relevant genetic regulation. Given the substantial enrichment of psychiatric GWAS signal within CNS expressed regulatory elements (*39*), we reasoned that our dataset would provide substantial increased power and specificity.

We identify 14,750 genes with heritable *cis*-regulated expression in brain in the PsychENCODE cohort, enabling increased transcriptomic coverage for detection of association signal (**Fig 4**). In ASD, TWAS prioritizes 12 genes across 3 genomic loci (Bonferroni-corrected $P<0.05$; **Fig 4**). This includes the 17q21.31 region, which showed multiple PRS associations as described above, but did not reach genome-wide significance in the largest GWAS to date (*38*), highlighting the complementarity of the TWAS approach. Of the seven TWAS-significant genes at 17q21, conditional analysis prioritizes one – *LRRC37A,* which is further supported by a Hi-C interaction in fetal brain (*38*). *LRRC37A* is intriguing due to its primate-specific evolutionary expansion, loss-of-function intolerance, and expression patterns in brain and testis (*45*). However, it is also possible that common variants in GWAS are indirectly tagging the known common inversions or other recurrent structural variants previously identified at this locus (*45*). TWAS additionally prioritizes *XKR6* and *PINX1* as well as *PLK1S1* and *NKX2-2* at ASD loci on chromosomes 8 and 20, respectively (**Fig 4**; **Table S4**; (*21*)).

In BD, TWAS prioritizes 17 genes across 14 distinct loci (Bonferroni-corrected $P<0.05$; **Fig 4**; **Table S4**), none of which were DE. These included *VPS45*, *TMEM258*, *CKMT1A*, *HLF*, *ILF3*, *LMAN2L*, *DCLK3, BMPR1B*, *SNAP91*, and *SYS1-DBNDD2*. At loci with multiple hits, we applied conditional and colocalization analyses (*21*) to further finemap these regions, permitting prioritization of single top candidate genes – *CDKN2C*, *UBE2Q2L*, and *HAPLN4*. The two isoforms showing PRS associations in BD (*NCALD*, *SNF8*) were not significant in TWAS, likely due to lack of a nearby genome-wide significant locus, or due to isoform-specific regulation, suggesting those expression changes may be driven by *trans*-acting factors.

Finally, TWAS identifies 193 genes in SCZ, of which 107 remain significant following conditional analysis at each gene within multi-hit loci. Excluding the MHC region, there remained 164 significant genes representing 78 genome-wide significant GWAS loci (**Fig 4**; **Table S4**). A previous TWAS study in SCZ primarily based on non-neural tissue prioritized 157 genes, of which 37 coincide with the current results, a highly significant overlap (OR 60.7, $p<10^{-42}$, Fisher's exact test). Moreover, 60 TWAS prioritized genes overlapped with the list of 321 'high confidence' SCZ risk genes identified in the companion manuscript (*18*), identified using gene regulatory networks and a deep learning approach (OR 34.7, $p<10^{-60}$, Fisher's exact test). Twenty one genes prioritized by TWAS were also concordantly DE in SCZ brain in the same direction as predicted by TWAS (**Table S4**).

Overall, these analyses prioritize 125 candidate risk genes whose *cis* expression regulation is associated with disease. Most genes show disease-specific effects, as only three genes showed overlap between SCZ and BD TWAS, including *VPS45*, *SNAP91*, *DCLK3*, while none overlapped with ASD. Considering genes independently identified by each method separately at genome-wide significance, only 2 genes are identified by both PRS association and TWAS, 1 in ASD (*LRRC37A*) and 1 in SCZ (*LRRC37A2*), which may represent structural variation at this locus (*45*). When we restrict our analysis to those genes that are captured by TWAS and replicated by association with the PRS (Likelihood Ratio Test; $p<0.01$), 11 genes are identified in SCZ (*JKAMP*, *SETD6*, *TMEM214*, *FTSJ2*, *BTN2A2*, *CLEC18B*, *CACNA1D*, *HIST1H4L*, *HLA-DOA*, *TRIM27*, *LRRC37A2*, *RP11-350N15.5*), and one, the *LRRC37A* locus, for ASD.

Co-expression Networks Refine Shared Cross-Disorder Signals

To place transcriptomic changes within a systems-level context and more fully interrogate the specific molecular neuropathology of these disorders, we performed weighted gene correlation network analysis (WGCNA) to create independent gene and isoform-level networks (*15*, *48*, *49*), which we then assessed for disease association and GWAS enrichment using stratified LD score regression ((*21*); see Resource.PsychENCODE.org for interactive visualization). Although calculated separately, gene and isoform-level networks generally reflect equivalent biological processes, as demonstrated by hierarchical clustering (**Fig 5A**). However, the isoform-level networks captured greater detail and a larger proportion were associated with disease GWAS than gene-level networks (61% *vs* 41% with nominal GWAS enrichment,

P=0.07, χ^2; **Fig 5A**). Consistent with expectations, modules showed enrichment for gene ontology pathways and we identified modules strongly and selectively enriched for markers of all major CNS cell types (**Fig 5A**-**B**; **Fig S12**), facilitating computational deconvolution of cell type specific signatures (*15*, *48*, *50*). For ease of subsequent presentation, we group gene-isoform module pairs that co-cluster, have overlapping parent genes, and represent the same biological processes.

The large sample sizes, coupled with the specificity of isoform-level quantifications, enabled refinement of previously identified gene networks related to ASD, BD and SCZ (*1*, *2*, *15*, *16*, *19*, *51*). Of a combined 90 modules, including 34 gene- (geneM) and 56 isoform-level (isoM) modules, 61 (68%) showed significant association with at least one disorder, demonstrating the pervasive nature of transcriptome dysregulation in psychiatric disease. Five modules are shared across all three disorders, 3 up and two downregulated; 22 modules are shared by 2 of the 3 disorders, and 36 demonstrate more specific patterns of dysregulation in either ASD, SCZ or BD (**Fig 5**; **Table S5**). It is notable that of these 61 co-expression modules with a disease-association, 41 demonstrate cell type enrichments, consistent with the strong cell type disease-related signal observed using both supervised and unsupervised methods in our companion paper (*18*). This demonstrates the importance of cell type specific changes in the molecular pathology of these major psychiatric disorders; the cell type relationships defined by the disease modules substantially enhance our knowledge of these processes, as we outline below.

The five modules shared between ASD, BD and SCZ can be summarized to represent 3 distinct biological processes. Two of these processes are upregulated, including an inflammatory NFkB signaling module pair (geneM5/isoM5; further discussed in neural-immune section below), and a module (geneM31) enriched primarily for genes with roles in the postsynaptic density, dendritic compartments, and receptor mediated presynaptic signaling that are expressed in excitatory neurons, and to a lesser extent, inhibitory neurons (**Fig 5C**). Remarkably, *DCLK3*, one of the hubs of geneM31, is a genome-wide significant TWAS hit in both SCZ and BD. The third biological process, geneM26/isoM22 (**Fig 5C**), is downregulated, and enriched for endothelial and pericyte genes, with hubs that represent markers of the blood-brain barrier, including *ITIH5*, *SLC38A5*, *ABCB1,* and *GPR124*, a critical regulator of brain-specific angiogenesis *(52, 53)*. This highlights specific, shared alterations in neuronal-glial-endothelial interactions across these neuropsychiatric disorders.

In contrast to individual genes or isoforms, no modules were significantly associated with PRS scores after multiple-testing correction. However, 19 modules were significantly (FDR < 0.05) enriched for SNP-heritability based on published GWAS ((*21*)**; Fig 5A**; **Fig S13**). A notable example is geneM2/isoM13, which is enriched for oligodendrocyte markers and neuron projection developmental pathways and is downregulated in ASD and SCZ, with a trend in BD (**Fig 5C**). This module pair showed the greatest overall significance of enrichment for SCZ and educational attainment GWAS, and was also enriched in BD GWAS to a lesser degree, suggesting that the processes represented by geneM2/isoM13 genes play a causal role in SCZ and BD disease risk. As additional causal evidence, this module is enriched for genes harboring ultra-rare variants identified in SCZ (*54*) (**Fig S13**). Finally, we also observe pervasive and distinct enrichments for syndromic genes and rare variants identified through whole exome sequencing in individuals with neurodevelopmental disorders (**Table S5**; **Fig S13**).

Neuronal Isoform Networks Capture Disease Specificity

Multiple neuronal and synaptic signaling pathways have been previously demonstrated to be downregulated in a diminishing gradient across ASD, SCZ, and BD brains without identification of clear disease-specific signals for these neuronal-synaptic gene sets (*1*, *2*, *16*, *19*, *55*, *56*). We do observe neuronal modules broadly dysregulated across multiple disorders, including a neuronal/synaptic module (isoM18) with multiple isoforms of the known ASD risk gene, *ANK2,* as hubs. However, the large sample size, coupled with the specificity of isoform-level qualifications, enabled us to identify synaptic modules containing unique

isoforms with distinct disease associations and to separate distinct signals from excitatory and inhibitory neurons (**Fig 5B**).

A particularly salient example of differential module membership and disease association of transcript isoforms is *RBFOX1,* a major neuronal splicing regulator implicated across multiple neurodevelopmental and psychiatric disorders (*16*, *32*, *57*, *58*). Previous work has identified downregulated neuronal modules in ASD and SCZ containing *RBFOX1* as a hub (*1*, *16*). Here, we identify two neuronal modules with distinct *RBFOX1* isoforms as hub genes (**Fig 6A**). The module pair geneM1/isoM2, downregulated only in ASD (**Fig 6B**), contains the predominant brain-expressed *RBFOX1* isoform (*44*) and includes several cation channels (e.g., *HCN1*, *SCN8A*). The second most abundant *RBFOX1* isoform is in another module, isoM17, which is downregulated in both ASD and SCZ (**Fig 6B**). Experiments in mouse indicate that *RBFOX1* has distinct nuclear and cytoplasmic isoforms with differing functions, the nuclear isoform primarily regulating pre-mRNA alternative splicing, and the cytoplasmic isoform binding to the 3' UTR to stabilize target transcripts involved in regulation of neuronal excitability (*28*, *32*, *57*, *59*). Here, we find that isoM17 shows greater enrichment for nuclear *RBFOX1* targets (**Fig 6C**), whereas isoM2 shows stronger overlap with cytoplasmic targets (*32*). Consistent with a predicted splicing-regulatory effect, isoM17 shows greater enrichment for genes exhibiting DS in ASD and SCZ (**Fig 6D**). In accordance with a predicted role in regulating excitability, isoM2 shows strong and selective enrichment for epilepsy risk genes (**Fig 6E**). Moreover, the two modules show differential association with common genetic risk (**Fig 6E**), with isoM2 exhibiting GWA signal enrichment across SCZ, BD, and MDD. This widespread enrichment of neurodevelopmental and psychiatric disease risk factors -- from rare variants in epilepsy to common variants in BD, SCZ, and MDD – is consistent with a model where broad neuropsychiatric liability emanates from myriad forms of dysregulation in neuronal excitability, all linked via *RBFOX1*. These results highlight the importance of further studies focused on understanding the relationship between human *RBFOX1* transcript diversity and functional divergence, as most of what is known is based on mouse, and the human shows far greater transcript diversity (*32*, *57*, *60*).

Previous transcriptional networks related to ASD, BD and SCZ did not separate inhibitory and excitatory neuron signals (*1*). The increased resolution here allowed us to identify several modules enriched in inhibitory interneuron markers (**Fig 5B**), including geneM23/isoM19, which is downregulated in ASD and SCZ, with a trend toward downregulation observed in BD; downsampling in the SCZ dataset suggests that the lack of significance in BD may be due to smaller sample size (**Fig S14**). This module pair contained as hubs the two major GABA synthesizing enzymes (*GAD1*, *GAD2)*, multiple GABA transporters (*SLC6A1*, *SLC24A3*), many other known interneuron markers (*RELN*, *VIP*), as well as *DLX1* and the lncRNA *DLX6-AS1*, both critical known regulators of inhibitory neuron development (*61*). This inhibitory neuron-related module is not enriched for common or rare genetic disease-associated variation, although other studies have found enrichment for SCZ GWAS signal among interneuron markers defined in other ways (*62*).

Several neuronal modules that distinguish between the disorders differentiate BD and SCZ from ASD, including the module pair geneM21/isoM30 (**Fig 5C**), which captures known elements of activity-dependent neuronal gene regulation, whose hubs include classic early-response (*ARC*, *EGR1*, *NPAS4*, *NR4A1)* and late-response genes (*BDNF, HOMER1*) (*40*). Although these modules were not significantly downregulated in ASD, subsampling indicates that the differences between disorders could be driven by sample size (**Fig S14**). These genes play critical roles in regulating synaptic plasticity and the balance of excitatory and inhibitory synapses (*40*). Remarkably, a nearly identical module was recently identified as a sex-specific transcriptional signature of major depression and stress susceptibility (*63*). Since psychiatric drug use is more prevalent in SCZ and BD than ASD, and the geneM21/isoM30 module pair are altered more substantially in these disorders, we explored whether these modules may be affected by medication exposure. Indeed, geneM21/isoM30 was associated with genes downregulated by chronic high-doses (but not low-doses) of haloperidol, as well as genes upregulated by the antidepressant fluoxetine (**Fig S11A**). Furthermore, geneM21/isoM30 expression was negatively correlated with the degree of lifetime antipsychotic exposure in the subset of patients for whom these data were available (P=0.001, Pearson; **Fig S11B**). As such, it will be

worthwhile to determine whether this module is a core driver of the therapeutic response, as has been suggested (*64*). Other neuronal modules distinguished SCZ and BD from ASD (**Fig 5B**), including geneM7, enriched for synaptic and metabolic processes with the splicing regulator *NOVA2* (**Fig 5C**). This neuronal module was significantly enriched for both BD and SCZ GWAS signals, supporting a causal role for this module.

Distinct Trajectories of Neural-Immune Dysregulation

Previous work has identified differential activation of glial and neural-immune processes in brain from patients with psychiatric disorders (*16*, *51*, *56*, *65*–*68*), including upregulation of astrocytes in SCZ and BD (*1*, *56*) and both microglia and astrocytes in ASD (*19*, *69*). Evidence supports hyperactive complement-mediated synaptic pruning in SCZ pathophysiology, presumably through microglia (*5*), although post-mortem microglial upregulation was observed only in ASD (*19*, *69*). We examined whether our large cohort including ~1000 control brains, capturing an age range from birth to 90 years, would enable refinement of the nature and timing of this neuroinflammatory signal and potential relationship to disease pathogenesis (**Fig 7A**). Four modules were directly related to neural-immune processes (**Fig 7A**-**C**), two of which are gene/isoform module pairs that correspond clearly to cell type specific gene expression; one representing microglia (geneM6/isoM15) and the other astrocytes (geneM3/isoM1), as they are strongly and selectively enriched for canonical cell type specific marker genes (**Fig 7C-E**). Two additional immune-related modules appear to represent more broadly expressed signaling pathways: interferon response (geneM32) and NFkB (geneM5/isoM5). The interferon response module (geneM32) contains critical components of the IFN-stimulated gene factor 3 (ISGF3) complex that activates the transcription of downstream interferon-stimulated genes (ISGs), which comprise a striking 59 of the 61 genes in this module (*70*). The NF*k*B module pair (geneM5/isoM5) includes four out of five of the NFkB family members (*NFkB1*, *NFkB2*, *REL*, *RELA*), as well as many downstream transcription factor targets and upstream activators of this pathway.

The dynamic trajectories of these processes in cases with respect to controls reveal distinct patterns across disorders that coincide with disease course (**Fig 7F**). The IFN-response and microglial modules are most strongly upregulated in ASD, peaking during early development, coincident with clinical onset. In contrast, in SCZ and BD, the microglial module is actually downregulated and driven by a later dynamic decrease, dropping below controls after age 30. The NF*k*B module, which is upregulated across all three disorders, maximally diverges from controls during early adulthood, coincident with typical disease onset in SCZ and BD (~25). Accordingly, this NF*k*B module contained *C4A* – the top GWAS-supported, and strongly upregulated, risk gene for SCZ (*5*). This pattern is clearly distinct from ASD, which shows a dynamic trajectory, but remains upregulated throughout (**Fig 7F**).

Non-coding Modules and lncRNA Regulatory Relationships

Given that many lncRNAs are predicted to have transcriptional regulatory roles, we next assessed whether mRNA-based co-expression networks could provide additional functional annotation for ncRNAs. As a subset of lncRNAs are thought to function by repressing mRNA targets (*71*), we applied csuWGCNA (*72*) to identify potential regulatory relationships (*21*). We identified 39 modules (csuM) using csuWGCNA, all preserved in the signed networks with strong cell type and GWAS enrichments, which captured 7186 negatively correlated lncRNA-mRNA pairs within the same module (**Fig S15**). We provide a table of putative mRNA targets for these brain expressed lncRNAs, including 209 exhibiting DE in ASD, 122 in BD and 241 in SCZ (**Table S6**).

A salient example of the power of this approach for functional annotation is *LINC00473,* a hub of the neuronal activity dependent gene regulation module (geneM21/isoM30; **Fig 5C**). Expressed in excitatory neurons and downregulated in SCZ (log_2FC -0.16, FDR<0.002), *LINC00473* is regulated by synaptic activity and downregulates immediate early gene expression (*73*), consistent with its hub status in this module.

Similarly, we identify the lncRNA *DLX6-AS1,* a known development regulator of interneuron specification (*61*), as the most central hub gene in the interneuron module (geneM23/isoM19), which is downregulated in ASD and SCZ. This interneuron module also contains *LINC00643* and *LINC01166*, two poorly annotated, brain enriched lncRNAs. *LINC00643* is downregulated in SCZ (log_2FC -0.06, FDR=0.04) whereas *LINC01166* is significantly downregulated in BD (log_2FC -0.17, FDR<0.05) with trends in ASD and SCZ (FDR's < 0.1). Our data suggest a role for these lncRNAs in interneuron development, making them intriguing candidates for follow-up studies. Using fluorescence *in situ* hybridization (FISH), we confirmed that both *LINC00643* and *LINC1166* are expressed in GAD1+ GABAergic neurons in area 9 of adult brain, present both in the cell nucleus and cytoplasm (**Fig 8A**; **Fig S16**), although expression was also detected in other non GAD1+ neurons as well.

Multiple ncRNAs including *SOX2-OT, MIAT*, and *MEG3* are enriched in oligodendrocyte modules (geneM2/isoM13/csuM1; **Fig 5C**) that are downregulated in both SCZ and ASD. *SOX2-OT* is a heavily spliced, evolutionarily-conserved lncRNA exhibiting predominant brain expression and a hub of these oligodendrocyte modules, without previous mechanistic links to myelination (*74*, *75*). The lncRNAs *MIAT* and *MEG3* are negatively correlated with most of the hubs in this module, including *SOX2-OT* (**Fig S15**). *MIAT* is also known to interact with *QKI,* an established regulator of oligodendrocyte-gene splicing also located in this module (*76*, *77*). These analyses predict critical roles for these or these often overlooked non-coding genes in oligodendrocyte function (*76*, *77*) and potentially in psychiatric conditions.

Isoform Network Specificity and Switching

To more comprehensively assess whether aspects of disease specificity are conferred by alternative transcript usage or splicing, versus DE, we surveyed genes exhibiting DTU across disorders (*21*). We identified 134 such 'switch isoforms', corresponding to 64 genes displaying different DTU between ASD and SCZ (**Table S7**). As an example, isoforms of *SMARCA2, a* member of the BAF-complex strongly implicated in several neurodevelopmental disorders including ASD (*78*), are up and downregulated in ASD and SCZ, respectively (**Fig S17**). Conversely, the isoforms of *NIPBL,* a gene associated with Cornelia de Lange Syndrome (*79*) are down and upregulated in ASD and SCZ, respectively (**Fig S17**). Such opposing changes in isoform expression of various genes may represent differences in disease progression or symptom manifestation in diseases as ASD and SCZ, mediated by genetic risk variants that create subtle differences in isoforms within the same gene that exhibit distinct biological effects in each disorder. A remarkable example is the ASD risk gene *ANK2* (*80*), whose two alternatively spliced isoforms, *ANK2-006* and *ANK2-013*, are differentially regulated in SCZ and ASD (**Fig 8B**). These switch isoforms show markedly different expression patterns, belonging to different co-expression modules, geneM3/isoM1 (**Fig 7C**) and isoM18, which are enriched in astrocyte and neuronal cell types, respectively (**Fig 5A**; **Fig S12**). The protein domain structure of these transcripts is also non-overlapping, with *ANK2-006* carrying exclusively ZU5 and DEATH domains, and *ANK2-013* carrying exclusively ankyrin repeat domains (**Fig 8C**). Both isoforms are impacted by a *de novo* ASD CNV, and *ANK-006* also carries *de novo* mutations from neurodevelopmental disorders. Both isoforms bind to the neuronal cell adhesion molecular *NRCAM*, but *ANK2-013* has two additional, unique partners – *TAF9* and *SCN4B* (**Fig 8D**), likely cell type specific interactions that suggest distinct functions of the isoforms of this genes in different neural cell types and diseases.

Several studies have demonstrated that genes carrying microexons are preferentially expressed in brain and their splicing is dysregulated in ASD (*30*, *81*, *82*). This PsychENCODE sample provided the opportunity to assess the role of microexons in a far larger cohort and across several disorders. Indeed, we find that switch isoforms with microexons (3-27 bp) are significantly enriched in both ASD (FDR=0.03) and SCZ (FDR=0.03, logistic regression) (**Fig 8E**; (*21*)). Genes with switch isoforms are also enriched for the regulatory targets of two ASD risk genes, *CHD8* and *FMRP*, as well as highly mutationally constrained genes (pLI>0.99), syndromic ASD genes, and in genes with *de novo* exonic mutations in ASD, SCZ and BD (**Fig 8F**; **Table S7**;

(*21*)). These data confirm the importance of microexon regulation in neuropsychiatric disorders beyond ASD, and their potential role in distinguishing among biological pathways differentially affected across conditions. This role for microexons further highlights local splicing regulation as a potential mechanism conferring key aspects of disease specificity, extending the larger disease signal observed at the isoform-level in co-expression and differential expression analyses.

Discussion

We present a large-scale RNA-Seq analysis of the cerebral cortex across three major psychiatric disorders, including extensive analyses of the non-coding and alternatively spliced transcriptome, as well as gene- and isoform-level co-expression networks. The scope and complexity of these data do not immediately lend themselves to simple mechanistic reduction. Nevertheless, at each level of analysis, we present concrete examples that provide proofs-of-principle and starting points for investigations targeting shared and distinct disease mechanisms to connect causal disease drivers with brain-level perturbations.

Broadly, we find that isoform-level changes exhibit the largest effect sizes in disease brain, are most enriched for genetic risk, and provide the greatest disease specificity when assembled into co-expression networks. Remarkably, disturbances in the expression of distinct isoforms of more than 50 genes are differentially observed in SCZ and ASD, which in the case of the ASD risk gene *ANK*2, is predicted to affect different cell types in each disorder. Moreover, we observe disease-associated changes in the splicing of dozens of RNA-binding proteins and splicing factors, most of whose targets and functions are unknown. Similarly, nearly 1000 ncRNAs are dysregulated in at least one disorder and most of these ncRNAs show significant CNS enrichment, but until now, have limited functional annotation.

As with any case/control association study, multiple potential factors contribute to gene expression changes in post-mortem human brain, many of which may represent reactive processes. At each step of analysis, we have attempted to mitigate the contribution of these factors through known and hidden covariate correction, assessment of age trajectories, and via enrichment for causal genetic variation. Supporting the generalizability of our findings, we find highly significant correlations of the $\log_2$FC between randomly split halves of the data (**Fig S3**). This likely varies by transcript class, and some of the modest correlations are likely due to low abundance genes, such as ncRNAs, which we prefer to include, while recognizing the inherent tension between expression level and measurement accuracy. We provide access to this extensive resource, both in terms of raw and processed data and as browsable network modules (Resource.PsychENCODE.org).

Several broad shared patterns of gene expression dysregulation have been observed in post mortem brain in previous studies, most prominently, a gradient of downregulation of neuronal and synaptic signaling genes, and upregulation of glial-immune or neuroinflammatory signals. Here, we are able to substantially refine these signals, by distinguishing both up and downregulated neuron-related processes that are differentially altered across these three disorders. Furthermore, we extend previous work that identified broad neuroinflammatory dysregulation in SCZ, ASD, and BD, by identifying specific pathways involving IFN-response, NFkB, astrocytes and microglia that manifest distinct temporal patterns across conditions. A module enriched for microglial-associated genes, for example, shows a clear distinction between disorders, with strong upregulation observed on ASD and significant downregulation in SCZ and BD. Overall, these results provide substantially increased specificity to the observations that ASD, BD, and SCZ are associated with elevated neuroinflammatory processes (*68*, *83*–*85*).

This work highlights isoform-level dysregulation as a critical, and relatively underexplored, proximal mechanism linking genetic risk factors with psychiatric disease pathophysiology. In contrast to local splicing changes, isoform-level quantifications require imputation from short-read RNA-Seq data guided by existing genomic annotations. Consequently, the accuracy of these estimates is hindered by incomplete annotations, as well as by limitations of short-read sequencing, coverage, and genomic biases like GC content (*86*, *87*). This may be particularly problematic in brain where alternative splicing patterns are more distinct than in other

organ systems (*81*). We present experimental validations for several specific isoforms, but try to focus on the class of dysregulated isoforms, and the modules and biologically processes they represent, rather than individual cases which may be more susceptible to bias. Longer read sequencing, which provides a more precise means for isoform quantification, will be of great utility as it becomes more feasible at scale.

By integrating transcriptomic data with genetic variation, we identify multiple disease-associated co-expression modules enriched for causal variation, as well as new mechanisms potentially underlying specific disease loci in each of the diseases. In parallel, by performing a well-powered brain-relevant TWAS in SCZ, and to a lesser extent in BD and ASD, we are further able to elucidate candidate molecular mechanisms through which disease-associated variants may act. TWAS prioritizes dozens of new candidate disease genes, including many dysregulated in disease brain. Similar to the eQTLs identified in the companion paper (*18*), the majority of these new loci do not overlap with disease GWAS association signals. Rather, most are outside of the LD block and quite distal to the original association signal, highlighting the importance of orthogonal functional data types, such as transcriptome or epigenetic data (*17*, *47*, *81*, *88*, *89*), in deciphering the underlying mechanisms of disease-associated genetic effects.

A large proportion of disease-associated co-expression modules are enriched for cell type specific markers, as is overall disease DE signal, indicating that transcriptomic alterations in disease are likely driven substantially by (even subtle) shifts in cell type proportions, or cell type specific pathways, consistent with our previous observations (*1*) and those in the companion PsychENCODE manuscript (*18*). Functional genomic studies often remove such cell type-specific signals, through use of large numbers of expression-derived principle components or surrogate variables as covariates, to remove unwanted sources of variation and maximize detection of *cis* eQTLs (*44*). We retain the cell type-specific signals as much as possible, reasoning that cell type-related alterations may directly inform the molecular pathology of disease in psychiatric disorders, in which there is no known microscopic or macroscopic pathology. This rationale is supported by the consistent observation of the dynamic and disease-specific microglial upregulation observed in ASD, and the shared astrocyte upregulation in SCZ and ASD. This approach, however, reduces the ability to detect genetic enrichment from GWAS, as current methods predominately capture *cis*-acting regulatory effects. The modesty of genetic enrichments among disease-associated transcriptomic alterations may also indicate that gene expression changes reflect an indirect cascade of molecular events triggered by environmental as well as genetic factors, or that genetic factors may act earlier such as during development.

Finally, these data, while providing a unique, large-scale resource for the field, also suggest that profiling additional brains, especially from other implicated brain regions from patients will continue to be informative. Similarly, these data suggest that isoform level analyses including the identification of isoform-specific PPI and cell type specificity, while posing major challenges for high-throughput studies, are likely to add substantial value to understanding brain function and neuropsychiatric disorders. Finally, as GWAS studies in ASD and BD increase in size and subsequently in power, their continued integration with these transcriptome data will likely prove critical in identifying the functional impact of disease-associated genetic variation.

References and Notes

1. M. J. Gandal *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. **359**, 693–697 (2018).

2. M. Fromer *et al.*, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).

3. H. A. Whiteford, A. J. Ferrari, L. Degenhardt, V. Feigin, T. Vos, The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One*. **10**, e0116820 (2015).

4. M. J. Gandal, V. Leppa, H. Won, N. N. Parikshak, D. H. Geschwind, The road to precision psychiatry: translating genetics into disease mechanisms. *Nat. Neurosci.* **19**, 1397–1407 (2016).

5. A. Sekar *et al.*, Schizophrenia risk from complex variation of complement component 4. *Nature*. **530**, 177–183 (2016).

6. S. J. Sanders, First glimpses of the neurobiology of autism spectrum disorder. *Curr. Opin. Genet. Dev.* **33**, 80–92 (2015).

7. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science*. **337**, 1190–1195 (2012).

8. L. D. Ward, M. Kellis, Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012).

9. A. Visel, E. M. Rubin, L. A. Pennacchio, Genomic views of distant-acting enhancers. *Nature*. **461**, 199–205 (2009).

10. GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. **348**, 648–660 (2015).

11. S. K. Reilly *et al.*, Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science*. **347**, 1155–1159 (2015).

12. The ENCODE Project Consortium, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. **447**, 799 (2007).

13. Roadmap Epigenomics Consortium *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature*. **518**, 317–330 (2015).

14. R. Andersson *et al.*, An atlas of active enhancers across human cell types and tissues. *Nature*. **507**, 455–461 (2014).

15. N. N. Parikshak, M. J. Gandal, D. H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).

16. I. Voineagu *et al.*, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. **474**, 380–384 (2011).

17. PsychENCODE Consortium *et al.*, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).

18. Daifeng Wang, Shuang Liu, Jonathan Warrell, Hyejung Won, Xu Shi, Fabio Navarro, Declan Clarke, Mengting Gu, Prashant Emani, Min Xu, Yucheng T. Yang, Jonathan J. Park, Suhn Kyong Rhie, Kasidet Manakongtreecheep, Holly Zhou, Aparna Nathan, Jing Zhang, Mette Peters, Eugenio Mattei, Dominic Fitzgerald, Tonya Brunetti, Jill Moore, PsychENCODE Consortium, Nenad Sestan, Andrew E. Jaffe, Kevin White, Zhiping Weng, Daniel H. Geschwind, James Knowles, Mark Gerstein, Comprehensive functional genomic resource and integrative model for the adult brain. *Submitted* (2018).

19. N. N. Parikshak *et al.*, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. **540**, 423–427 (2016).

20. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature*. **478**, 483–489 (2011).

21. See supplemental methods.

22. A. E. Jaffe *et al.*, qSVA framework for RNA quality correction in differential expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7130–7135 (2017).

23. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **12**, 323 (2011).

24. A. E. Jaffe *et al.*, Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).

25. Y. I. Li *et al.*, RNA splicing is a primary link between genetic variation and disease. *Science*. **352**, 600–604 (2016).

26. P. J. Batista, H. Y. Chang, Long noncoding RNAs: cellular address codes in development and disease. *Cell*. **152**, 1298–1307 (2013).

27. J. di Iulio *et al.*, The human noncoding genome defined by genetic diversity. *Nat. Genet.* (2018), doi:10.1038/s41588-018-0062-7.

28. C. K. Vuong, D. L. Black, S. Zheng, The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016).

29. Y. I. Li *et al.*, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

30. M. Irimia *et al.*, A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. **159**, 1511–1523 (2014).

31. N. Akula *et al.*, RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol. Psychiatry*. **19**, 1179–1185 (2014).

32. J.-A. Lee *et al.*, Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron*. **89**, 113–128 (2016).

33. J. C. Darnell *et al.*, FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. **146**, 247–261 (2011).

34. E. Sebestyén *et al.*, Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26**, 732–744 (2016).

35. J. Gauthier *et al.*, Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum. Genet.* **130**, 563–573 (2011).

36. F. F. Hamdan *et al.*, Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).

37. B. Treutlein, O. Gokce, S. R. Quake, T. C. Südhof, Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1291–E1299 (2014).

38. J. Grove *et al.*, Common risk variants identified in autism spectrum disorder. *bioRxiv* (2017), p. 224774.

39. H. K. Finucane *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

40. A. E. West, M. E. Greenberg, Neuronal Activity–Regulated Gene Transcription in Synapse Development and Cognitive Function. *Cold Spring Harb. Perspect. Biol.* **3** (2011), doi:10.1101/cshperspect.a005744.

41. Y. Zhu, L. Wang, Y. Yin, E. Yang, Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* **7**, 5435 (2017).

42. J. Z. Li *et al.*, Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum. Mol. Genet.* **13**, 609–616 (2004).

43. J. T. Leek, J. D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS*

Genet. **3**, 1724–1735 (2007).

44. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature*. **550**, 204 (2017).

45. M. C. Zody *et al.*, Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat. Genet.* **40**, 1076–1083 (2008).

46. A. Gusev *et al.*, Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

47. E. R. Gamazon *et al.*, A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).

48. M. C. Oldham *et al.*, Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).

49. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).

50. J. A. Miller, M. C. Oldham, D. H. Geschwind, A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* **28**, 1410–1420 (2008).

51. C. Chen *et al.*, Two gene co-expression modules differentiate psychotics and controls. *Mol. Psychiatry*. **18**, 1308–1314 (2013).

52. R. Daneman *et al.*, The mouse blood-brain barrier transcriptome: a new resource for understanding the development and function of brain endothelial cells. *PLoS One*. **5**, e13741 (2010).

53. B. Obermeier, R. Daneman, R. M. Ransohoff, Development, maintenance and disruption of the blood-brain barrier. *Nat. Med.* **19**, 1584–1596 (2013).

54. G. Genovese *et al.*, Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).

55. S. E. Ellis, R. Panitch, A. B. West, D. E. Arking, Transcriptome analysis of cortical tissue reveals shared sets of downregulated genes in autism and schizophrenia. *Transl. Psychiatry*. **6**, e817 (2016).

56. R. C. Ramaker *et al.*, Post-mortem molecular profiling of three psychiatric disorders. *Genome Med.* **9**, 72 (2017).

57. C. K. Vuong *et al.*, Rbfox1 Regulates Synaptic Transmission through the Inhibitory Neuron-Specific vSNARE Vamp1. *Neuron*. **98**, 127–141.e7 (2018).

58. A. F. Pardiñas *et al.*, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* (2018), doi:10.1038/s41588-018-0059-2.

59. L. T. Gehman *et al.*, The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).

60. B. L. Fogel *et al.*, RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).

61. A. M. Bond *et al.*, Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Neurosci.* **12**, 1020–1027 (2009).

62. N. G. Skene *et al.*, Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).

63. B. Labonté *et al.*, Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).

64. A. de Bartolomeis *et al.*, Immediate-Early Genes Modulation by Antipsychotics: Translational Implications for a Putative Gateway to Drug-Induced Long-Term Brain Changes. *Front. Behav. Neurosci.* **11**, 240 (2017).

65. R. Birnbaum *et al.*, Investigating the neuroimmunogenic architecture of schizophrenia. *Mol. Psychiatry* (2017), doi:10.1038/mp.2017.89.

66. S. G. Fillman *et al.*, Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol. Psychiatry*. **18**, 206–214 (2013).

67. R. Pacifico, R. L. Davis, Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Mol. Psychiatry*. **22**, 441–449 (2017).

68. J. S. Rao, G. J. Harry, S. I. Rapoport, H. W. Kim, Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients. *Mol. Psychiatry*. **15**, 384–392 (2010).

69. S. Gupta *et al.*, Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014).

70. I. Rusinova *et al.*, Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41**, D1040–6 (2013).

71. A. Necsulea *et al.*, The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. **505**, 635–640 (2014).

72. R. Dai, Y. Xia, C. Liu, C. Chen, csuWGCNA: a combination of signed and unsigned WGCNA to capture negative correlations. *bioRxiv* (2018), p. 288225.

73. P. Pruunsild, C. P. Bengtson, H. Bading, Networks of Cultured iPSC-Derived Neurons Reveal the Human Synaptic Activity-Regulated Adaptive Gene Program. *Cell Rep.* **18**, 122–135 (2017).

74. P. P. Amaral *et al.*, Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. *RNA*. **15**, 2013–2027 (2009).

75. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 716–721 (2008).

76. K. Aberg, P. Saetre, N. Jareborg, E. Jazin, Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7482–7487 (2006).

77. G. Barry *et al.*, The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry*. **19**, 486–494 (2014).

78. M. Koga *et al.*, Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum. Mol. Genet.* **18**, 2483–2494 (2009).

79. I. D. Krantz *et al.*, Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat. Genet.* **36**, 631–635 (2004).

80. S. J. Sanders *et al.*, Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. **87**, 1215–1233 (2015).

81. M. Melé *et al.*, Human genomics. The human transcriptome across tissues and individuals. *Science*. **348**, 660–665 (2015).

82. M. Quesnel-Vallières *et al.*, Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol. Cell*. **64**, 1023–1034 (2016).

83. M. L. Estes, A. K. McAllister, Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 469–486 (2015).

84. U. Meyer, J. Feldon, O. Dammann, Schizophrenia and autism: both shared and disorder-specific pathogenesis via perinatal inflammation? *Pediatr. Res.* **69**, 26R–33R (2011).

85. J. D. Rosenblat *et al.*, Inflammation as a neurobiological substrate of cognitive impairment in bipolar disorder: Evidence, pathophysiology and treatment implications. *J. Affect. Disord.* **188**, 149–159 (2015).

86. T. Steijger *et al.*, Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods*. **10**, 1177–1184 (2013).

87. M. I. Love, J. B. Hogenesch, R. A. Irizarry, Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat. Biotechnol.* **34**, 1287–1291 (2016).

88. T. Lappalainen, Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.* **25**, 1427–1431 (2015).

89. M. Brandt, T. Lappalainen, SnapShot: Discovering Genetic Regulatory Variants by QTL Analysis. *Cell*. **171**, 980–980.e1 (2017).

90. D. M. Ruderfer *et al.*, Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*. **173**, 1705–1715.e16 (2018).

91. B. B. Lake *et al.*, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* (2017), doi:10.1038/nbt.4038.

92. T. Goldmann *et al.*, Origin, fate and dynamics of macrophages at central nervous system interfaces. *Nat. Immunol.* **17**, 797–805 (2016).

93. H. Keren-Shaul *et al.*, A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell*. **169**, 1276–1290.e17 (2017).

94. A. Zeisel *et al.*, Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. **347**, 1138–1142 (2015).

95. Y. Zhang *et al.*, Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*. **89**, 37–53 (2016).

96. B. Wilkinson *et al.*, The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Transl. Psychiatry*. **5**, e568 (2015).

97. K. E. Samocha *et al.*, A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).

98. I. Iossifov *et al.*, Low load for disruptive mutations in autism genes and their biased transmission. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5600–7 (2015).

99. K. J. Karczewski *et al.*, The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* **45**, D840–D845 (2017).

100. D. Zhang *et al.*, Genetic control of individual differences in gene-specific methylation in human brain. *Am. J. Hum. Genet.* **86**, 411–419 (2010).

101. A. E. Jaffe *et al.*, Developmental And Genetic Regulation Of The Human Cortex Transcriptome In Schizophrenia. *bioRxiv* (2017), p. 124321.

102. M. C. Oldham, P. Langfelder, S. Horvath, Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst. Biol.* **6**, 63 (2012).

103. J. Reimand, T. Arak, J. Vilo, g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* **39**, W307–W315 (2011).

104. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. **536**, 285–291 (2016).

105. B. B. Lake *et al.*, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*. **352**, 1586–1590 (2016).

106. S. Darmanis *et al.*, A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7285–7290 (2015).

107. J. P. Doyle *et al.*, Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*. **135**, 749–762 (2008).

108. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

109. F. Hahne, R. Ivanek, Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).

110. I. Letunic, P. Bork, 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).

111. R. D. Finn *et al.*, The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–85 (2016).

112. A. G. Baltz *et al.*, The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*. **46**, 674–690 (2012).

113. A. Castello *et al.*, Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. **149**, 1393–1406 (2012).

114. S. C. Kwon *et al.*, The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013).

115. E. L. Huttlin *et al.*, The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*. **162**, 425–440 (2015).

116. T. S. Keshava Prasad *et al.*, Human Protein Reference Database--2009 update. *Nucleic Acids Res.* **37**, D767–72 (2009).

117. K. Lage *et al.*, A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).

118. J. Das, H. Yu, HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* **6**, 92 (2012).

119. A. Chatr-Aryamontri *et al.*, The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).

120. K. Zuberi *et al.*, GeneMANIA prediction server 2013 update. *Nucleic Acids Res.* **41**, W115–22 (2013).

121. D. Szklarczyk *et al.*, The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

122. A. Ruepp *et al.*, CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* **38**, D497–D501 (2010).

123. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

124. A. Siepel *et al.*, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

125. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

126. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. **155**, 1008–1021 (2013).

127. M. J. Mason, G. Fan, K. Plath, Q. Zhou, S. Horvath, Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*. **10**, 327 (2009).

128. C. S. Benton *et al.*, Evaluating genetic markers and neurobiochemical analytes for fluoxetine response using a panel of mouse inbred strains. *Psychopharmacology* . **221**, 297–315 (2012).

129. N. R. Wray *et al.*, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* (2018), doi:10.1038/s41588-018-0090-3.

130. A. Okbay *et al.*, Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism

identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).

131. A. Okbay *et al.*, Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. **533**, 539–542 (2016).

132. A. P. Morris *et al.*, Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

133. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. **511**, 421 (2014).

134. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism*. **8**, 21 (2017).

135. Psychiatric GWAS Consortium Bipolar Disorder Working Group, Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).

136. International HapMap 3 Consortium *et al.*, Integrating common and rare genetic variation in diverse human populations. *Nature*. **467**, 52–58 (2010).

137. B. J. Vilhjálmsson *et al.*, Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).

138. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

139. S. Mostafavi *et al.*, Normalizing RNA-sequencing data by modeling hidden covariates with prior knowledge. *PLoS One*. **8**, e68141 (2013).

140. H. T. Nguyen *et al.*, Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Med.* **9**, 114 (2017).

141. D. Polioudakis *et al.*, A single cell transcriptomic analysis of human neocortical development. *bioRxiv* (2018), p. 401885.

142. A. Untergasser *et al.*, Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

143. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

144. Mingfeng Li, Gabriel Santpere, Yuka Imamura Kawasawa, Oleg V. Evgrafov, Forrest O. Gulden, Sirisha Pochareddy, Susan M. Sunkin, Zhen Li, Yurae Shin, Robert R. Kitchen, Ying Zhu, Donna M. Werling, Andre M.M. Sousa, Hyojung Kang, Mihovil Pletikos, Jinmyung Choi, Sydney Muchnik, Xuming Xu, Daifeng Wang, Shuang Liu, Paola Giusti-Rodríguez, Christiaan A de Leeuw, Antonio Pardinas, BrainSpan Consortium, PsychENCODE Consortium: Developmental Subgroup, Ming Hu, Fulai Jin, Yun Li, Michael Owen, Michael O'Donovan, James Walters, Danielle Posthuma, Patrick Sullivan, Patt Levitt, Daniel R. Weinberger, Joel E. Kleinman, Daniel H. Geschwind, Stephan Sanders, Michael J. Hawrylycz, Matthew State, Mark B. Gerstein, Ed S. Lein, James A. Knowles, Nenad Sestan, Integrative Functional Genomic Analysis of Human Brain Development and Neuropsychiatric Risk Convergence. *Submitted* (2018).

Acknowledgements

We would like to acknowledge the National Institute of Mental Health (NIMH) for funding. Also, we acknowledge program staff, in particular T. Lehner, L. Bingaman, D. Panchision, A. Arguello and G. Senthil, for providing institutional support and guidance for this project The authors also acknowledge A Silva and M López-Aranda for helpful discussions. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, the Banner Sun Health Research Institute, the Harvard Brain Bank as part of the Autism Tissue Project (ATP), the Stanley Medical Research Institute, and the NIMH Human Brain Collection Core. Tissue for the RNAscope experiments was obtained from the Mount Sinai Neuropathology Research Core and Brain Bank.

Funding: The work is funded by the U.S. National Institute of Mental Health (NIMH) (grants P50-MH106438, D.H.G.; R01-MH094714, D.H.G.; U01-MH103339, D.H.G.; R01-MH110927, D.H.G.; R01-MH100027, D.H.G.; R01-MH110920, C.L.; U01-MH103340, C.L.; MH109885, LMI; MH104766, LMI; MH105524, LMI; MH108528, LMI; R21-MH105881, D.P.; R01-MH110555, D.P.; R01-MH109715, D.P.), the Simons Foundation for Autism Research Initiative (SFARI grant 206733, D.H.G.; SFARI grant 345469, LMI; SFARI Bridge to Independence Award, M.J.G.), the National Natural Science Foundation of China (grants 81401114, CC; 31571312, CC), the National Key Plan for Scientific Research and Development of China (2016YFC1306000, CC), the Innovation-Driven Project of Central South University (No. 2015CXS034, CC; 2018CX033, CC), and the Icahn School of Medicine at Mount Sinai computational resources by the Office of Research Infrastructure of the NIH (award number S10OD018522, D.P.). Data were generated as part of the PsychENCODE Consortium, supported by grants U01MH103339, U01MH103365, U01MH103392, U01MH103340,U01MH103346, R01MH105472, R01MH094714, R01MH105898,R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale),Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (SUNY Upstate), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman(Yale), Kevin White (UChicago), and Peter Zandi (JHU). RNA-Seq data from the CommonMind Consortium used in this study (Synapse accession no. syn2759792) was supported by funding from Takeda Pharmaceuticals Company, F. Hoffman-La Roche and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, R01MH075916, P50M096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692.

Data and Materials Availability: Study data is available on the Sage Bionetworks developed Synapse platform: https://doi.org/10.7303/syn4921369 with access governed by the NIMH Repository and Genomics Resource.

‡The PsychENCODE Consortium:

Allison E Ashley-Koch, Duke University; Gregory E Crawford, Duke University; Melanie E Garrett, Duke University; Lingyun Song, Duke University; Alexias Safi, Duke University; Graham D Johnson, Duke University; Gregory A Wray, Duke University; Timothy E Reddy, Duke University; Fernando S Goes, Johns Hopkins University; Peter Zandi, Johns Hopkins University; Julien Bryois, Karolinska Institutet; Andrew E Jaffe, Lieber Institute for Brain Development; Amanda J Price, Lieber Institute for Brain Development; Nikolay A Ivanov, Lieber Institute for Brain Development; Leonardo Collado-Torres, Lieber Institute for Brain Development; Thomas M Hyde, Lieber Institute for Brain Development; Emily E Burke, Lieber Institute for Brain Development; Joel E Kleiman, Lieber Institute for Brain Development; Ran Tao, Lieber Institute for Brain Development; Joo Heon Shin, Lieber Institute for Brain Development; Schahram Akbarian, Icahn School of Medicine at Mount Sinai; Kiran Girdhar, Icahn School of Medicine at Mount Sinai; Yan Jiang, Icahn School of Medicine at Mount Sinai; Marija Kundakovic, Icahn School of Medicine at Mount Sinai; Leanne Brown, Icahn School of Medicine at Mount Sinai; Bibi S Kassim, Icahn School of Medicine at Mount Sinai; Royce B Park, Icahn School of Medicine at Mount Sinai; Jennifer R Wiseman, Icahn School of Medicine at Mount Sinai; Elizabeth Zharovsky, Icahn School of Medicine at Mount Sinai; Rivka Jacobov, Icahn School of Medicine at Mount Sinai; Olivia Devillers, Icahn School of Medicine at Mount Sinai; Elie Flatow, Icahn School of Medicine at Mount Sinai; Gabriel E Hoffman, Icahn School of Medicine at Mount Sinai; Barbara K Lipska, Human Brain Collection Core, National Institutes of Health, Bethesda, MD; David A Lewis, University of Pittsburgh; Vahram Haroutunian, Icahn School of Medicine at Mount Sinai and James J Peters VA Medical Center; Chang-Gyu Hahn, University of Pennsylvania; Alexander W Charney, Mount Sinai; Stella Dracheva, Mount Sinai; Alexey Kozlenkov, Mount Sinai; Judson Belmont, Icahn School of Medicine at Mount Sinai; Diane DelValle, Icahn School of Medicine at Mount Sinai; Nancy Francoeur, Icahn School of Medicine at Mount Sinai; Evi Hadjimichael, Icahn School of Medicine at Mount Sinai; Dalila Pinto, Icahn School of Medicine at Mount Sinai; Harm van Bakel, Icahn School of Medicine at Mount Sinai; Panos Roussos, Mount Sinai; John F Fullard, Mount Sinai; Jaroslav Bendl, Mount Sinai; Mads E Hauberg, Mount Sinai; Lara M Mangravite, Sage Bionetworks; Mette A Peters, Sage Bionetworks; Yooree Chae, Sage Bionetworks; Junmin Peng, St. Jude Children's Hospital; Mingming Niu, St. Jude Children's Hospital; Xusheng Wang, St. Jude Children's Hospital; Maree J Webster, Stanley Medical Research Institute; Thomas G Beach, Banner Sun Health Research Institute; Chao Chen, Central South University; Yi Jiang, Central South University; Rujia Dai, Central South University; Annie W Shieh, SUNY Upstate Medical University; Chunyu Liu, SUNY Upstate Medical University; Kay S. Grennan, SUNY Upstate Medical University; Yan Xia, SUNY Upstate Medical University/Central South University; Ramu Vadukapuram, SUNY Upstate Medical University; Yongjun Wang, Central South University; Dominic Fitzgerald, The University of Chicago; Lijun Cheng, The University of Chicago; Miguel Brown, The University of Chicago; Mimi Brown, The University of Chicago; Tonya Brunetti, The University of Chicago; Thomas Goodman, The University of Chicago; Majd Alsayed, The University of Chicago; Michael J Gandal, University of California, Los Angeles; Daniel H Geschwind, University of California, Los Angeles; Hyejung Won, University of California, Los Angeles; Damon Polioudakis, University of California, Los Angeles; Brie Wamsley, University of California, Los Angeles; Jiani Yin, University of California, Los Angeles; Tarik Hadzic, University of California, Los Angeles; Luis De La Torre Ubieta, UCLA; Vivek Swarup, University of California, Los Angeles; Stephan J Sanders, University of California, San Francisco; Matthew W State, University of California, San Francisco; Donna M Werling, University of California, San Francisco; Joon-Yong An, University of California, San Francisco; Brooke Sheppard, University of California, San Francisco; A Jeremy Willsey, University of California, San Francisco; Kevin P White, The University of Chicago; Mohana Ray, The University of Chicago; Gina Giase, SUNY Upstate Medical University; Amira Kefi, University of Illinois at Chicago; Eugenio Mattei, University of Massachusetts Medical School; Michael Purcaro, University of Massachusetts Medical School; Zhiping Weng, University of Massachusetts Medical School; Jill Moore, University of Massachusetts Medical School; Henry Pratt, University of Massachusetts Medical School; Jack Huey, University of Massachusetts Medical School; Tyler Borrman, University of Massachusetts Medical School; Patrick F Sullivan, University of North Carolina - Chapel Hill; Paola Giusti-Rodriguez, University of North Carolina - Chapel Hill; Yunjung Kim, University of North Carolina - Chapel Hill; Patrick Sullivan, University of North Carolina - Chapel Hill; Jin Szatkiewicz, University of North Carolina - Chapel Hill; Suhn Kyong Rhie, University of Southern California; Christoper Armoskus, University of Southern California; Adrian Camarena, University of Southern California; Peggy J Farnham, University of Southern California; Valeria N Spitsyna, University of Southern California; Heather Witt, University of Southern California; Shannon Schreiner, University of Southern California; Oleg V Evgrafov, SUNY Downstate Medical Center; James A Knowles, SUNY Downstate Medical Center; Mark Gerstein, Yale University; Shuang Liu, Yale University; Daifeng Wang, Yale University; Fabio C. P. Navarro, Yale University; Jonathan Warrell, Yale University; Declan Clarke, Yale University; Prashant S. Emani, Yale University; Mengting Gu, Yale University; Xu Shi, Yale University; Min Xu, Yale University; Yucheng T. Yang, Yale University; Robert R. Kitchen, Yale University; Gamze Gürsoy, Yale University; Jing Zhang, Yale University; Becky C Carlyle, Yale University; Angus C Nairn, Yale University; Mingfeng Li, Yale University; Sirisha Pochareddy, Yale University; Nenad Sestan, Yale University; Mario Skarica, Yale University; Zhen Li, Yale University; Andre M.M. Sousa, Yale University; Gabriel Santpere, Yale University; Jinmyung Choi, Yale University; Ying Zhu, Yale University; Tianliuyun Gao, Yale University; Daniel J Miller, Yale University; Adriana Cherskov, Yale University; Mo Yang, Yale University; Anahita Amiri, Yale University; Gianfilippo Coppola, Yale University; Jessica Mariani, Yale University; Soraya Scuderi, Yale University; Anna Szekely, Yale

University; Flora M Vaccarino, Yale University; Feinan Wu, Yale University; Sherman Weissman, Yale University; Tanmoy Roychowdhury, Mayo Clinic Rochester; Alexej Abyzov, Mayo Clinic Rochester;

Figure Legends

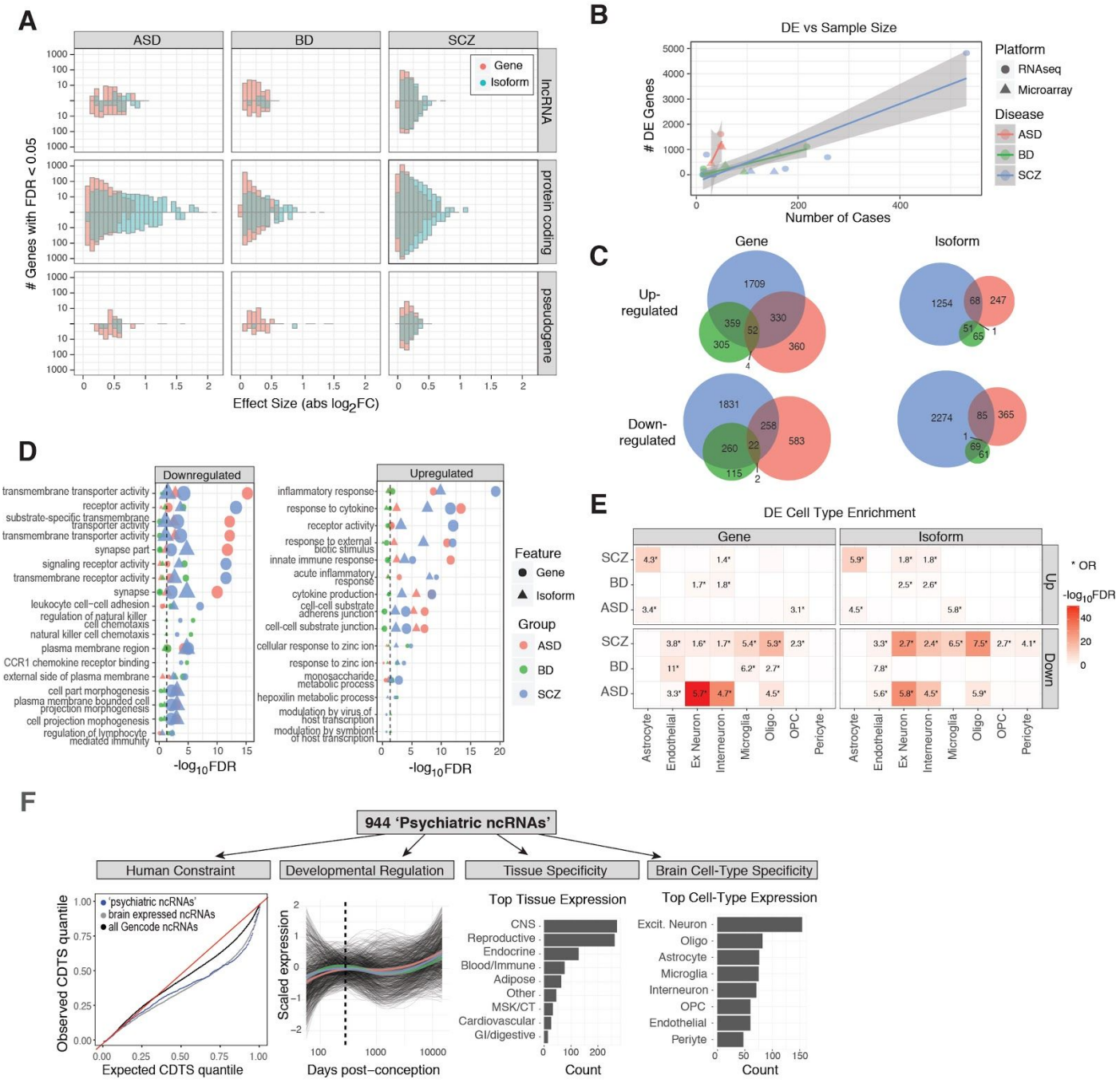


Figure 1. Gene and isoform expression dysregulation in psychiatric brain

A) Differential expression effect size ($|\log_2 FC|$) histograms are shown for protein-coding, lncRNA, and pseudogene biotypes up or downregulated (FDR<0.05) in disease. Isoform-level changes (DTE; blue) show larger effect sizes than at gene level (DGE; red), particularly for protein-coding biotypes in ASD and SCZ. B) A literature-based comparison shows that the number of DE genes detected is dependent on study sample size for each disorder. C) Venn diagrams depict overlap among up or downregulated genes and isoforms across disorders. D) Gene ontology enrichments are shown for differentially expressed genes or isoforms. The top 5 pathways are shown for each disorder. E) Heatmap depicting cell type specificity of enrichment signals. Differentially expressed features show substantial enrichment for known CNS cell type markers, defined at the gene level from single cell RNA-Seq. F) Annotation of 944 unique non-coding RNAs DE in at least one disorder. From left to right: Sequence-based characterization of ncRNAs for measures of human selective constraint; brain developmental expression trajectory are similar across each disorder (colored lines represent mean trajectory across disorders); tissue, and CNS cell type expression patterns.

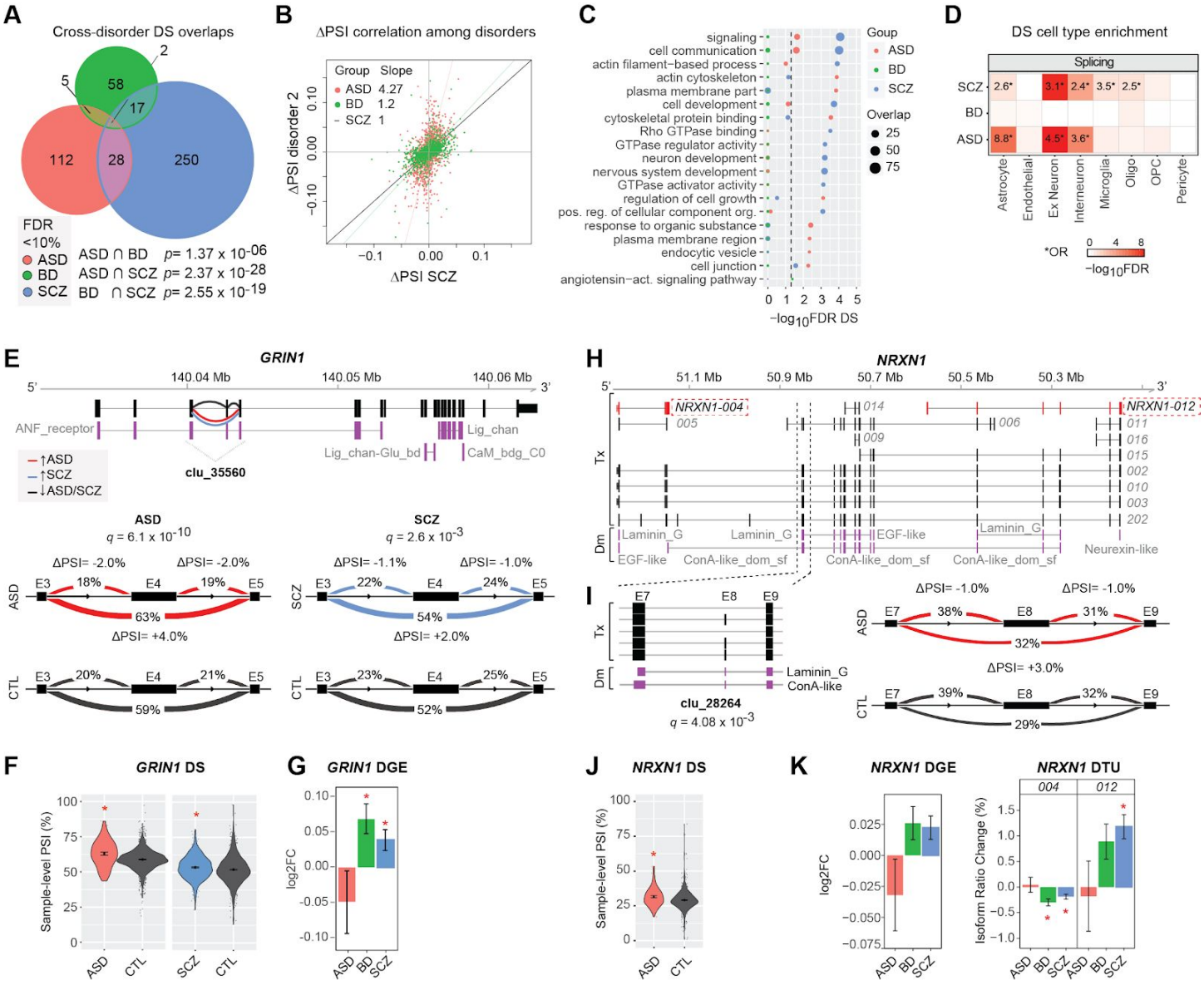


Figure 2. Aberrant local splicing and isoform usage in ASD, SCZ and BD

A) Venn diagram showing cross-disorder overlap for 472 genes with significant differentially spliced (DS) intron clusters (FDR< 10%) identified by LeafCutter. P values for hypergeometric tests of pairwise overlaps between each disorder are shown at the bottom. **B)** Scatter plots comparing percent spliced-in (PSI) changes for all 1,287 introns in 515 significant DS clusters in at least one disorder, for significant disease pairs SCZ *vs* ASD and SCZ *vs* BD (Spearman's ρ=0.52 and ρ=0.59, respectively). Principal component regression lines are shown in red, with regressions slopes for ASD and BD delta PSI compared to SCZ in the top-left corner. **C)** Top 10 gene ontology (GO) enrichments for DS genes in each disorder (see also **Fig S8C**). **D)** Significant enrichment for neuronal and astrocyte markers (ASD and SCZ), as well as oligodendrocyte and microglia (SCZ) cell type markers in DS genes. *Odds Ratio (OR) is given only for FDR< 5% and OR> 1. Oligo - oligodendrocytes; OPC - oligodendrocyte progenitor cells. **E)** A significant DS intron cluster in *GRIN1* (clu_35560; chr9:140,040,354-140,043,461) showing increased exon 4 (E4) skipping in both ASD and SCZ. Increased or decreased intron usage in ASD/SCZ cases compared to controls are highlighted in red and blue, respectively. Protein domains are annotationed as ANF_receptor - Extracellular receptor family ligand binding domain; Lig_chan - Ionotropic glutamate receptor; Lig_chan-Glu_bd - Ligated ion channel L-glutamate- and glycine-binding site; CaM_bdg_C0 - Calmodulin-binding domain C0 of NMDA receptor NR1 subunit. Visualization of splicing events in cluster clu_35560 with the change in PSI (ΔPSI) for ASD (left) and SCZ (right) group comparisons. FDR-corrected p-values (q) are indicated for each comparison. Covariate-adjusted average PSI levels in ASD or SCZ (red) *vs* CTL (blue) are indicated at each intron. **F)** Violin-plots with the distribution of covariate-adjusted PSI per sample for the intron skipping E4 are shown for each

disease group comparison. **G)** DGE for *GRIN1* in each disorder (*FDR< 5%). **H)** Whole-gene view of *NRXN1* highlighting (dashed lines) the intron cluster with significant DS in ASD (clu_28264; chr2:50,847,321-50,850,452), as well as transcripts NRXN1-004 and NRXN1-012 that show significant DTU in SCZ and/or BD. Protein domain mappings are shown in purple. DM - Protein domains; Tx - Transcripts. ConA-like_dom_sf - Concanavalin A-like lectin/glucanase domain. EGF-like - Epidermal growth factor-like domain; Laminin_G - Laminin G domain; Neurexin-like - Neurexin/syndecan/glycophorin C domain. **I)** Left: close-up of exons and protein domains mapped onto the DS cluster, and FDR-corrected *p*-value (*q*). Right: visualization of introns in cluster clu_28264 with their change in percent spliced in (ΔPSI). Covariate-adjusted average PSI levels in ASD (red) *vs* CTL (blue) are indicated for each intron. **J)** Violin-plots with the distribution of covariate-adjusted PSI per sample for the largest intron skipping exon 8 (E8). **K)** Bar plots for changes in gene expression and transcript usage for NRXN1-004 and NRXN1-012 (*FDR< 5%).

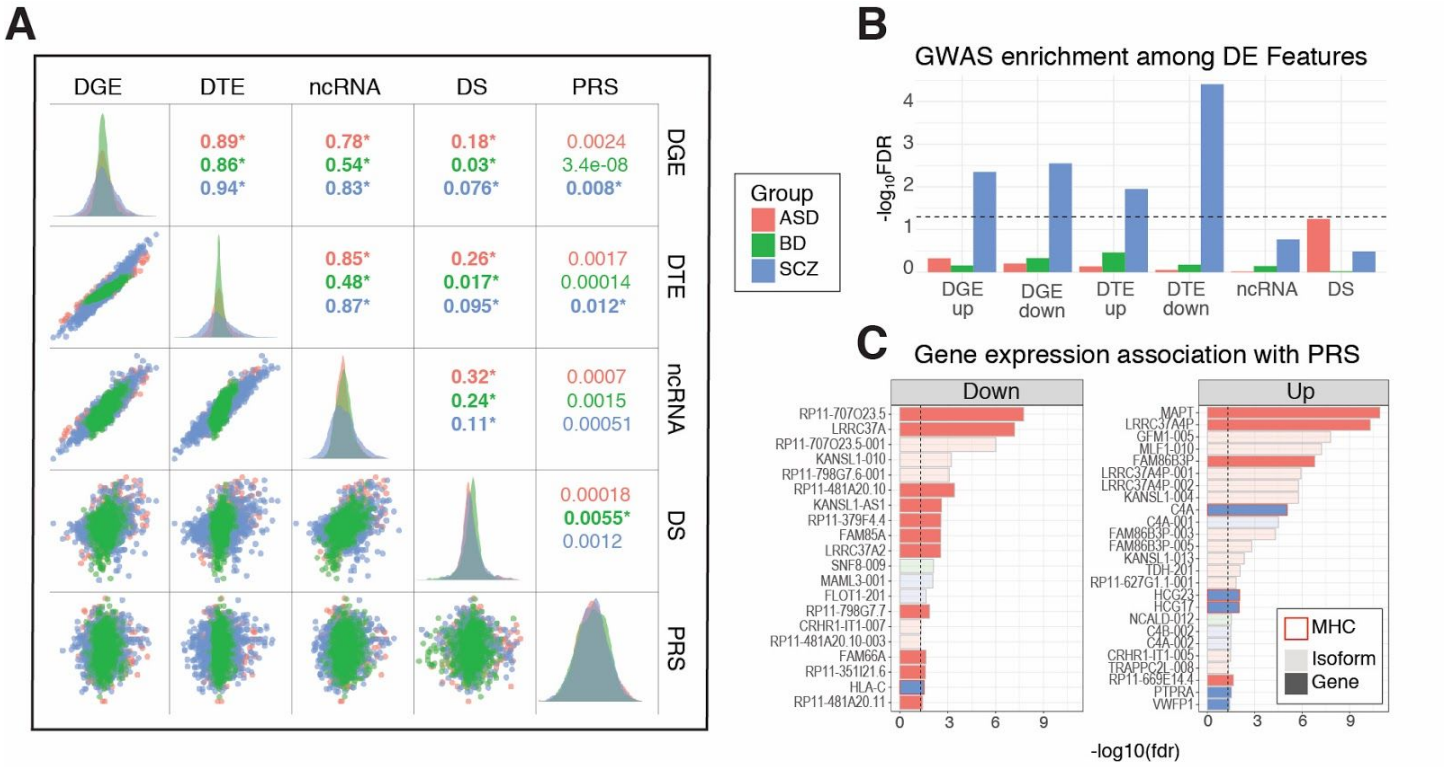


Figure 3. Overlaps and genetic enrichment among dysregulated transcriptomic features

A) Scatterplots demonstrate overlap among dysregulated transcriptomic features, summarized by their first principle component across case and control subjects (R^2 values; *P<0.05). Polygenic risk shows greatest association with differential transcript signal in SCZ, although the magnitude is small . B) Stratified LD-score regression identifies enrichment of GWAS SNP-heritability in SCZ among multiple differentially expressed transcriptomic features, with downregulated isoforms showing must substantial association. C) Polygenic risk scores created separately for each disorder are significantly (FDR<0.05) associated with expression for multiple individual genes and isoforms. Plots are split between upregulated and downregulated associations with increasing PRS. Several associations with SCZ PRS are located within the MHC region of the genome, which harbors the largest GWAS association signal but also highly complex LD structure.

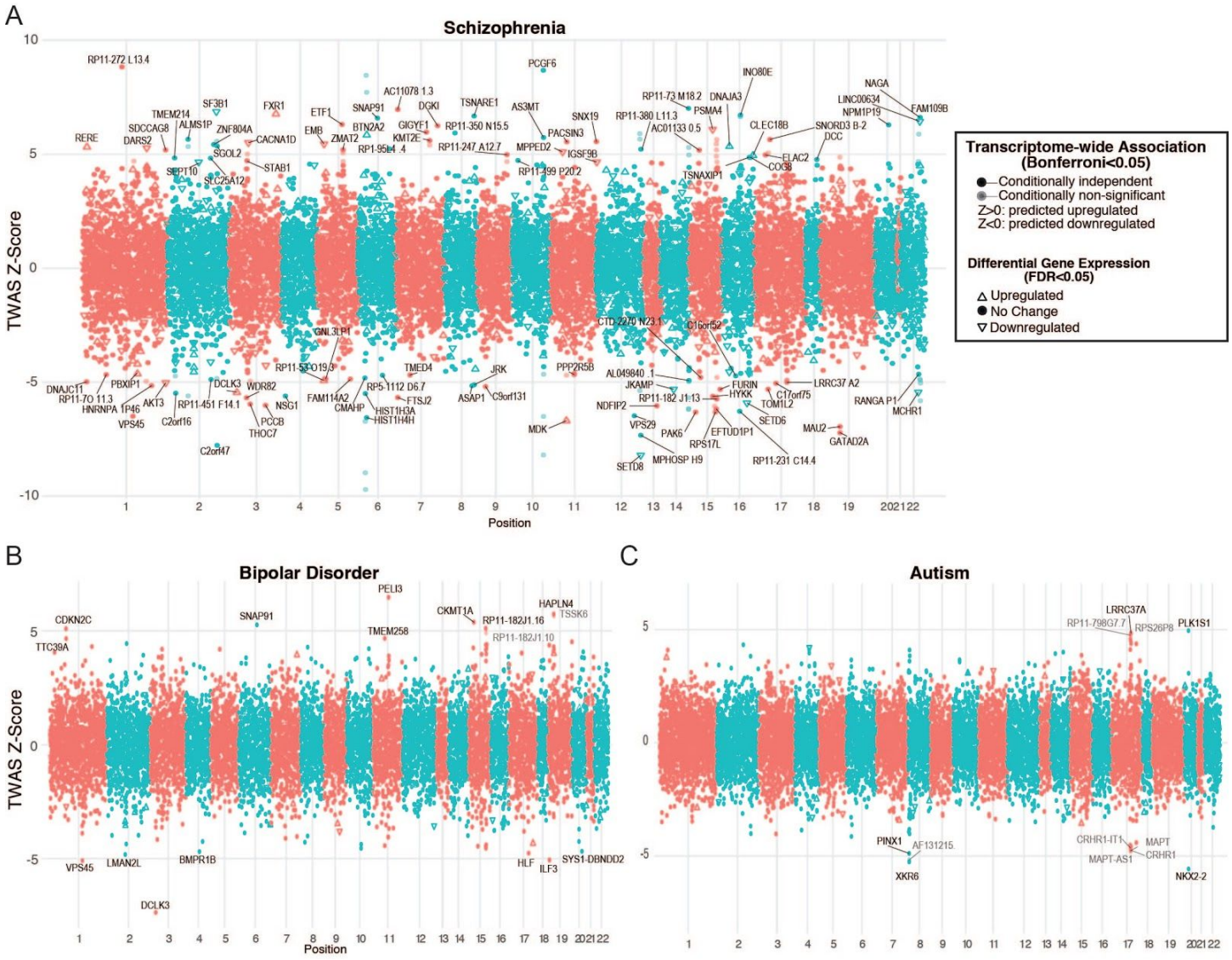


Figure 4. Transcriptome-wide association

Results from TWAS prioritize genes whose *cis*-regulated expression in brain is associated with disease. Plots show conditionally-independent TWAS prioritized genes, with lighter shade depicting marginal associations. The sign of TWAS Z-scores indicates predicted direction of effect. Genes significantly up or downregulated in disease brain are shown with arrows, indicating directionality. A) In SCZ, 193 genes (164 outside of MHC) are prioritized by TWAS at Bonferroni-corrected P<0.05, including 107 genes with conditionally independent signals. Of these, 23 are also differentially expressed in SCZ brain with 11 in the same direction as predicted. B) Seventeen genes are prioritized in BD, of which 15 are conditionally independent. Three TWAS associations overlap between SCZ and BD: *SNAP91*, *DCLK3*, *VPS45*. C) In ASD, TWAS prioritizes 12 genes, of which 5 are conditionally independent.

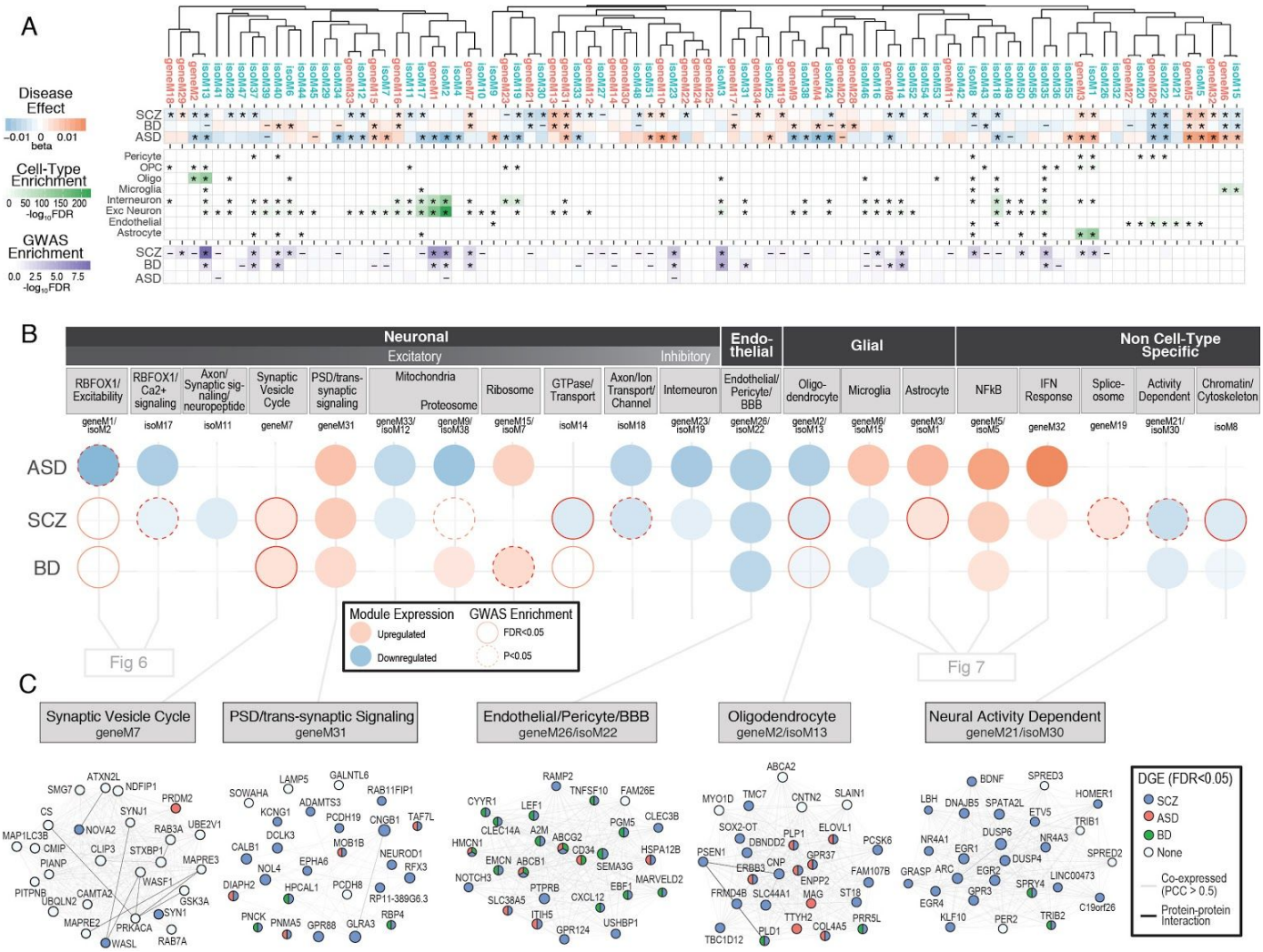


Figure 5. Gene and isoform coexpression networks capture shared and disease-specific cellular processes and interactions

A) Gene and isoform co-expression networks demonstrate pervasive dysregulation across psychiatric disorders. Hierarchical clustering shows that separate gene- and isoform-based networks are highly overlapping, with greater specificity conferred at the isoform level. Disease associations are shown for each module (linear regression β value, * FDR<0.05, – P<0.05). Module cell type enrichments (*FDR < 0.05) are shown for major CNS cell types defined from PsychENCODE UMI single cell clusters. Enrichments are shown for GWAS results from SCZ (*58*), BD (*90*), and ASD (*38*), using stratified LD score regression (* FDR<0.05, – P<0.05). B) Co-expression modules capture specific cellular identities and biological pathways. Colored circles represent module differential expression effect size in disease, with red outline representing GWAS enrichment in that disorder. Modules are organized and labeled based on CNS cell type and top-gene ontology enrichments. However, we recognize that these annotations are imprecise with respect to complex neurobiological processes, such as those dysregulated in disease. C) Examples of specific modules dysregulated across disorders, with top 25 hub genes shown. Edges represent co-expression (Pearson correlation > 0.5) and known protein-protein interactions. Nodes are colored to represent disorders in which that gene is differentially expressed (*FDR<0.05).

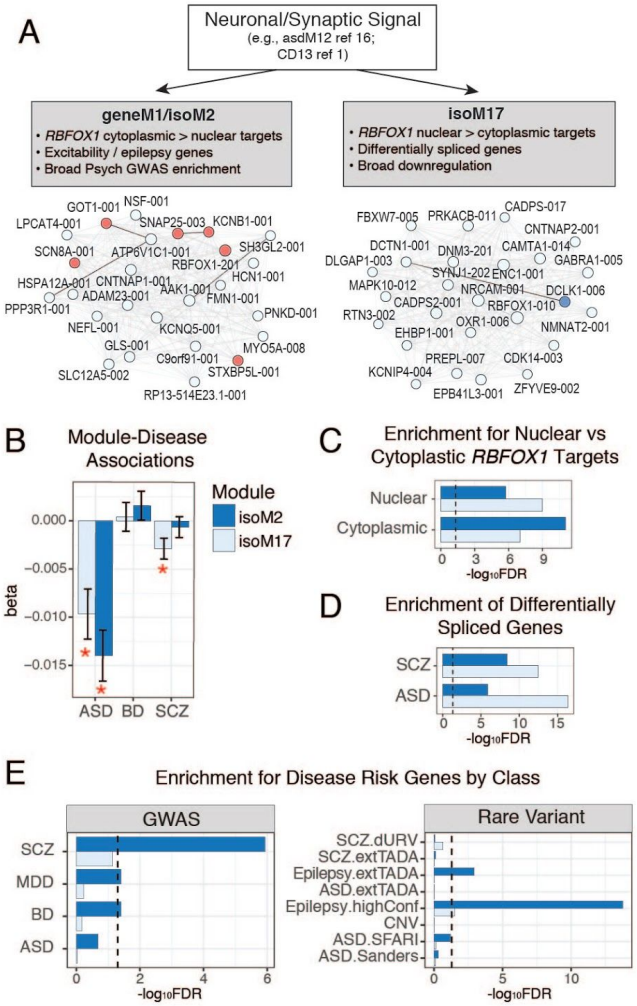


Figure 6. Two *RBFOX1* isoform modules capture distinct biological and disease associations.

A) Previous studies have identified *RBFOX1* as a critical hub of neuronal and synaptic modules downregulated across multiple psychiatric disorders (*1*, *16*, *19*, *32*). Here, we identify two pairs of modules with distinct *RBFOX1* isoforms as hub genes. Plots show the top 25 hub genes of modules isoM2 and isoM17, following the same coloring scheme as **Fig 5C**. B) Distinct module-eigengene trait associations are observed for isoM2 (downregulated in ASD only) compared with isoM17, which is downregulated in ASD and SCZ. C) Modules show distinct enrichments for nuclear and cytoplasmic *RBFOX1* targets, defined experimentally in mouse (*32*)**.** D) Genes harboring differential splicing events observed in ASD and SCZ show greater overlap with isoM17, consistent with its association with nuclear *RBFOX1* targets. E) Modules show distinct patterns of genetic association. isoM2 exhibits broad enrichment for GWAS signal in SCZ, BD, and MDD, as well as for epilepsy risk genes, whereas isoM17 shows no apparent genetic enrichment. GWAS enrichments show FDR-corrected P-values calculated using stratified-LDSC, and rare-variant associations were calculated using logistic regression, controlling for gene length and GC content (*21*).

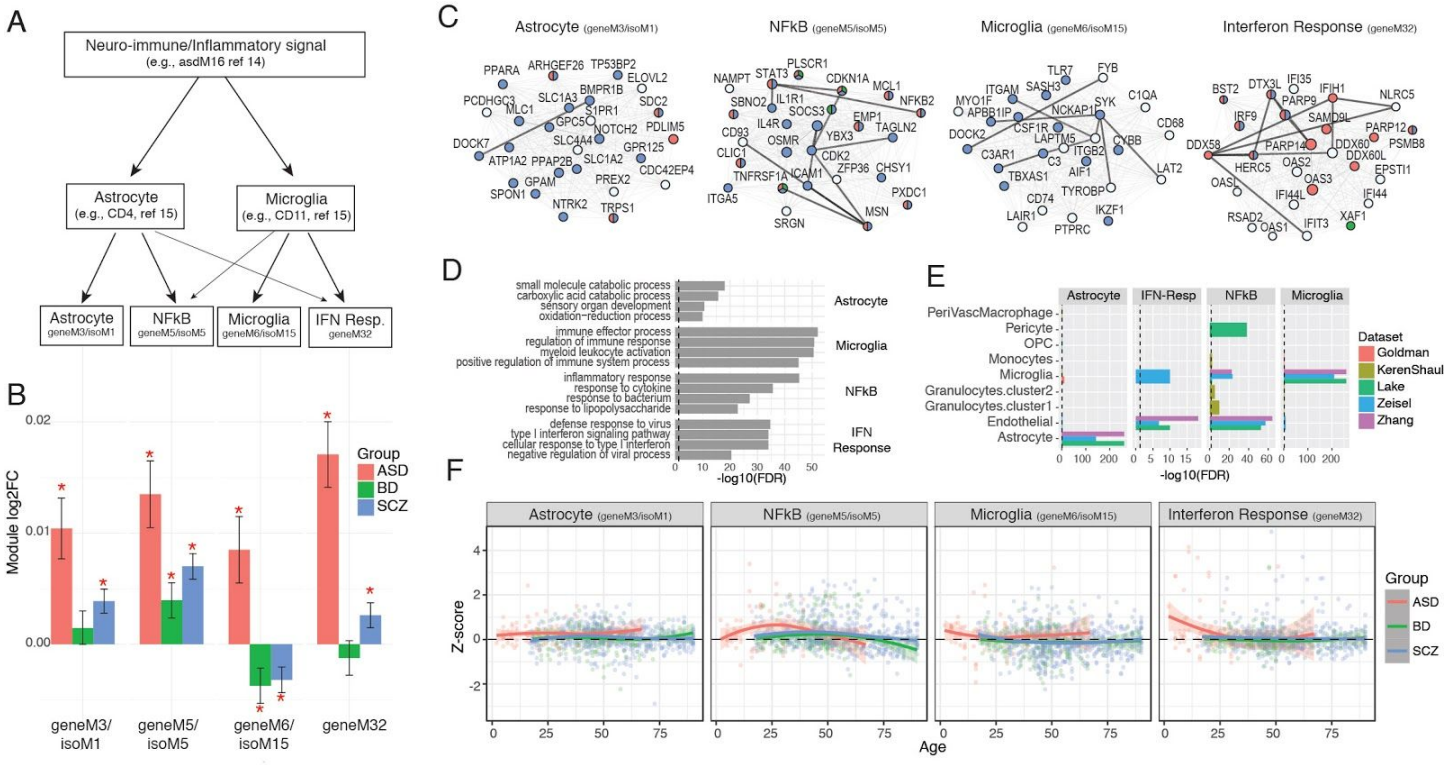


Figure 7. Distinct neural-immune trajectories in disease

A) Coexpression networks provide substantial refinement of the neuro-immune/inflammatory processes upregulated in ASD, SCZ, and BD. Previous work has identified specific contributions to this signal from astrocyte and microglial populations (*1*, *19*). Here, we further identify additional, distinct interferon (IFN)-response and NFkB signaling modules. B) Eigengene-disease associations are shown for each of 4 identified neural-immune module pairs. The astrocyte and IFN-response modules are upregulated in ASD and SCZ. NFkB signaling is elevated across all three disorders. The microglial module is upregulated in ASD and downregulated in SCZ and BD. C) Top hub genes for each module are shown, along with edges supported by co-expression (light grey; Pearson correlation > 0.5) and known protein-protein interactions (dark lines). Nodes follow same coloring scheme as in Fig 5C. Hubs in the astrocyte module (geneM3/isoM1) include several canonical, specific astrocyte markers, including *SOX9*, *GJA1*, *SPON1*, and *NOTCH2*. Microglial module hub genes include canonical, specific microglial markers, including *AIF1*, *CSF1R*, *TYROBP*, *TMEM119*. The NFkB module includes many known downstream transcription factor targets (*JAK3*, *STAT3*, *JUNB*, *FOS*) and upstream activators (*IL1R1*, 9 TNF receptor superfamily members) of this pathway. D) The top 4 GO enrichments are shown for each module. E) Module enrichment for known cell type-specific marker genes, collated from sequencing studies of neural-immune cell types (*91*–*95*). F) Module eigengene expression across age demonstrates distinct and dynamic neural-immune trajectories for each disorder.

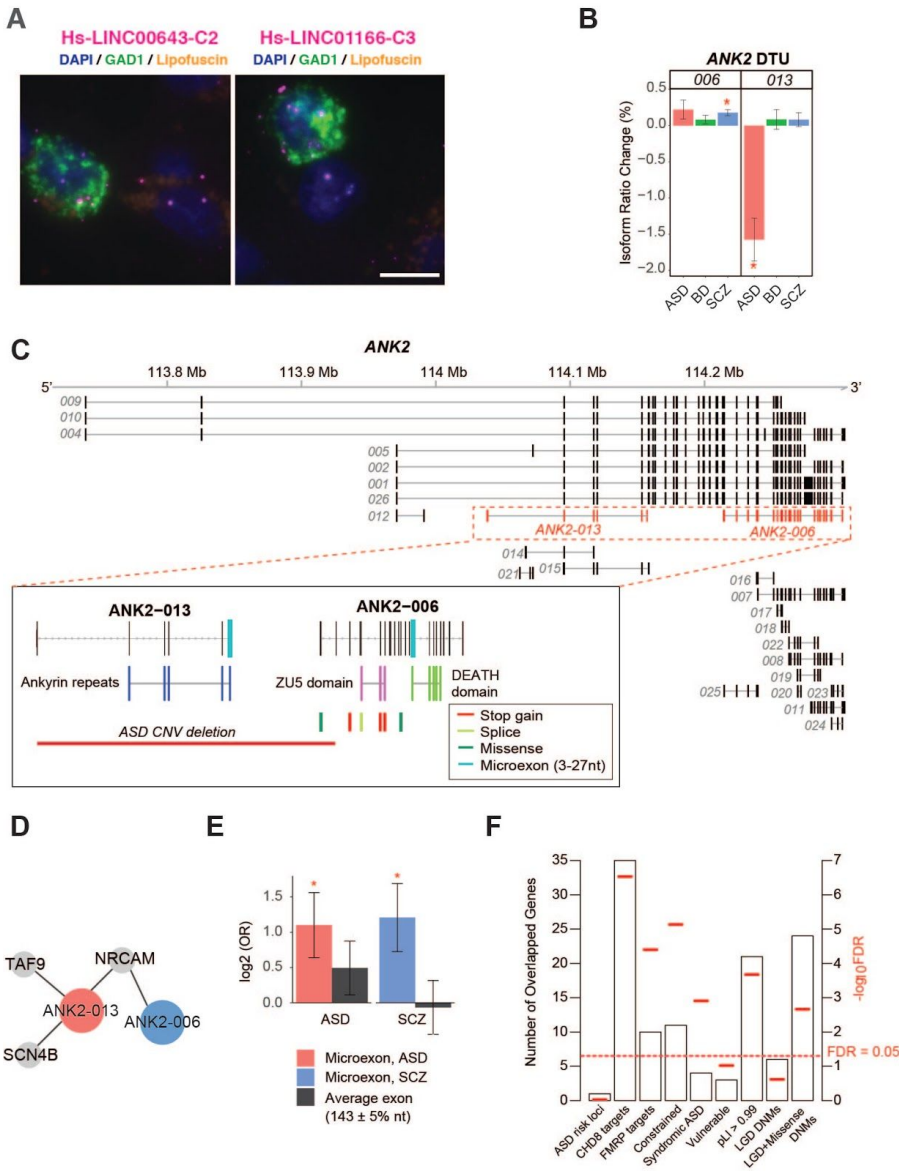


Figure 8. LncRNA annotation, *ANK2* isoform switching & microexon enrichment

A) FISH images demonstrate interneuron expression for two poorly annotated lincRNAs – *LINC00643* and *LINC01166* – in area 9 of adult human prefrontal cortex. Sections were labeled with *GAD1* probe (green) to indicate GABAergic neurons and lncRNA (magenta) probes for *LINC00643* (left) or for *LINC01166* (right). All sections were counterstained with DAPI (blue) to reveal cell nuclei. Lipofuscin autofluorescence is visible in both the green and red channels and appears yellow/orange. Scale bar, 10 µm. FISH was repeated at least twice on independent samples (**Table S9** (*21*)) with similar results (see also **Fig S16**). **B**) *ANK2* isoforms *ANK2-006* and *ANK2-013* show significant DTU in SCZ and ASD, respectively (*FDR<0.05). **C**) Exon structure of *ANK2* highlighting (dashed lines) the *ANK2-006* and *ANK2-013* isoforms. Inset, these isoforms have different protein domains and carry different microexons. *ANK2-006* is hit by multiple ASD DNMs while *ANK2-013* could be entirely eliminated by a *de novo* CNV deletion in ASD. **D**) Disease-specific co-expressed PPI network. Both ANK2-006 and ANK2-013 interact with NRCAM. The ASD-associated isoform ANK2-013 has two additional interacting partners, SCN4B and TAF9. **E**) As a class, switch isoforms are significantly enriched in microexon(s). In contrast, exons of average length are not enriched among switch isoforms. Y-axis displays odds ratio on $\log_2$ scale. P-values are calculated using logistic regression and corrected for multiple comparisons. **F**) Enrichment of 64 genes with switch isoforms in: ASD risk loci (*80*); CHD8 targets (*96*); FMRP targets (*33*); Mutationally constraint genes (*97*); Syndromic and highly ranked (1 and 2) genes from SFARI Gene database; Vulnerable ASD genes (*98*); Genes with

probability of loss-of-function intolerance (pLI) > 0.99 as reported by the Exome Aggregation Consortium (*99*); Genes with likely-gene-disruption (LGD) or LGD plus missense *de novo* mutations (DNMs) found in patients with neurodevelopmental disorders (*21*).

Supplementary Material

Materials and Methods

Data Generation

The data generated for this manuscript represent Freeze 1 and 2 of the PsychENCODE consortium dataset. Post-mortem human brain samples were collected as part of eight studies, detailed below and in **Fig S1**. RNA-Seq and genotype array data was generated by each site and then processed together through a unified pipeline by a central data analysis core. For this capstone analysis, we restricted analysis to frontal and temporal cortex brain samples from postnatal timepoints. We provide a description of each individual study below, derived from the PsychENCODE website. All data are available at doi.org/10.7303/syn12080241.

Study 1 - BrainGVEX

For the BrainGVEX study, RNA-Seq data was generated from 427 post-mortem prefrontal cortex samples from subjects with schizophrenia (n=95), bipolar disorder (n=73), and non-psychiatric controls (n=259). RNA samples were collected from the Stanley Medical Research Institute (SMRI) as part of the “Array Collection”, “Consortium Collection”, “New Collection” and ”Extra Collection”. Array collection and consortium collection samples were from the superior frontal gyrus (Brodmann’s area (BA) 9) whereas those from extra and new collections were from the mid frontal gyrus (BA46). Another 184 controls were obtained as fresh-frozen brain tissue from the Banner Sun Health Research Institute (BSHRI). All BSHRI samples were from the frontal cortex. RNA was extracted from BSHRI samples by first homogenizing 20-50 mg of tissue in QIAzol (Qiagen) using the Lysin Matrix D and FastPrep-24 system (MPBiomedicals). Total RNA were then isolated using the miRNeasy Kit (Qiagen) according to manufacturer’s instructions. RNA integrity was assessed with Agilent Technologies RNA 600 nano kit. Samples with RNA Integrity Number (RIN) lower than 5.5 were excluded from the study. RNA sequencing libraries were prepared using TruSeq Stranded Total RNA sample prep kit with RiboZero Gold HMR (Illumina). Libraries were multiplexed (3 per lane) for paired-end 100 bp sequencing on Illumina HiSeq2000 with read depth >70 million reads on average. Genotyping was performed using two different platforms. 144 samples (SMRI Consortium and Array Collections) were genotyped using the Affymetrix GeneChip Mapping 5.0K Array. Genotypes were called with the BRLMM-p algorithm (Affymetrix) on all arrays simultaneously (*100*). The remaining samples (SMRI New and Extra Collection, and BSHRI samples) were genotyped on the Human PsychChip platform, which is a custom version of the Illumina Infinium CoreExome-24 v1.1 BeadChip (#WG-331-1111). However, PsychChip data were not yet available for this study.

Study 2 - BrainSpan

For the BrainSpan study, RNA-Seq data was generated from 606 brain samples from 41 unique individuals. RNA was extracted using RNeasy Plus Mini Kit (Qiagen) for mRNA. Either approximately 30 mg of pulverized tissue (12 PCW – 40 Y specimens) or entire amount of dissected brain piece (8 – 9 PCW, smaller than 30 mg) was processed. Tissue was pulverized with liquid nitrogen in a chilled mortar and pestle and transferred to a chilled safe-lock microcentrifuge tube (Eppendorf). Per tissue mass, equal mass of chilled stainless steel beads (Next Advance, cat# SSB14B) along with two volumes of lysis buffer were added. Tissue was homogenized for 1 min in Bullet Blender (Next Advance # SSB14B) at speed 6 and incubated at 37°C for 5 min. Lysis buffer up to 0.6 ml was again added, tissue homogenized for 1 min and incubated at 37°C for 1 min. Extraction was further carried out according to manufacturer’s protocol. Genomic DNA was removed by a proprietary column provided in RNeasy Plus Mini Kit (Qiagen) or by DNase treatment using TURBO DNA-free

Kit (Ambion/ Life technologies). 260:A280 ratio and RNA Integrity Number (RIN) were determined for each sample with NanoDrop (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer system, respectively. The mRNA-sequencing (mRNA-Seq) sample preparation Kit (Illumina) was used to prepare cDNA libraries per manufacturer instructions with some modifications. Briefly, polyA RNA was purified from 1 to 5 µg of total RNA using Oligo (dT) beads. Quaint-IT RiboGreen RNA Assay Kit (Invitrogen) was used to quantitate purified mRNA with the NanoDrop 3300. Following mRNA quantitation, 2.5 µl spike-in master mixes, containing five different types of RNA molecules at varying amounts (2.5 × 10-7 to 2.5 × 10-14 mol), were added per 100 ng of mRNA. Spike-in RNAs were synthesized by the External RNA Control Consortium (ERCC) by *in vitro* transcription of *de novo* DNA sequences or DNA derived from B. subtilis or the deep-sea vent microbe M. jannaschii and were a generous gift of Dr. Mark Salit at The National Institute of Standards and Technology (NIST). Each sample was tagged by adding two spike-in RNAs unique to the region from which the sample was taken. Further, three common spike-in RNAs with gradient concentrations were added to each sample, to enable the assessment of sequencing quality. Spike-in sequences are available at http://archive.gersteinlab.org/proj/brainseq/spike_in/spike_in.fa. The mixture of mRNA and spike-in RNAs was subjected to fragmentation, reverse transcription, end repair, 3' end adenylation, and adapter ligation to generate libraries of short cDNA molecules, followed by PCR amplification. The PCR enriched product was assessed for its size distribution and concentration using Bioanalyzer DNA 1000 Kit. Genotype data were not used in this study.

Study 3 - CommonMind

Full details of the CommonMind study have been published (*2*), although the data here were processed separately according to the uniform RNA-Seq pipeline described below. Samples were acquired through brain banks at three institutions:The Mount Sinai NIH Brain Bank and Tissue Repository, University of Pennsylvania Brain Bank of Psychiatric illnesses and Alzheimer's Disease Core Center, and the University of Pittsburgh NIH NeuroBioBank Brain and Tissue Repository. Details about brain banks, inclusion/exclusion criteria, and sample collection and processing have been previously described (https://www.synapse.org/#!Synapse:syn2759792/wiki/71104). RNA-Seq data from 613 total human post-mortem dorsolateral prefrontal cortex (DLPFC) brain samples were obtained from 603 subjects with schizophrenia (n=263), bipolar disorder (n=47), affective disorder (8), and neurotypical controls (n=285), where 10 neurotypical controls were sequenced as technical replicates. Subjects with affective disorder were not used in this study. Total RNA was extracted from 50 mg of homogenized dorsolateral prefrontal cortex tissue using RNeasy kit. Samples with RIN < 5.5 (n=51) were excluded. The remaining samples had a mean RIN of 7.7. RNA-Seq library preparation was performed using ribosomal RNA depletion, with the RiboZero Magnetic Gold Kit. Samples were barcoded, multiplexed (n=10/lane), and sequenced across two lanes as 100 bp paired end sequencing on the Illumina HiSeq 2500 with an average of 85 million reads. Data are provided for those samples that passed all of the following QC filters: samples were required to have had a minimum of 50 million total reads and less than 5% rRNA alignment. For genotyping, DNA was isolated from approximately 10 mg dry homogenized tissue coming from the same dissected samples as the RNA isolation using the Qiagen DNeasy Blood and Tissue Kit according to manufacturer's protocol. Genotyping was performed using the Illumina Infinium HumanOmniExpressExome platform (Catalog #: WG-351-2301). All data were checked for discordance between nominal and genetically-inferred sex using Plink software to calculate the mean homozygosity rate across X-chromosome markers and to evaluate the presence or absence of Y-chromosome markers. In addition, pairwise comparison of samples across all genotypes was done to identify potentially duplicate samples (genotypes > 99% concordant) or related individuals using Plink.

Study 4 - Yale-ASD

For the Yale-ASD study, RNA-Seq data was generated from 45 brain samples from 37 unique individuals, including 9 with ASD and 28 controls. Total RNA was extracted using mirVana kit (Ambion) with some modifications to the manufacturer's protocol. Approximately 60 mg of tissue was pulverized with liquid nitrogen in a pre-chilled mortar and pestle and transferred to a chilled safe-lock microcentrifuge tube (Eppendorf). Per tissue mass, equal mass of chilled stainless steel beads (Next Advance, catalog # SSB14B) along with one volume of lysis/binding buffer were added. Tissue was homogenized for 1 min in Bullet Blender (Next Advance) and incubated at 37°C for 1 min. Another nine volumes of the lysis/binding buffer were added, homogenized for 1 min, and incubated at 37°C for 2 min. One-tenth volume of miRNA Homogenate Additive was added and extraction was carried out according to the manufacturer's protocol. RNA was treated with DNase using TURBO DNA-free Kit (Ambion/ Life Technologies) and RNA integrity was measured using Agilent 2200 TapeStation System. Barcoded libraries for RNA-Seq were prepared with 5 ng of RNA using TruSeq Stranded Total RNA with Ribo-Zero Gold kit (Illumina) per manufacturer's protocol. Paired-end sequencing (100bp x 2) was performed on HiSeq 2000 sequencers (Illumina) at Yale Center for Genome Analysis. Genotype data was not yet available as part of Freeze 1 or 2 of the PsychENCODE dataset.

Study 5 - UCLA-ASD

For the UCLA-ASD study, RNA-Seq data was generated from 253 brain samples from 97 unique individuals, across prefrontal cortex (BA9/46), temporal cortex (BA41/42/22), and cerebellum. Full details of the UCLA-ASD study have been published (*19*). Brain samples were obtained from the Harvard Brain Bank as part of the Autism Tissue Project (ATP). Frozen brain regions were dissected on dry ice in a dehydrated dissection chamber to reduce degradation effects from sample thawing or humidity. Approximately 50-100 mg of tissue across the cortical region of interest was isolated from each sample using the miRNeasy kit with no modifications (Qiagen). For each RNA sample, RNA quality was quantified using the RNA Integrity Number (RIN) on an Agilent Bioanalyzer. Strand-specific, rRNA-depleted RNA-Seq libraries were prepared using TruSeq Stranded Total RNA sample prep kit with RiboZero Gold (Illumina) kits. Libraries were randomly pooled to multiplex 24 samples per lane using Illumina TruSeq barcodes. Each lane was sequenced five times on an Illumina HiSeq 2500 instrument using high output mode with standard chemistry and protocols for 50 bp paired-end reads to achieve a target depth of 70 million reads. Genotyping data was generated at the UCLA Neurogenomics Core (UNGC) on the Illumina Omni 2.5 8v1 platform (Human Exome). Illumina Genome Studio files were clustered using Illumina's standard HapMap cluster file. SNP genotypes were exported from the Illumina GenomeStudio Software as forward strand in PLINK format. SNP marker names were updated with a conversion file from Illumina which converts local marker name to rsID (plink --update-map --update-name). All quality filtering was performed using PLINK v1.07. SNPs missing more than 99.99% data were excluded (--geno 0.9999). Individuals missing > 5% data, SNPs missing > 5% data, and SNPs with HW p<0.0000001 were also excluded. The order of filtering was performed according to PLINK default procedures (plink --mind 0.05 --geno 0.05 --hwe 0.0000001).

Study 6 - CMC_HBCC

Brain specimens for the CMC_HBCC study were obtained from the the NIMH Human Brain Collection Core (HBCC; https://www.nimh.nih.gov/labs-at-nimh/research-areas/research-support-services/hbcc/human-brain-collection-core-hbcc.shtml) under protocols approved by the CNS IRB (NCT00001260), with the permission of the next-of-kin through the Offices of the Chief Medical Examiners in the District of Columbia, Northern Virginia, and Central Virginia. All specimens were characterized neuropathologically, clinically and toxicologically. A clinical diagnosis was obtained through family interviews and review of medical records by two psychiatrists

based on DSM-IV criteria. Non-psychiatric controls were defined as having no history of a psychiatric condition or substance use disorder. Brain samples were dissected at the NIMH Human Brain Collection Core and shipped to Icahn School of Medicine at Mount Sinai (ISMMS) for sample preparation and RNA-sequencing. Samples for the study were dissected from either the left or right hemisphere of fresh frozen coronal slabs cut at autopsy from the dorsolateral prefrontal cortex. Total RNA from 468 HBCC samples was isolated from approximately 100 mg homogenized tissue from each sample by TRIzol/chloroform extraction and purification with the Qiagen RNeasy kit (Cat#74106) according to manufacturer's protocol. Samples were processed in randomized batches of 12. The order of extraction was assigned randomly with respect to diagnosis and all other sample characteristics. The mean total RNA yield was 24.2 ug. The RNA Integrity Number (RIN) was determined by fractionating RNA samples on the 4200 Agilent TapeStation System. Sixty nine samples with RIN <5.5 were excluded from the study. An additional 12 samples were removed post sequencing due to evidence of sample swap or contamination, resulting in a final dataset of 387 samples (70 BD, 97 SCZ, 220 neurotypical controls) with a mean RIN of 7.5 and a mean ratio of 260/280 of 2.0. RNA sequencing raw and quantified expression data is provided for these 387 samples from 387 unique individuals. Data was generated, QCed, processed and quantified as follows: All samples submitted to the New York Genome Center for RNA-Seq were prepared for sequencing in randomized batches of 94. The sequencing libraries were prepared using the KAPA Stranded RNA-Seq Kit with RiboErase (KAPA Biosystems). rRNA was depleted from 1ug of RNA using the KAPA RiboErase protocol that is integrated into the KAPA Stranded RNA-Seq Kit. The insert size and DNA concentration of the sequencing library was determined on Fragment Analyzer Automated CE System (Advanced Analytical) and Quant-iT PicoGreen (Thermo Fisher Scientific) respectively. A pool of 10 barcoded libraries were layered on a random selection of two of the eight lanes of the Illumina flow cell at appropriate concentration and bridge amplified to ~ 250 million raw clusters. One-hundred base pair paired end reads were obtained on a HiSeq 2500. Genotyping was performed using Illumina_1M, Illumina_h650, and Illumina_Omni5 platforms.

Studies 7+8 - BipSeq & LIBD_szControl

Post-mortem tissue homogenates of dorsolateral prefrontal cortex (DLPFC) approximating BA46/9 in postnatal samples and the corresponding region of PFC in fetal samples were obtained from all subjects. Total RNA was extracted from ~100 mg of tissue using the RNeasy kit (Qiagen) according to the manufacturer's protocol. The poly-A containing RNA molecules were purified from 1 µg DNAse treated total RNA and sequencing libraries were constructed using the Illumina TruSeq© RNA Sample Preparation v2 kit. Sequencing indices/barcodes were inserted into Illumina adapters allowing samples to be multiplexed across lanes in each flow cell. These products were then purified and enriched with PCR to create the final cDNA library for high throughput sequencing using an Illumina HiSeq 2000 with paired end 2x100bp reads. Further details are available in (*101*). SNP genotyping with HumanHap650Y_V3, Human 1M-Duo_V3, and Omni5 BeadChips (Illumina, San Diego, CA) was carried out according to the manufacturer's instructions with DNA extracted from cerebellar tissue. Genotype data were processed and normalized with the crlmm R/Bioconductor package separately by platform.

RNA-sequencing Data Processing Pipeline

All sample FASTQ files were run through a unified RNA-Seq processing pipeline (**Fig S1**) run at the University of Chicago on an OpenStack cloud system and modeled after the long-rna-seq-pipeline used by the ENCODE Consortium. Fastqs were trimmed for adapter sequence and low base call quality (Phred score < 30 at ends) using cutadapt (v1.12). Trimmed reads were then aligned to the GRCH37.p13 (hg19) reference genome via STAR (2.4.2a) using comprehensive gene annotations from Gencode (v19). BAM files were produced in both genomic and transcriptome coordinates and sorted using samtools (v1.3). Gene and isoform-level quantifications were calculated using RSEM (v1.2.29). Quality control metrics were calculated

using RNA-SeQC (v1.1.8), featureCounts (v1.5.1), PicardTools (v1.128), and Samtools (v1.3.1). Pipeline source code can be found on Synapse at doi:10.7303/syn12026837. RNA-Seq data was processed in two batches: Freeze 1 consisted of re-processed RNA-Seq data from the following studies: BrainGVEX, BrainSpan, CMC, UCLA-ASD, Yale-ASD, iPSC; Freeze 2 consisted of BipSeq, LIBD_szControl, CMC_HBCC, EpiGABA. Data from EpiGABA and iPSC studies was not used in this Capstone project.

Genotyping and QTL Pipeline

Genotype calls were generated at each data production site separately, as described above, and centralized for imputation. Genotype imputation and QTL analyses were performed as described in our companion manuscript (*18*) and on the PsychENCODE website using a uniform genotype QC and imputation pipeline for all studies. To generate high-quality observed genotypes (removing low quality and rare variants), initial QC was performed using Plink to remove SNPs with zero alternate alleles, MAF <1%, genotyping call rate < 0.95, Hardy-Weinberg p-value < $1x10^{-6}$, individuals with genotyping call rate < 0.95, and to correct strand flips. Parallel haplotype pre-phasing and imputation were done using Beagle2, Minimac3 with the HRC reference panel for imputation. Calculation of gene-level expression QTLs (eQTL) and isoform-level expression QTLs (isoQTL) was done using QTLtools, as described in our companion manuscript (*18*). Imputation of *C4A* structural variation for each genotyped sample of European ancestry was performed using Beagle5 with a custom HapMap3 CEU reference panel as described (*5*). Inferred copy number of C4 structural elements (C4A, C4B, C4L, and C4S) based on the imputed C4 alleles was then associated with normalized C4A expression using a linear model.

RNA-Seq Quality Control and Normalization

Expected counts were compiled from gene and isoform-level RSEM quantifications and imported into R for downstream analyses. Genes were filtered to include those with TPM > 0.1 in at least 25% of samples. We removed all transcripts derived from mitochondrial DNA and Y-chromosome pseudoautosomal regions (“ENSR”) as well as transcripts with immunoglobulin (IG or TR) biotypes or those shorter than 250 bp. Downstream analyses were performed on the resulting 25,774 transcribed genes based on GENCODE V19 annotations. We restricted our analysis to frontal and temporal cortex brain samples obtained from subjects at postnatal time points (**Fig S2**). We removed samples with an ambiguous diagnosis or a diagnostic label other than ASD, SCZ, BD, or CTL (n=11). We removed samples with unspecified or ambiguous age (n=2), or sex (n=2) as well as samples with less than 10 million total reads. Each individual study was then assessed for outlier samples (**Fig S2C**), defined as those with standardized sample network connectivity Z scores < -2, as published, which were removed (*102*). We further removed 8 samples whose documented sex was discordant from that predicted by gene expression, based on hierarchical clustering of samples using expression of *XIST* and the first principal component of genes on the Y chromosome.

Covariate Selection

We compiled a set of 187 RNA-Seq quality control metrics as the aggregate sample-level outputs from RNA-SeQC, cutadapt, featureCounts, PicardTools (CollectAlignmentSummaryMetrics, CollectInsertSizeMetrics, CollectRnaSeqMetrics, MarkDuplicates), and STAR (**Fig S2**). As many of these metrics were highly overlapping, we summarized these measures by the top 29 principal components which collectively explained 99% of the total variance. To determine which covariates to include in the final differential expression model, we performed multivariate adaptive regression as implemented in the *earth* package in R. This builds a model in two phases using a forward pass to capture maximal amount of variance explained by an underlying set of covariates, followed by a backward (pruning) pass to remove potential redundant terms. The superset of potential covariates available for all samples included: diagnosis, age, study/batch, sex, PMI, RIN, libraryPrep, sequencing platform, strand specificity, brain bank, brain region, ethnicity, along with all 29

seqPCs. For continuous variables, we also included squared terms. These covariates with input into the *earth* model along with gene expression data (limma voom normalized, centered, and scaled). The model was run using linear predictors and otherwise default parameters. As the model fits a maximum of 1000 features (genes) simultaneously, we performed 1000 permutations randomly subsetting 1000 genes at a time. From this, we chose as a set of known covariates those present in at least half of the resulting pruned models, which consisted of: diagnosis, age, age^2, study/batch, sex, PMI, RIN, RIN^2, brain bank, brain region, seqPCs (1-3, 5-8, 10-14, 16, 18-25, 27-29) and seqPC3^2.

The above set contained known covariates (or those derived from known sequencing quality metrics) that contributed uniquely to variance in gene expression. However, these do not capture potential underlying hidden factors or confounders that may also influence gene expression. To ensure that DGE signal is not being driven by such hidden confounding factors, we performed surrogate variable analysis (SVA) on gene expression measurements (*43*). To determine the optimal number of SVs to include in our final model, we randomly split our dataset into equal halves and calculated differential expression for each gene and disorder using a fixed number of SVs (**Fig S3A**). We then compared the replicability of differential expression (log_2FC) effect sizes between the two split halves of the dataset, quantified using spearman's correlation. This analysis was repeated 1000 times each for a fixed number of SVs increasing from 0 to 25. We found that including 4 SVs in addition to the final set of known covariates above maximized this split-dataset replicability (**Fig S3**). As such, our final model used for all differential gene expression, isoform expression, and splicing analyses consisted of: diagnosis, age, age^2, study/batch, sex, PMI, RIN, RIN^2, brain bank, brain region, seqPCs (1-3, 5-8, 10-14, 16, 18-25, 27-29), seqPC3^2, and SVs (1-4).

Differential Gene and Transcript Expression/Usage

Count level quantifications were corrected for library size using TMM normalization in *edgeR* and were transformed as log_2(CPM+0.5). DGE was then calculated using a linear mixed-effects model using the *nlme* package in R. The covariates specified in the previous section were included as fixed effects in the model. In addition, we included a random effect term for each unique subject to account for subject overlap across sequencing studies. Resulting P-values were FDR-corrected using the Benjamini-Hochberg method, to control for multiple comparisons. Differential transcript expression (DTE) was calculated similarly as for DGE except that the transcript-level quantifications from RSEM were used as inputs for the linear mixed-effects model. Finally, differential transcript usage (DTU) was calculated similarly as for DGE except that isoform percentage data reported by RSEM was used as inputs for the linear mixed-effects model.

To ensure the robustness of DGE results, we compared log_2FC effect size measurements for genes identified as significantly differentially expressed in several previous studies profiling gene expression using cortical brain samples from ASD, SCZ, and BD (**Fig S4A**-**D**). Finally, to ensure that differential gene expression in disease was not being driven by subtle differences in RNA quality or degradation, we compared differential expression T-statistics with those experimentally derived from brain tissue samples allowed to degrade for fixed intervals of time (**Fig 4E**) (*22*). We did not observe substantial concordance between these RNA degradation metrics and psychiatric disease DGE summary statistics.

Enrichment Analysis of Gene Sets

Enrichment for Gene Ontology (GO; biological process, molecular function and cellular component) and KEGG pathways was performed using the *gProfileR* R v0.6.4 package (*103*). Only pathways containing less than 1000 genes were assessed. Background was restricted to brain expressed genes. An ordered query was used, ranking genes by log_2FC for DE analyses or by kME for coexpression module enrichment analyses. P-values were FDR corrected to account for multiple comparisons.

Enrichment analyses were also performed using several established, hypothesis-driven gene sets including: high confidence ASD risk loci (*80*); CHD8 targets (*96*); FMRP targets (*33*); mutationally constrained

genes (*97*); syndromic and highly ranked (1 and 2) genes from the SFARI GENE database; 'vulnerable' ASD genes (*98*); genes with probability of loss-of-function intolerance (pLI) > 0.99 as reported by the Exome Aggregation Consortium (*104*). Statistical enrichment analyses were performed using logistic regression, correcting for both gene length and GC content. All results were FDR-corrected for multiple comparisons.

Cell Type Enrichment Analyses

Cell type enrichment analyses were performed using uniformly processed human brain single-cell RNA-Seq datasets, compiled by the companion manuscript (*18*) which combined multiple published datasets (*91*, *105*, *106*) with newly generated data from PsychENCODE. Clustering was performed separately for single-cell datasets using TPM and UMI quantifications. See (*18*) for further details. Enrichment was performed for cell type specific marker genes using Fisher's exact test, followed by FDR-correction for multiple testing.

For neural-immune modules (**Fig 7)**, we additionally assessed several mouse experimentally derived cell type specific expression datasets. These included: a translating ribosome affinity purification (TRAP) dataset profiling 24 genetically identified populations of CNS cell types in mouse using microarray (*107*); a large-scale single-cell RNA-Seq dataset of mouse somatosensory cortex and hippocampus (*94*); a MARS-seq dataset of FACS-sorted CD45+ cells from mouse brain tissue, representing the major CNS immune cell populations (*93*); and a single cell RNA-Seq dataset of cells derived from meninges and choroid plexus in mouse (*92*).

Differential Local Splicing (DS) analysis

Local splicing analysis used LeafCutter (*29*), which detects splicing variation using the sequencing reads that span an intron (or spliced reads) to quantify intron usage across samples, without relying on existing reference annotations and without estimation of isoform abundance or exon inclusion levels. The same BAM alignment files to the hg19 genome assembly produced by STAR (version 2.4.2a) (*108*) for the DGE/DTE analyses were used as input for leafcutter intron clustering. The BAM files included the XS strand tags to all canonically spliced alignments based on their intron motifs (parameters: alignSJoverhangMin =8, outSAMstrandField =intronMotif). We used LeafCutter to first call clusters of variable spliced introns across all our samples and then to identify differential splicing between each disorder (ASD, SCZ, and BD) and the control (CTL) group by jointly modeling intron clusters using the Dirichlet-Multinomial generalized linear model (GLM) (*29*). We controlled for the same technical, biological covariates and hidden confounds as described above in the DGE/ DTE analyses, except that we did not incorporate a random term for individuals (random effects are not supported by the Dirichlet-Multinomial GLM). Accordingly, we also removed tissue sample replicates that were sequenced in more than one study, randomly retaining only one sample in our analysis. The dataset for LeafCutter analysis numbered 944 controls, 79 ASD, 531 SCZ and 217 BD samples (1,771 total).

We used LeafCutter to call intron clusters as follows: overlapping introns, defined as spliced reads, were clustered and filtered to keep intron clusters supported by at least 50 split reads across all 1,771 samples, retaining introns of up to 100 kb and accounting for at least 1% of the total number of reads in the entire cluster. This yielded 37,215 clusters encompassing 120,921 introns in 17,342 genes that were used for further analysis. This intron count file was then used in the differential splicing (DS) analysis.

DS intron clusters were identified in pairwise analyses comparing each psychiatric disorder (ASD, BD, SCZ) to the common set of 944 control samples. After discarding introns that were not supported by at least one read in 5 or more samples, clusters were analyzed for DS if at least 3 samples in each comparison group (*i.e*. cases or controls) had an overall coverage of 20 or more reads. *P*-values were corrected for multiple testing using the Benjamini-Hochberg (BH) method and used to select clusters with significant splicing differences (FDR $q < 0.1$).

Percent-spliced-in (PSI) values were corrected for covariates using the *quantify_PSI* function provided in the LeafCutter *psi* branch (https://github.com/davidaknowles/ leafcutter/tree/psi). Violin plots of intron PSI values were prepared using ggplot2. Principal component analysis (PCA) plots were evaluated before and after covariate-correction (**Fig S6**). Schematic visualization of significant intron clusters was done using the *leafviz* R shiny package [https://davidaknowles.github.io/leafcutter/articles/Visualization.html]. All DS events were further annotated using *leafviz* and custom R code, manually inspected, and classified into single- or multi-exon skipping events (changes in cassette splicing), alternative 5' and 3' exon usage, and alternative 5' (donor) or 3' (acceptor) splice site usage. Intron clusters that did not match any of these categories were classified as complex events involving multiple changes. DS intron clusters were mapped onto transcripts using *gViz* (v3.7) (*109*) and *ensembldb* (v3.7) (https://github.com/jotsetung/ensembldb) bioconductor R packages. SMART (*110*) and PFAM (*111*) protein domains were mapped onto transcript structures using the *proteinToGenome* function of *ensembldb*.

DeltaPSI (ΔPSI) Correlation Across Disorders

To determine significance for the correlation of ΔPSI across disorders for significant intron clusters identified by LeafCutter, we permuted the case/control status within each disorder 3,000 times and repeated the LeafCutter analysis with the same GLM described above. In each permutation we assessed the ΔPSI correlation between disorders using Spearman's correlation (ρ) to yield a null distribution of ρ values that was used to assess the significance of the observed correlations.

Cross-disorder DS Overlaps

For cross-disorder DS overlaps we selected all genes associated with significant intron clusters identified by LeafCutter at FDR <10%. Venn diagrams area-proportional to the number of genes with significant DS clusters in each disorder were then created using the *eulerr* R package. Hypergeometric p-values for pairwise overlaps between disorders were calculated using the *phyper* function in R and setting the size of the 'universe' to all genes with intron clusters meeting the LeafCutter clustering criteria.

Functional Enrichment of DS Genes

Gene set enrichment for Gene Ontology (GO) biological process, molecular function and cellular component aspects was performed using the *gProfileR* v0.6.4 package in R (*103*) with moderate hierarchical filtering and using an ordered query, after ranking genes in increasing order of the LeafCutter p-value (i.e. most significant at the top). In case a gene had multiple significant intron clusters, the most significant cluster was used for the ranking. The custom background set for each disorder consisted of all 10,677 genes with intron clusters that were evaluated in pairwise LeafCutter analyses between each disorder and control groups, as described above. Visualization of enriched GO terms used custom ggplot2 functions.

Gene-set enrichment for RBFOX1 targets (*32*), FMRP targets (*33*), and a curated list of genes coding for RNA binding proteins (RBPs) (*34*), used the same custom background of genes with LeafCutter intron clusters as in the GO enrichment analysis, and was assessed using Fisher's exact test and correcting for multiple testing by FDR. The curated list of RBP genes included those with high confidence for RNA binding (*112–114*) and those annotated as RNA-binding in Ensembl including known and potentially auxiliary splice factors (*34*).

The comparison of LeafCutter DS events with Parishak *et al*. 2016 (*19*) MATS events was based on genomic coordinates overlap using BEDtools, irrespective of the event type assigned by each algorithm.

Microexon Enrichment

Transcripts that carry at least one exon of 3-27 nucleotides in length (*i.e.* microexon) (*30*) were extracted from the Gencode V19 database. Statistical enrichment analyses were performed using logistic

regression, correcting for both gene and transcript length on linear and log10 scales. As an additional control, transcripts that carry exon(s) of average length (143±5% nucleotides) were also extracted and their overlap with switch transcripts was also tested with logistic regression. All results were FDR-corrected for multiple comparisons by the Benjamini-Hochberg method.

Construction of Disease-specific Isoform-level Co-expressed PPI Networks

Pairwise spearman correlation coefficients (SCC) between transcript of interest and all other transcripts were calculated using either ASD samples or SCZ samples. To obtain age-balanced datasets, only samples from donors of age 17-67 years old were used. A cutoff of SCC > 0.5 was used to filter for the co-expressed partners of the transcript of interest in either ASD samples or SCZ samples. PPI data was compiled from well-characterized PPI databases, including Bioplex (*115*), HPRD (*116*), Inweb (*117*), HINT (*118*), Biogrid (*119*), GeneMANIA (*120*), STRING (*121*) and CORUM (*122*). Only physical interactions and co-complex associations were kept. Co-expressed partners which are also supported by PPI were used to construct the co-expressed PPI network.

ncRNA Annotation

To identify ncRNAs that may be relevant to neuropsychiatric disorders, we compiled a list of non-protein-coding genes exhibiting differential gene (DGE) or transcript expression (DTE) at FDR < 0.05 in at least one disorder (**Table S2**). As ncRNA designation can change based on genomic annotation, we filtered out genes that were designated as protein coding in the most recent version of Gencode v27, yielding a total of 944 unique ncRNAs, many of which were differentially expressed across more than one disorder or at both gene and transcript-level features. Differentially expressed ncRNAs were annotated according to sequence and expression characteristics. Human tissue-specific expression was assessed using data from GTEX v6. Median RPKM values per tissue were obtained and averaged into broad categories (**Fig 1F**). To identify ncRNAs broadly expressed across human tissues, we ran an ANOVA on log2(RPKM +1) values across tissues, and selected those with uncorrected P > 0.05. Brain-specific expression was defined as $RPKM_{brain}$ / sum($RPKM_{all\ tisues}$) > 0.8. CNS cell type specificity was assessed in a similar fashion using single-nucleus RNA-Seq from the Lake dataset (*91*). Expression counts were CPM normalized and then averaged together across defined cell clusters. We ran an ANOVA on log2(CPM + 1) values across cell clusters, and report those ncRNAs with P>0.05 as “broadly expressed” with regard to cell type. Cell type specificity was quantified by $CPM_{max\ cluster}$ / sum($CPM_{all\ cell\ clusters}$) > 0.8.

Evolutionary conservation was assessed using phastCons and phyloP scores (*123*, *124*). Both methods assign a score to each base in the human genome, quantifying its degree of conservation across selected species. Whereas phastCons base scores are smoothed according to scores of neighboring bases, phyloP evaluates each base independently. We downloaded phastCons and phyloP per-base scores for hg19 from UCSC, computed from 17-way (primate), 30-way (mammalian), and 100-way (vertebrate) Multiz alignments, to calculate a mean base score for each ncRNA across 1) the gene, and 2) its exonic regions only. The per-exon scores were averaged over all exons belonging to a gene to produce a more robust metric for gene conservation.

Context-dependent tolerance (CDTS) scores were used to quantify patterns of human selective constraint (*27*). CDTS scores are computed for each 10bp window in high-confidence regions of the genome, which we intersected with exonic coordinates for ncRNAs using Bedtools (*125*). To produce a per-gene score, we first computed the mean across all 10bp windows intersecting a single exon, then averaged the mean exon scores across all exons for a gene.

Signed Gene and Isoform Co-Expression Network Analysis

To place results from individual genes within their systems-level network architecture, we performed Weighted Gene Co-Expression Network Analysis (WGCNA) separately for gene- and isoform-level quantifications (*49*). All covariates except for diagnostic group were first regressed from our expression dataset. Network analysis was performed with the WGCNA package using signed networks. A soft-threshold power of 7 was used for all studies to achieve approximate scale-free topology (R^2>0.8). Networks were constructed using the blockwiseModules function. The network dendrogram was created using average linkage hierarchical clustering of the topological overlap dissimilarity matrix (1-TOM). Modules were defined as branches of the dendrogram using the hybrid dynamic tree-cutting method. Modules were summarized by their first principal component (ME, module eigengene) and modules with eigengene correlations of >0.9 were merged together. A robust version of WGCNA (rWGCNA) was run to reduce the influence of potential outlier samples on network architecture (*126*). Module robustness was ensured by randomly resampling (2/3 of the total) from the initial set of samples 100 times followed by consensus network analysis, a meta-analytic approach, to define modules using a consensus quantile threshold of 0.2. Modules were defined using biweight midcorrelation (bicor), with a minimum module size of 50, deepsplit of 4, merge threshold of 0.1, and negative pamStage. Module (eigengene)-disease associations were evaluated using a linear mixed-effects model, using a random effect of subject, to account for subject overlap across datasets. Significance values were FDR-corrected to account for multiple comparisons. Results from module-eigengene association tests are shown in **Fig 5**. Genes within each module were prioritized based on their module membership (kME), defined as correlation to the module eigengene. The top 'hub' genes for several of the modules are shown in **Figs 5**-**7** and through an interactive portal on our companion website (Resource.PsychENCODE.org).

The robustness of all network modules were tested as described previously (*48*). In brief, each module's density (defined as the average intramodular topological overlap) was compared to the density of modules of equivalent size selected randomly from the same network (n = 5,000 permutations). Density p-values were determined for each initial module by calculating the percentage of trials in which the density of the "random" modules exceeded the density of the initial module. All modules have density p-values less than 0.05.

csuWGCNA

We also used a modified version of WGCNA named *Combination of Signed and Unsigned WGCNA* (csuWGCNA), which captures strong and moderate negative correlations in the coexpression network (*72*). Current versions of WGCNA allow unsigned, signed and signed hybrid options for network types, but have disadvantages when trying to capture moderate negatively correlated features such as lncRNA-mRNA regulatory relationships. Signed and signed hybrid networks down-weight negatively correlated pairs in network. Unsigned networks highlight strong positive and negative correlations, but has worse performance on identifying functionally-related gene pathways than its signed counterparts (*127*). To address these limitations, we modified two functions for picking soft thresholding power and calculating the network adjacency. The core modification of csuWGCNA is in its definition of adjacency, $a_{ij} = ((1 + |cor(x_i , x_j)|)/2)^{\beta}$, which integrates the advantages of signed networks ($a_{ij} = |(1 + cor(x_i , x_j))/2|^{\beta}$) and unsigned networks ($a_{ij} = |cor(x_i , x_j)|^{\beta}$). Using this adjacency function, csuWGCNA then constructs a topological overlap matrix (TOM) and follows the procedure described above for clustering, tree cutting, and network module detection. Using this method, csuWGCNA can detect modules containing genes with negative correlations, which may be more useful when lncRNAs and miRNAs are included in the network (**Fig S14**).

Assessment of Psychiatric Medications

To assess the potential impact of medications on differential expression and co-expression results, we analyzed several published datasets of animal models exposed to multiple classes of psychiatric medications. These included: 1) a published RNA-Seq dataset of the DLPFC from non-human primates exposed for six months to haloperidol, clozapine, or placebo (*2*); 2) a published microarray dataset (GSE66276) of cortex from mice exposed to the SSRI fluoxetine for 21 days (*128*); 3) a microarray dataset (GSE66276) of rats exposed lithium, lamotrigine or placebo for 21 days. All datasets were reprocessed and analyzed as described below.

The antipsychotic dataset consisted of ~6 year old rhesus macaques treated with medications or placebo orally for six months, including high doses of haloperidol (4 mg/kg/d; n=7), low doses of haloperidol (0.14 mg/kg/d; n=10), clozapine (5.2 mg/kg/d; n=9), or vehicle (n=8). DLPFC tissue was extracted and RNA-Seq was run using rRNA-depleted libraries. Genes were kept that had expression greater than 0.1 cpm (counts per million) in at least half of samples. Limma voom normalization using TMM normalization factors was used for subsequent differential gene expression analysis, including the covariates: age, sex, sequencing batch factors, RNA quality statistics (RIN and RNA concentration) and sequencing statistics. In accordance with results from (*2*), none of the groups (clozapine, haloperidol_low_dose, and haloperidol_high_dose) had any genes significantly differentially expressed from placebo after FDR-correcting for multiple comparisons. We therefore used an unadjusted p-value threshold of 0.01 for downstream analyses, resulting in 133, 120, and 188 genes for clozapine, haloperidol_low_dose, and haloperidol_high_dose, respectively. Genes were grouped based on direction of effect (up or downregulated) and mapped to human orthologues using Ensembl. Overlap with PsychENCODE disease gene sets (DE and DS genes, gene and isoform-level coexpression modules) was assessed using Fisher's exact test followed by FDR correction. Although some gene sets showed nominal overlap with antipsychotic genes, no enrichments were significant after correction for multiple comparisons (**Fig S11**).

In the SSRI dataset, thirty mouse strains were treated for 21 days with fluoxetine (18 mg/kg per day) or vehicle, cortical RNA was extracted and profiled with an Affymetrix expression microarray (GeneChip Mouse Genome 430 2.0 array). Raw microarray data was normalized using the RMA function from the 'affy' package in R. Batch correction was performed using ComBat, and differential expression was detected using the lmFit and eBayes functions from the 'limma' R package (covariates for the linear model: fluoxetine treatment, strain, and RNA degradation score). In our analysis, only two genes were found to be significantly differentially expressed (downregulated) in the fluoxetine group after correction for multiple comparisons (FDR p-value < 0.05): SST, a hormone regulating factor, and FDFT1, an enzyme involved in cholesterol biosynthesis. For downstream enrichment analyses, we used the relaxed threshold of $p < 0.01$ (uncorrected), corresponding to 558 genes.

In the third dataset, rats (n=5 per group) were administered lithium in chow (0.2%) or lamotrigine via subcutaneous injection (30 mg/kg) and compared to a vehicle chow group or vehicle injection group. All regimens were administered once daily for 21 days and tissues were collected from frontal cortex, striatum, and hippocampus for analysis via an Affymetrix expression microarray (Affymetrix Rat Genome 230 2.0 Array). Differential expression analysis was performed as described above. For lamotrigine, no genes were differentially expressed following FDR-correction, so for downstream enrichment analyses we used the relaxed threshold of $p < 0.01$ corresponding to 121 genes. For lithium, 2338 genes at FDR-corrected $p < 0.05$ were used for downstream enrichment analyses.

Assessment of Non-linear Age Effects

To assess the influence of age on the magnitude of differential expression, and to account for potentially non-linear effects of age, we performed a local regression analysis using the locfit package in R. For

each gene expression measure, a local regression function was fit to model the effect of age on expression in control samples, as follows: `fit = locfit(Expr ~ Age, data=df[df$Group=="CTL",])`
For each non-control sample, expression was then converted to a z-score using the interpolated mean expression in controls at the same age. We then assessed the correlation between z-transformed expression and age within each disease group (ASD, SCZ, BD) separately, to identify those genes whose magnitude of differential expression was associated with age. Several examples are shown in **Fig S10**.

GWAS Datasets

We performed a number of GWAS enrichment analyses as described in the following sections. In each analysis, we used summary statistics from the largest publicly available GWAS in SCZ (*58*), ASD (*38*), and BD (*90*). Additional secondary analyses were performed using a variety of relevant traits, including major depressive disorder (MDD; ref (*129*)), neuroticism (*130*), educational attainment (*131*), diabetes (*132*), as well as previous GWA studies of SCZ (*133*), ASD (*134*), and BD (*135*).

GWAS Enrichment in DE Genes and Modules

We used stratified LD score regression (s-LDSR) (*39*) to investigate whether differentially expressed or spliced genes, and/or co-expression modules, are enriched for disease-associated genetic variation using the summary statistics described above. SNPs were assigned to these custom gene categories if they fell within ±10 kb of a gene in the set. These categories were added to a 'full baseline model' that includes 53 functional categories capturing a broad set of genomic annotations, as published (*39*). Enrichment was calculated as the proportion of SNP heritability accounted for by each module divided by the proportion of total SNPs within the module. Significance was assessed using a block jackknife procedure, followed by FDR correction of P values.

Polygenic Risk Score Calculation

Polygenic risk scores (PRS) were calculated using the same GWAS summary statistics as above, for SCZ (*58*), BD (*90*) and ASD datasets (*38*). Samples were restricted to those of European ancestry based on clustering with samples from HapMap3 (*136*). Genotypes were additionally filtered as follows, using plink: `plink --bfile PECDC_EUR --geno 0 --maf 0.05 --hardy --hwe 1e-40 --make-bed –out PECDC_EUR_PRSfilter.` To calculate PRS, we used LDpred (*137*) with the 1000 Genomes phase 3 European subset as a reference panel. The first five genotype principal components (gPC1-5, as calculated with plink) were included in the PRS calculation, to account for ancestry and technical effects. We then compared PRS for each diagnostic group with the strict set of non-psychiatric controls, contrasting baseline and full models. PRS significance was measured with a likelihood ratio test and Nagelkerke's pseudo-R^2.

```
mod.baseline=glm(dx~study+sex+age+gPC1+gPC2+gPC3+gPC4+gPC5,family=binomial)
mod.full=glm(dx~PRS+study+sex+age+gPC1+gPC2+gPC3+gPC4+gPC5,family=binomial)
adjustedR2=NagelkerkeR2(mod.full)$R2-NagelkerkeR2(mod.baseline)$R2
prs.significance=lrtest(mod.baseline, mod.full)
```

The default LDpred GWAS p-value thresholds were used (.001, .003, .01, .03, .1, .3, 1, and Inf). Maximal Nagelkerke pseudo-R^2 values were achieved for prediction of psychiatric diagnosis using thresholds of 0.001 for ASD, 0.01 for BD, and 1 for SCZ.

Association between PRS and measures of gene, isoform, or module (eigengene) expression was performed as described above, except using linear regression analogs. Associations were repeated for each p-value threshold in the 3 GWAS studies and the resulting association p-values were then FDR-corrected for multiple testing. Full results are compiled in **Table S4**.

Transcriptome-wide Association Study (TWAS)

To identify genes whose *cis*-regulated expression is associated with disease, we performed a transcriptome wide association study (TWAS) to identify putative molecular (e.g., *cis*-eQTL) phenotypes in brain underlying disease GWAS associations (*46*). TWAS was implemented using the fusion package (https://github.com/gusevlab/fusion_twas; (*46*)) with custom SNP-expression weights generated from brain using our dataset of 1321 unique individuals of European ancestry with imputed genotypes. Using the AI-REML algorithm implemented in GCTA (*138*) by the fusion package, we first identified the subset (n=14,750) of total expressed genes found to have significant *cis* SNP-heritability in our dataset (*cis*- h^2_g P<0.05 within 1 Mb window around the gene body). SNP-expression weights were calculated in a 1Mb region around all heritable genes using expression measurements adjusted for diagnosis, study, age, age^2, RIN, RIN^2, sex, tissue, PMI, 20 ancestry PCs, and 100 hidden covariates (*139*). Accuracy of five expression prediction models were tested (best *cis*-eQTL, best linear unbiased predictor, Bayesian linear mixed model, Elastic-net regression, LASSO regression) using the most accurate model for final weight calculations as implemented in fusion. TWAS disease-association statistics were computed using these custom weights, LD structure calculated from our PsychENCODE samples' genotypes, and disease GWAS summary statistics described above. For each disease, TWAS association statistics were Bonferroni-corrected for multiple comparisons. Full results are compiled in **Table S4**.

Rare Variant Enrichment Analyses

Gene and isoform co-expression modules were also assessed for enrichment of rare variants identified in disease, compiled from several datasets. These included: 71 risk loci harboring rare *de novo* variants associated with ASD through the transmission and *de novo* association test (TADA) (*80*); Syndromic and highly ranked (1 and 2) genes from SFARI Gene database; genes harboring recurrent *de novo* copy-number variants associated with ASD or SCZ, as defined in (*1*); genes harboring an excess of rare exonic variants in ASD, SCZ, intellectual disability (ID), developmental delay (DD), and epilepsy as assessment through an extended version of TADA (extTADA) (*140*); genes harboring disruptive and damaging ultra-rare variants (dURVs) in SCZ (*54*); a list of high confidence epilepsy risk genes, compiled in (*141*). For binary gene sets, enrichment among gene and isoform modules was calculated using logistic regression, correcting for linear- and log-transformed gene and transcript lengths as well as GC content. For dURVs, a two step procedure was used, first creating a logistic regression model for dURV genes identified in controls and a second model for those affected in cases and controls. A likelihood ratio test was used to calculate significance. Finally, for the extTADA datasets, the posterior-probability (PP) was used in the logistic regression model in place of a binary annotation. P-values were FDR-corrected for multiple comparisons. Results are shown in **Fig S13** and compiled in **Table S5**.

Experimental Validation

Initial optimization of the PCR conditions for all splicing and isoform primers used cDNA samples derived from total brain or cortex RNA (Clontech), and were performed on a Mastercycler Nexus Gradient Thermal Cycler (Eppendorf) and amplicons resolved on precast 96-well 2% agarose E-Gels (Invitrogen) stained with SYBR safe.

Splicing validation

For differential splicing (DS) analysis, selected exon-skipping events were validated by semiquantitative RT–PCR in ASD, BD, SCZ, and control samples. Total RNA (1-2 µg) was treated with 1 unit of Baseline-ZERO DNase (Lucigen), cleaned up with 1.8x AMPure XP (Beckman Coulter), and reverse-transcribed using

SuperScript III reverse transcriptase and random hexamer primers (Invitrogen). After clean-up with 1.8x AMPure XP, DS events were PCR amplified from 20 ng of cDNA for 30 cycles in 25 µl volume containing exon-specific primers at a concentration 0.5 µM each, and ChoiceTaq Blue MasterMix (Denville) according to manufacturer instructions. Exon-specific PCR primers (**Table S8**) were designed in the flanking exons of each skipping event using Primer3 (*142*) and BLAST (*143*). PCR products were cleaned up with 1.8x AMPure XP (Beckman Coulter) and analyzed on DNA 1000 chips on an Agilent 2100 Bioanalyzer system. Peaks corresponding to the amplicon including or excluding the skipped exon were quantified using the Bioanalyzer Expert software, and percent spliced in (PSI) ratios were calculated by dividing the molarity of the lower band (exon skipped) by the sum of the molarity of the lower and upper band (exon included). The ΔPSI between cases and control for each event was calculated as the difference between the average PSI in cases and average PSI in controls. Sample details and primers are reported in **Table S8**.

Isoforms Validation

For DTE analysis, selected isoforms were validated by semiquantitative RT–PCR using a similar approach as for DS. Each isoform was PCR amplified from 20 or 40 ng of cDNA for 30 or 35 cycles in a 25 µL volume containing isoform-specific primers at a concentration 0.5 µM each and ChoiceTaq Blue MasterMix with DNA polymerase (Denville), or 0.4 µM each and LongAmp Hot Start Taq DNA polymerase (New England Biolabs) (**Table S8**), according to manufacturer instructions. Isoform-specific PCR primers (**Table S8**) were designed using Primer3 (*142*) and BLAST (*143*), and based on GENCODE v19 annotations. PCR products were resolved on 1.5 or 2% agarose gels, counterstained with GelStar Nucleic Acid Gel Stain (Lonza) for visualization, and *GAPDH* and *ACTB* were used as loading controls. Gels were quantified using ImageLab (BioRad). The intensity of each isoform was first normalized to the average expression levels of *GAPDH* and *ACTB* in each sample. The intensity ratio between cases and controls for each isoform was then calculated by dividing the average intensity of cases by the average intensity of controls. The log2 intensity ratios were then compared to the log2 ratio differences from the DTE analysis. Sample details and primers are reported in **Table S8**.

Fluorescent *in situ* hybridization (FISH)

Fresh-frozen tissue blocks from the Brodmann's area 9 of the prefrontal cortex of five neurologically normal control donors were obtained from the Mount Sinai Neuropathology Research Core and Brain Bank and stored at -80ºC. Clinical information on the subjects is summarized in **Table S9A**. The blocks were embedded in O.C.T. compound, frozen at -20ºC, 10 µm-thick sections were cut using a cryostat (Leica), and the sections were collected onto Superfrost Plus slides. The slides were stored in an airtight box at -80ºC until FISH was conducted.

The *in situ* hybridization probes for detecting human *GAD1*, *LINC00643*, and *LINC01166* as well as the positive and negative control probes were designed by Advanced Cell Diagnostics (ACD; see **Table S9B** for RNAscope probe information). For the assay, we used RNAscope Multiplex Fluorescent Reagent Kit v2 (ACD), that provided the hydrogen peroxide, protease IV, amplification reagents (Amp1-3), HRP reagents, and wash buffer for probe hybridization. DAPI, TSA buffer (ACD) and TSA Plus fluorophores (PerkinElmer) were used for detection of the signal. We used a modified version of the manufacturer's protocol for sample preparation, probe hybridization, and signal detection. Briefly, the fresh frozen sections on slides were retrieved from -80ºC and immediately fixed by immersion in freshly prepared cold 4% paraformaldehyde for 2 h. After fixation, the sections were rinsed briefly with phosphate buffered saline (PBS) and then dehydrated in an ethanol series (5 min each in 50%, 70%, and two changes of 100% ethanol) at room temperature (RT). The sections were air-dried for 5 min and a hydrophobic barrier was created around the section using an Immedge pen (Vector Laboratories). After the barrier had completely dried, the sections were treated with hydrogen peroxide for 10 min at RT, washed twice with PBS, treated with protease IV for 15 min at RT, and washed twice with PBS. The

LINC-C2 or LINC-C3 probes for detecting lncRNAs were diluted at 1:50 in the GAD1-C1 probe. The sections were then hybridized with the probes at 40ºC for 2 h in the HybEZ Hybridization System (ACD), washed twice with wash buffer, and stored overnight at RT in 5x SSC buffer. The next day, the slides were rinsed twice with wash buffer, followed by the three amplification steps (AMP 1, AMP 2, and AMP 3 at 40ºC for 30, 30, and 15 min respectively, with two washes of 2 min each with wash buffer after each amplification step). The signal was developed by treating the sections in sequence with the HRP reagent corresponding to each channel (e.g. HRP-C1) at 40ºC for 15 min, followed by the TSA Plus fluorophore assigned to the probe channel (fluorescein for GAD1-C1 probe and cyanine 5 or Cy5 for LINC-C3 probes, prepared at a dilution of 1:750) at 40ºC for 30 min, and HRP blocker at 40ºC for 15 min, again with two wash steps after each of the incubation steps. As autofluorescence due to lipofuscin was detected in both the green and the red channels whereas the far red channel was relatively free of background, the highly expressed GAD1-C1 probe was assigned to the green fluorescein channel, the red cyanine 3 channel was left empty and the lncRNAs were probed on separate sections in the far red Cy5 channel. The sections were treated with TrueBlack Lipofuscin Autofluorescence Quencher (Biotium) for 30 s, rinsed twice with PBS, counterstained with DAPI for 30 s, mounted using ProLong Gold mounting medium (Thermo Fisher Scientific) and slides were stored at 4 ºC until ready for imaging. Two experiments were performed with two to three biological replicates each, and using positive and negative control probes to test for RNA quality and background signal respectively.

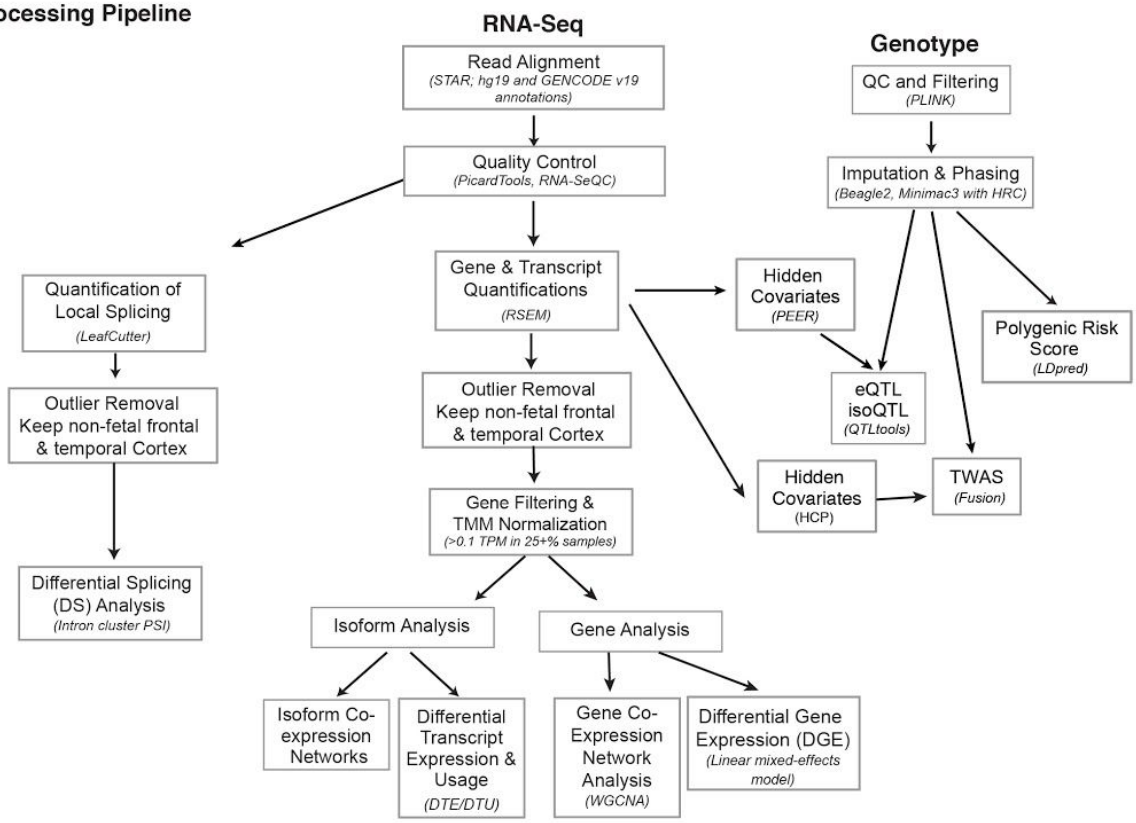
Layer III of area 9 was identified using a 5x/0.16 N.A. objective and the sections were imaged using a 63x/1.4 N.A. or 100x/1.4 N.A. oil DIC Plan Apochromat objectives on an AxioImager.M2 microscope (Carl Zeiss), equipped with a motorized stage (MBF Biosciences) and an Orca-R^2 digital camera (Hamamatsu), and operated using Neurolucida software (version 11.11.3 64-bit, MBF Biosciences). Camera exposure times were set for each of the four channels (red for lipofuscin, blue for DAPI, green for fluorescein, and magenta for Cy5) and were kept similar among the cases imaged in each experiment in order to enable comparison. The images at 100x magnification were presented as maximum intensity projections of Z-stacks imaged at 0.5 µm intervals. Adobe Photoshop was used for adjusting brightness/contrast and sharpness of the images.

Figs S1 to S17

A Individual Studies

Dataset	Groups	Samples in Dataset	Samples in final analysis	Tissue	Library	Stranded	RNAseq protocol	Platform	Genotyping Platform	Published / Reference
BrainGVEX	SCZ BD CTL	430	409	DLPFC	rRNA - dep	stranded (reverse)	100bp Paired End	HiSeq 2000 / 2500 / 4000	Affy 5.0 and Illumina PsychChip	153 samples included as replication in PMID 29439242
BrainSpan	CTL	606	163	many	polyA	unstranded	76bp Single End	GAIIx	N/A	Companion Manuscript
Common Mind	SCZ BD CTL	613	572	BA9/46	rRNA - dep	unstranded	100bp Paired End	HiSeq2000	Infinium Human Omni Express Exome	PMID 27668389
UCLA -ASD	ASD CTL	253	151	BA9/46, BA41, CBL	rRNA - dep	unstranded	50bp Paired End	HiSeq2000	Infinium Omni 2.5 Exome, Infinium Omni 2.5	PMID 27919067
Yale -ASD	ASD CTL	45	12	DLPFC, TC, V1, CBL	rRNA - dep	stranded (reverse)	100bp Paired End	HiSeq2500	N/A	Unpublished
BipSeq	BD	69	65	BA9/46	polyA	unstranded	100bp Paired End	HiSeq2000	Illumina_1M, Illumina_h650	Unpublished
LIBD szControl	SCZ CTL	495	422	DLPFC	polyA	unstranded	100bp Paired End	HiSeq2000	Illumina 1M, Illumina_h650, Illumina_Omni5	PMID 30050107
CMC_HBCC	SCZ BD CTL	387	366	DLPFC	rRNA - dep	stranded	100bp Paired End	HiSeq2500	Illumina 1M, Illumina_h650, Illumina_Omni5	Unpublished

B Processing Pipeline



C RNA-Seq Sample Overlap

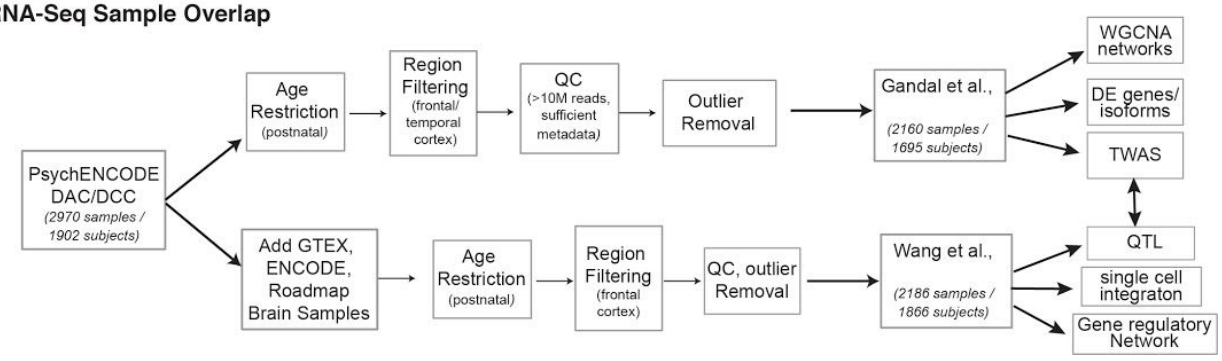


Figure S1 -- Dataset composition, analysis and integration pipeline

A) Description of individual studies contributing to this PsychENCODE analysis. B) Analysis pipeline through which all samples were uniformly processed. C) Comparison of samples overlapping between this manuscript and our companion paper (*18*).

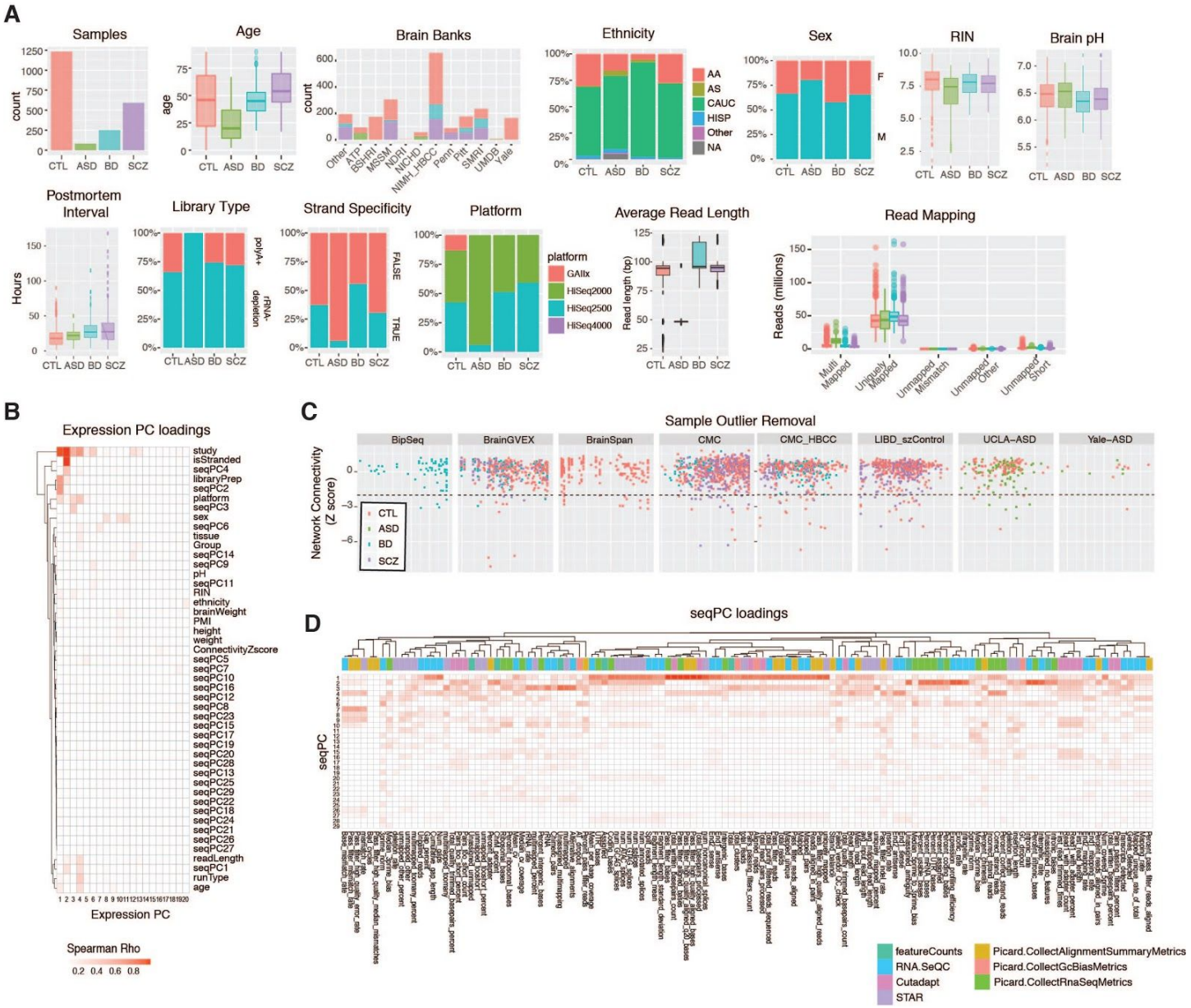


Figure S2 -- Dataset demographics and quality control

A) Sample information, subject demographics, and sequencing characteristics are shown for each group. This study only used frontal and temporal cortex samples from subjects at postnatal time points. B) Spearman's ρ values are shown for correlations between dataset covariates with the top 20 expression PCs. C) Sample outlier removal was performed individually for each study before combining data, based on Z-scores of standardized network connectivity (**Methods**). D) Sequencing surrogate variables ('seqPCs') were calculated as the top 29 principal components of the matrix of sequencing QC metrics. Loadings are shown between seqPCs and individual metrics, colored by the source of the QC metrics.

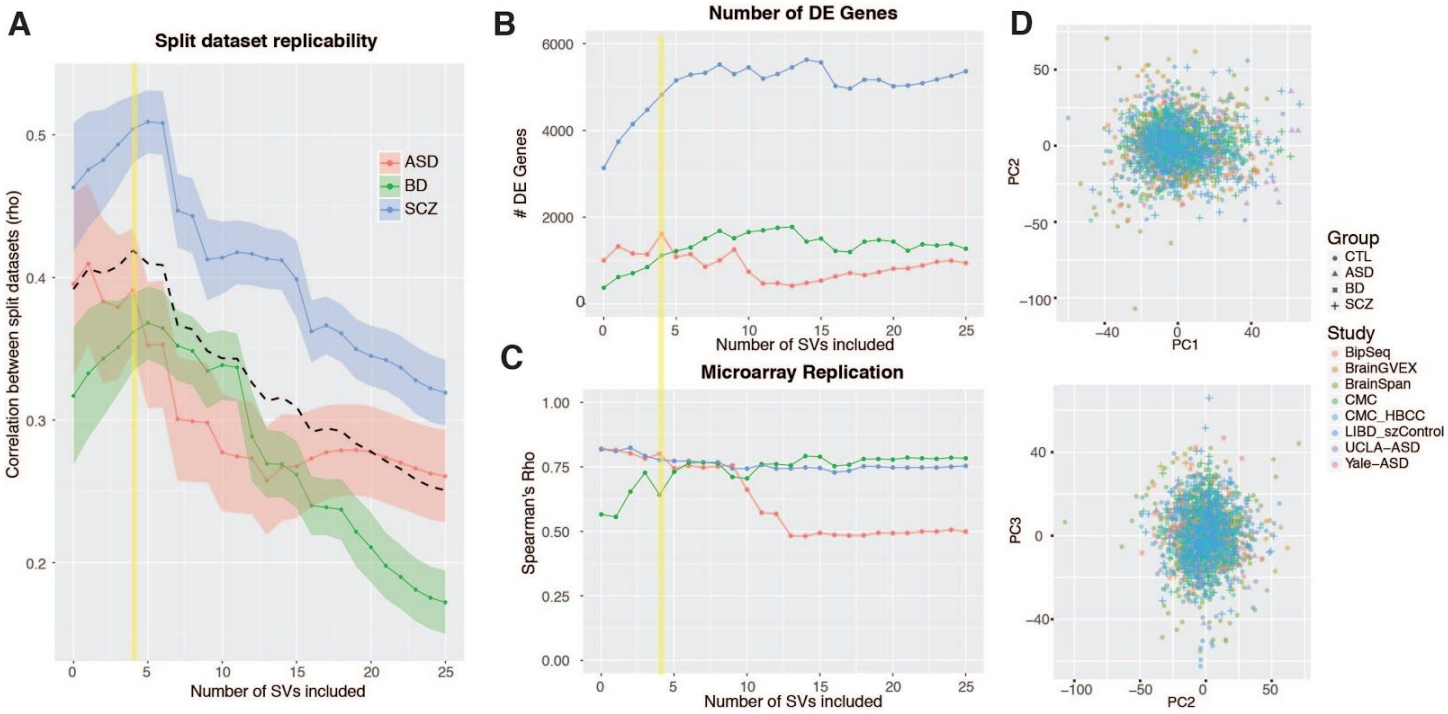


Figure S3 -- Selection of covariates

To capture the full range of factors influencing gene expression in our dataset, the final differential expression model included known covariates, aggregate sequencing metrics (seqPCs), and surrogate variables (SVs) calculated using SVA to correct for unmeasured sources of variation. A) To determine the appropriate number of SVs to use, we randomly split our dataset into two halves and calculated differential gene expression for each disorder using a fixed number of SVs ranging from 0 to 25. We compared DGE for each disorder between the split datasets using spearman's correlation of log_2FC effect sizes for all brain-expressed genes (N=25,774 genes). We then repeated this analysis 1000 times and compared results across the range of SVs included. Addition of 4 SVs yielded the greatest cross-dataset replicability. B) Here, we plot the number of genes considered differentially expressed as a function of the number of SVs included in the differential expression model. C) DE results from this study are compared with published microarray datasets for each disorder (*1*) as a function of the number of SVs included. Spearman's correlation is shown for DGE log_2FC effect sizes for genes previously identified as DE (FDR<0.05) in the microarray dataset, as described in **Fig S4**. D) Multidimensional scaling plots are shown for the top 3 PC's of the covariate-corrected dataset, colored by study/batch and diagnosis.

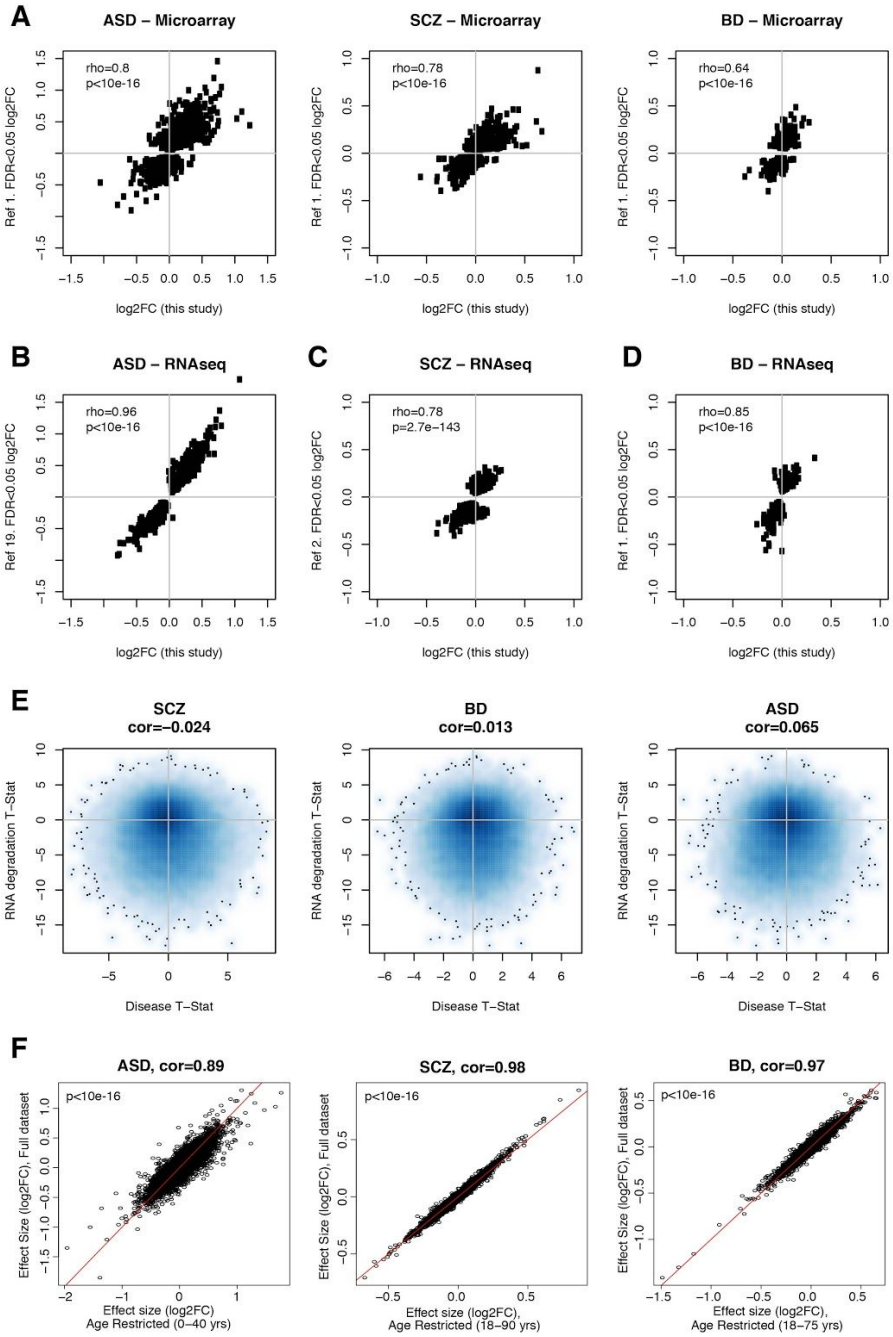
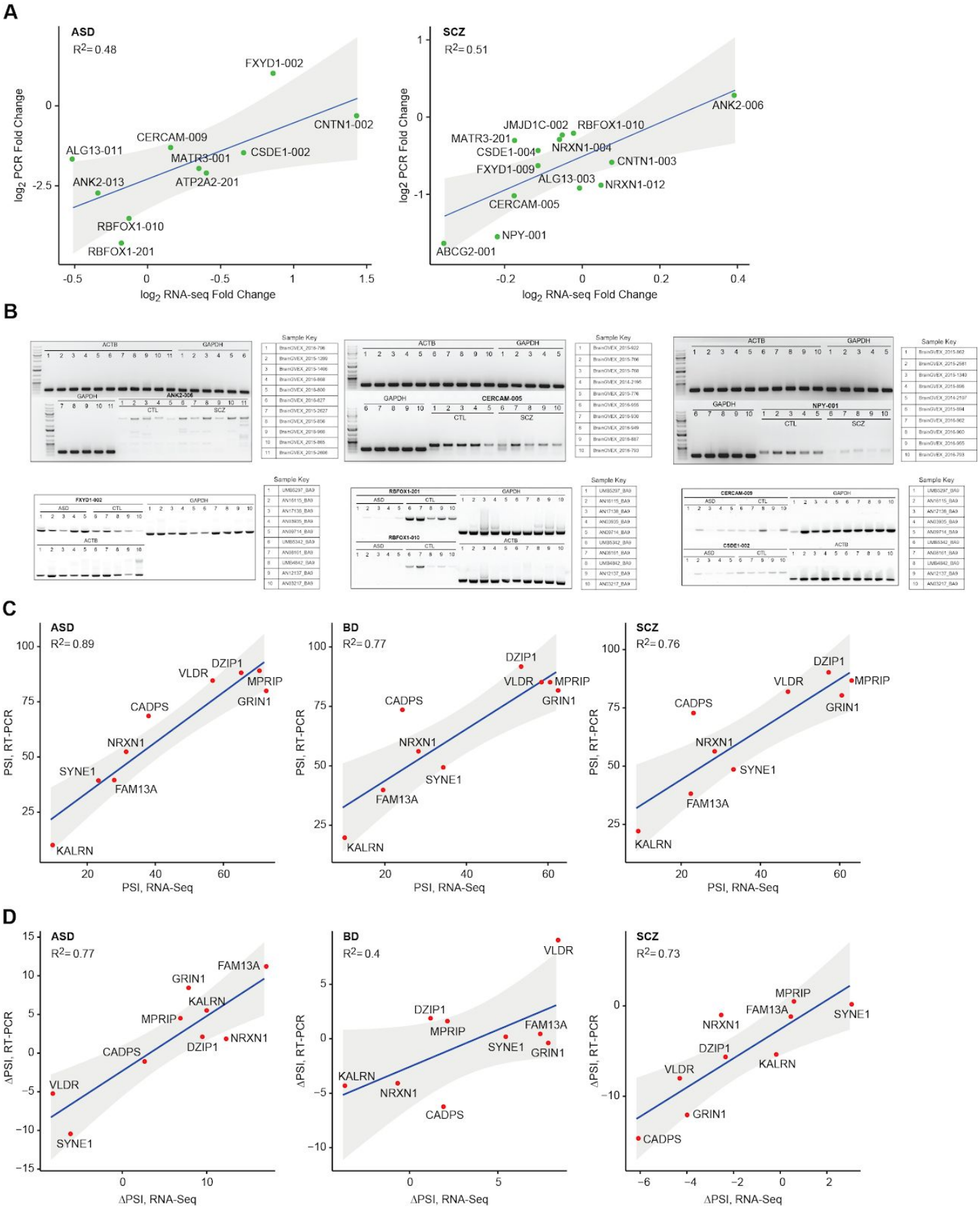


Figure S4 -- Validation of DGE results

Differential gene expression results from this study were compared with several published microarray and RNA-Seq datasets. A) Log_2FC effect sizes are plotted in comparison to a microarray meta-analysis of ASD, SCZ, and BD for genes identified as DGE (FDR<0.05) (*1*). We see substantial concordance of gene-level effect sizes across studies and platforms. Similar concordance is observed in comparison to results from RNA-Seq studies in B) ASD (*19*), C) SCZ (*2*), and D) BD (*1*). There is some overlap in samples across studies, due to the limited availability of post-mortem brain tissue from subjects with psychiatric disease. E) To ensure that differential gene expression in disease was not being driven by differences in RNA quality or degradation, we compared differential expression T-statistics with those experimentally derived from brain tissue samples allowed to degrade for fixed intervals of time (*22*). We did not observe substantial concordance between these RNA degradation metrics and psychiatric disease DGE summary statistics. F) Age balancing of case-control comparisons (0-40 years for ASD/CTL; 18-90 years for SCZ/CTL; 18-75 years for BD/CTL) does not substantially alter disease DGE signal.


A
ASD
R2= 0.48
FXYD1-002
CNTN1-002
CERCAM-009
ALG13-011
MATR3-001
CSDE1-002
ANK2-013
ATP2A2-201
RBFOX1-010
RBFOX1-201
log2 PCR Fold Change
log2 RNA-seq Fold Change
SCZ
R2= 0.51
ANK2-006
JMJD1C-002
RBFOX1-010
MATR3-201
NRXN1-004
CSDE1-004
CNTN1-003
FXYD1-009
ALG13-003
NRXN1-012
CERCAM-005
NPY-001
ABCG2-001
B
ACTB
GAPDH
ANK2-006
CERCAM-005
NPY-001
FXYD1-002
RBFOX1-201
RBFOX1-010
CSDE1-002
CTL
SCZ
ASD
Sample Key
C
ASD
R2= 0.89
BD
R2= 0.77
SCZ
R2= 0.76
DZIP1
VLDR
MPRIP
GRIN1
CADPS
NRXN1
SYNE1
FAM13A
KALRN
PSI, RT-PCR
PSI, RNA-Seq
D
ASD
R2= 0.77
BD
R2= 0.4
SCZ
R2= 0.73
FAM13A
GRIN1
KALRN
MPRIP
NRXN1
DZIP1
CADPS
VLDR
SYNE1
ΔPSI, RT-PCR
ΔPSI, RNA-Seq


E
ASD
BD
SCZ
CADPS
KALRN
FAM13A
NRXN1
VLDR
DZIP1
GRIN1
MPRIP
SYNE1
UMB5308_BA9
AN15451_BA9
UMB5278_BA9
AN16115_BA9
UMB5297_BA9
UMB5242_BA9
UMB5342_BA9
AN08161_BA9
UMB5079_BA9
UMB5168_BA9
BrainGVEX_2015-898
BrainGVEX_2015-854
BrainGVEX_2015-899
BrainGVEX_2015-860
BrainGVEX_2015-858
BrainGVEX_2015-246
BrainGVEX_2015-892
BrainGVEX_2015-895
BrainGVEX_2015-515
BrainGVEX_2015-857
BrainGVEX_2015-859
BrainGVEX_2016-834
BrainGVEX_2015-897
BrainGVEX_2016-791
BrainGVEX_2016-870
BrainGVEX_2015-775
BrainGVEX_2015-246
BrainGVEX_2015-893
BrainGVEX_2015-863
BrainGVEX_2015-737

Figure S5 -- Validation of differential transcript expression and differential splicing

A) Comparison of fold changes obtained from RSEM-based isoform quantification of RNA-Seq data (*21*) to semiquantitative PCR results for 10 isoforms tested in ASD and control samples (left) and 13 isoforms tested in SCZ and control samples (right). Fold changes were calculated between cases and control samples. B) Representative 1.5 to 2% agarose gel images obtained for isoform validation. C) Scatter plots comparing the average percent spliced-in (PSI) of exon-skipping events called by LeafCutter from RNA-Seq data (*21*) to semi-quantitative PCR. A total of 9 genes were tested in 5 cases and 5 controls in ASD, BD and SCZ. An additional 5 cases and 5 controls were tested for *FAM13A* and *SYNE1* in BD and SCZ to resolve outliers. D) Same as C, but now comparing the change in average PSI (∆PSI) between cases and controls in each disorder. E) Representative Agilent 2100 Bioanalyzer gel images (DNA 1000 chips) obtained for splicing validation. A-E) Gene or isoform names are indicated at each point. Regression lines with 95% confidence intervals are shown in blue and grey, respectively and the corresponding R^2 values are shown at the top-left in each plot.

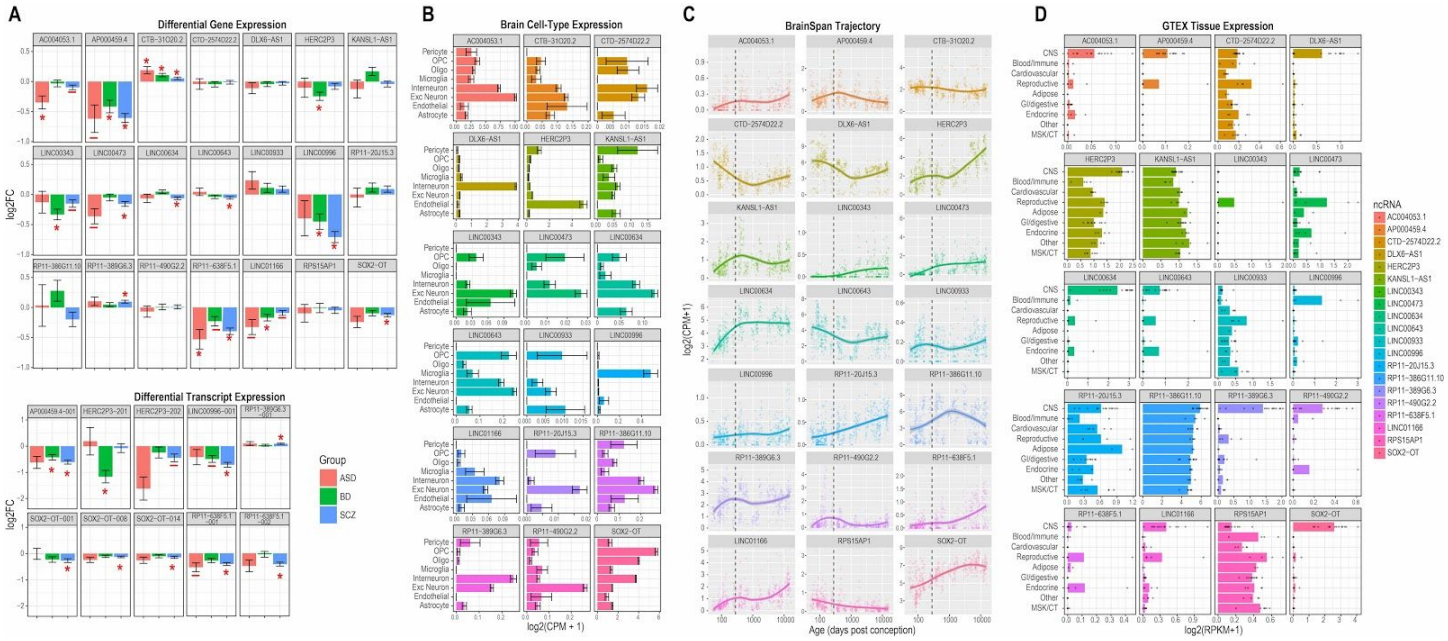


Figure S6 -- Annotation of individual ncRNAs

We highlight several individual ncRNAs differentially expressed in psychiatric disease or identified as hubs of relevant co-expression modules. A) We show differential gene expression (DGE; top) and differential transcript expression (DTE; bottom) in SCZ, BD, and ASD. *FDR<0.05, -- FDR<0.1. B) Human brain cell type expression patterns are shown for each ncRNA using data from Ref (*91*). Plots show mean expression for cells identified in specific clusters. C) Developmental expression trajectory is shown for each ncRNA using data from BrainSpan (*144*). Plots show expression as a function of age (days post-conception) on a $\log_{10}$ scale, with the dotted line denoting birth. D) Human tissue-specific expression levels are shown for each ncRNA using data from GTEX (*81*).

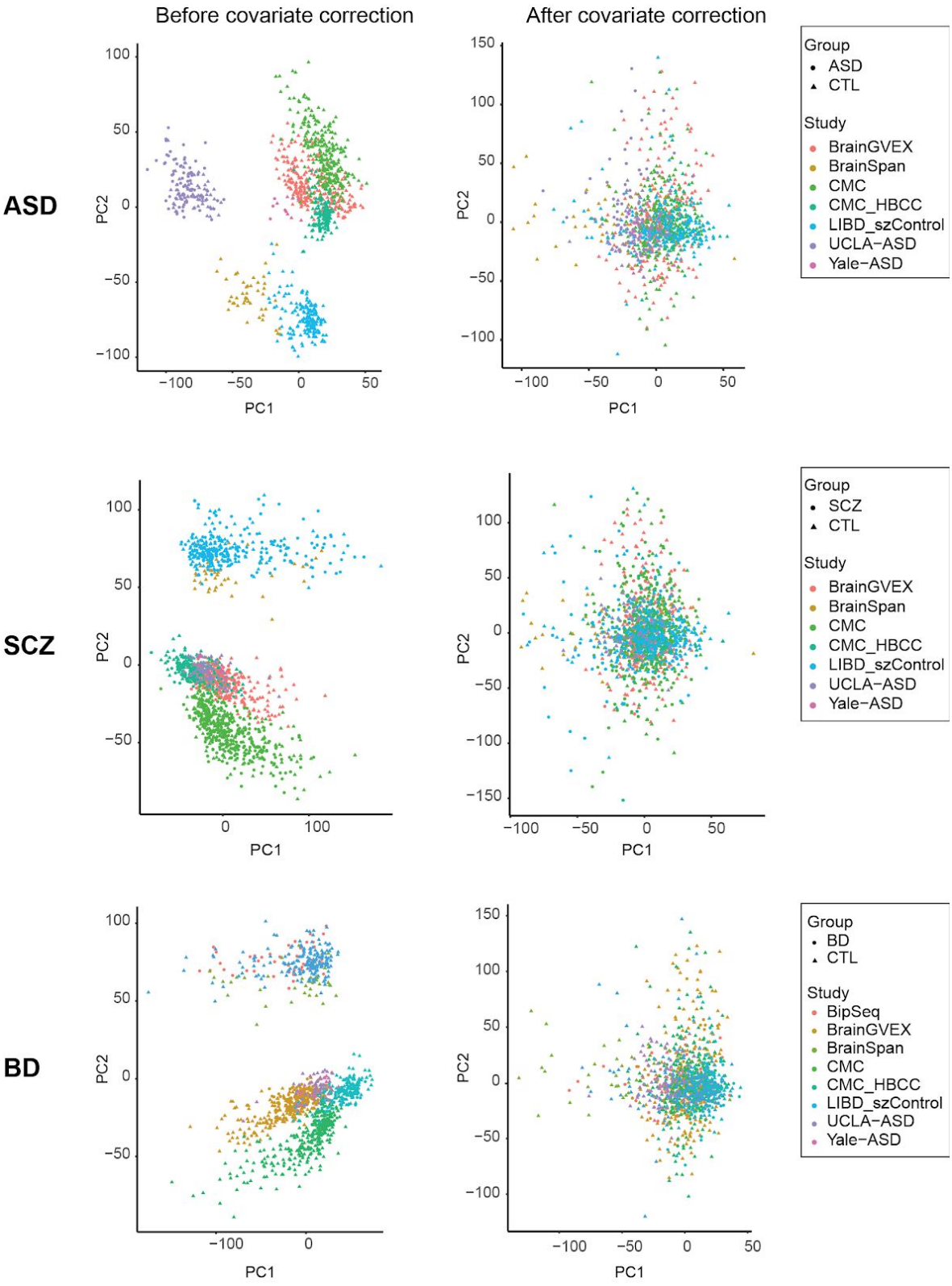


Figure S7 -- Covariate correction of DS

Top two principal components (PCs) of percent spliced-in (PSI) values are shown before (left) and after (right) covariate correction for the ASD, BD and SCZ datasets. Points are colored according to study origin and shape denotes disorder (circle) or control (CTL) status (triangle). See inset legends for further details.

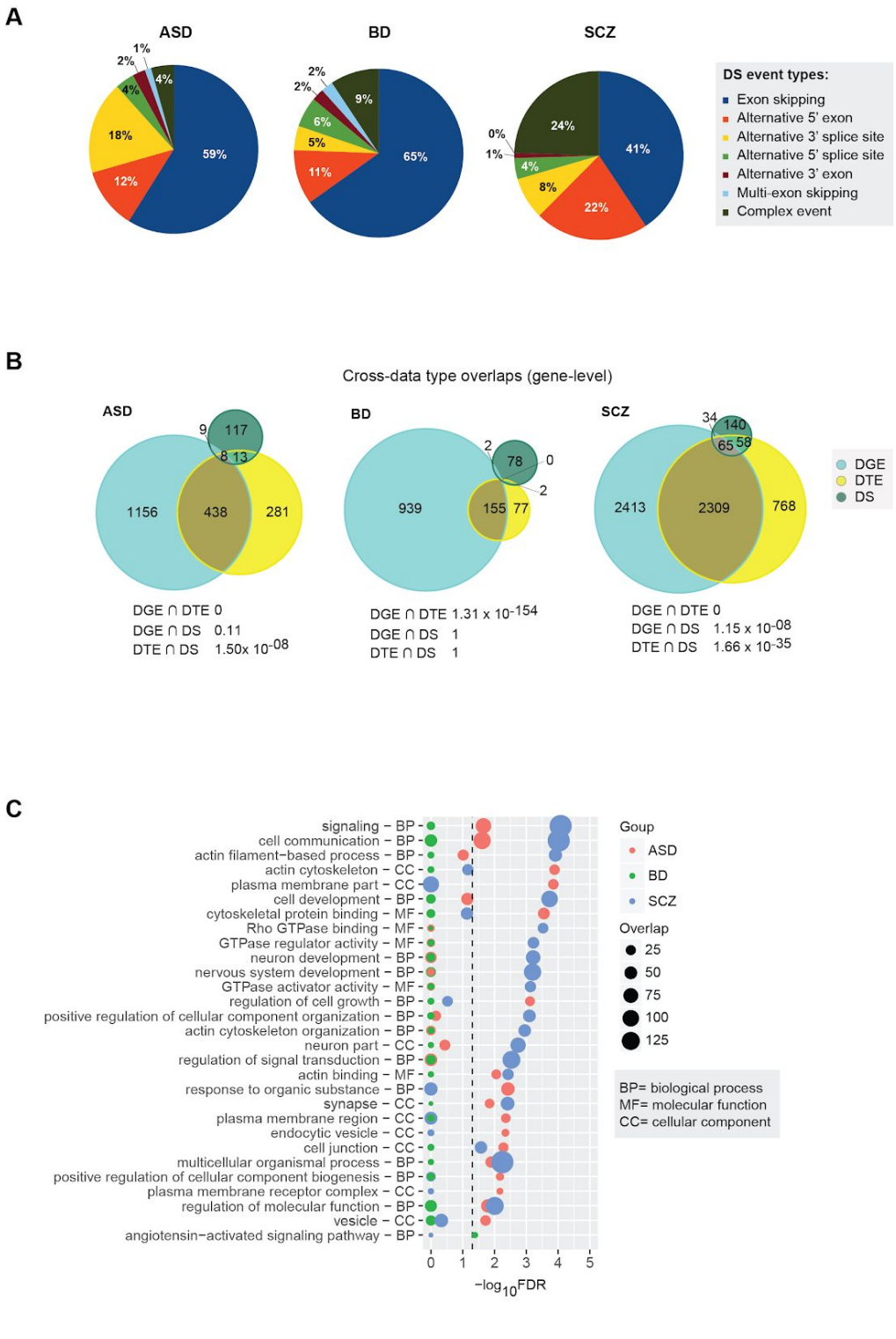


Figure S8 -- Annotation of DS events, Cross data DGE-DTE-DS overlaps.

A) Pie charts with breakdown of DS event types identified in each disorder. B) Venn diagrams showing overlaps between genes with significant DGE, DTE or DS changes for each disorder. P values for hypergeometric tests of pairwise overlaps between data types are shown at the bottom of the venn diagrams for each disorder. C) Top 20 gene ontology (GO) enrichments for DS genes in each disorder.

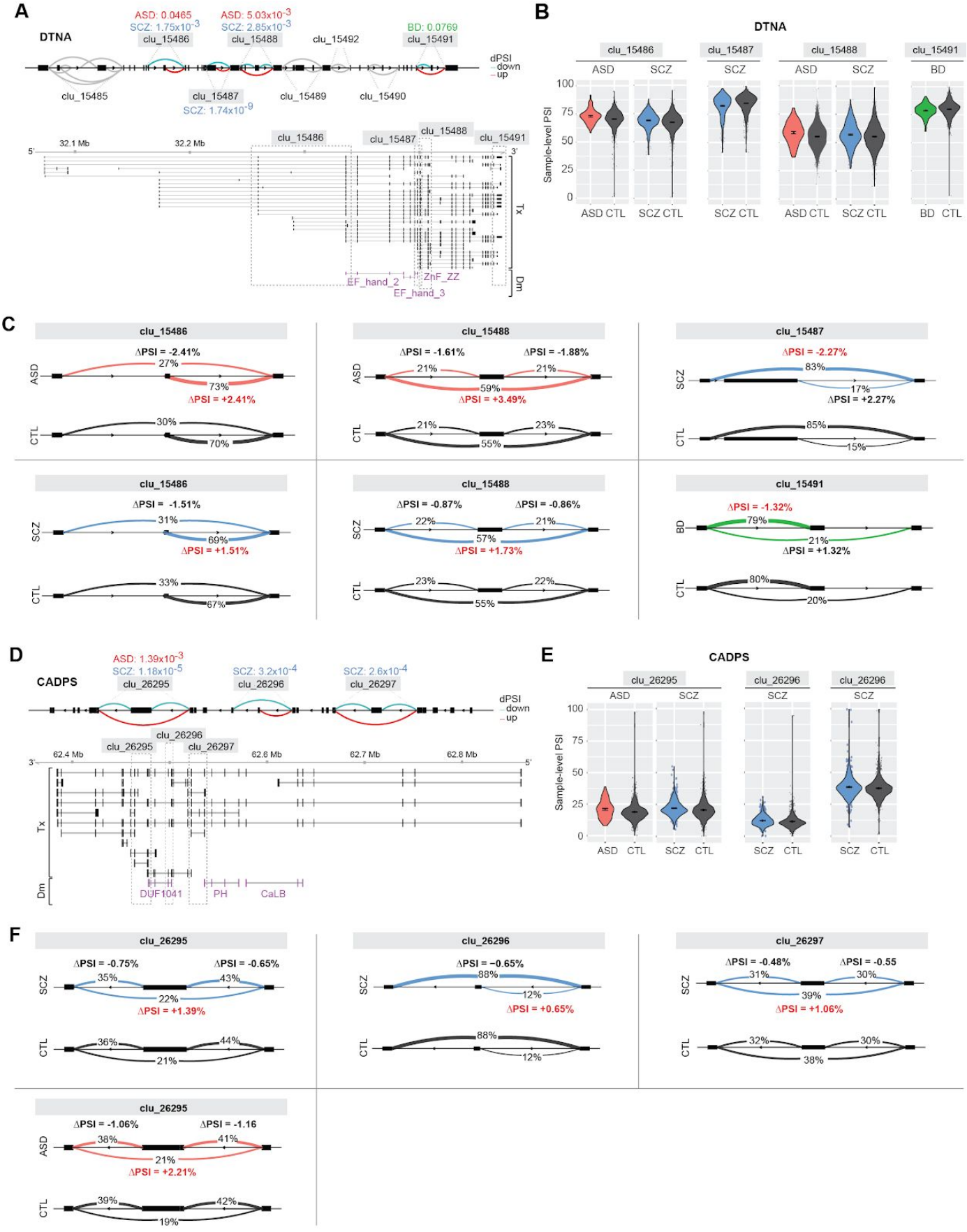

A
DTNA
ASD: 0.0465
SCZ: 1.75x10-3
clu_15486
ASD: 5.03x10-3
SCZ: 2.85x10-3
clu_15488
clu_15492
BD: 0.0769
clu_15491
dPSI
down
up
clu_15485
clu_15487
SCZ: 1.74x10-9
clu_15489
clu_15490
32.1 Mb
32.2 Mb
5'
3'
Tx
Dm
EF_hand_2
EF_hand_3
ZnF_ZZ
B
DTNA
Sample-level PSI
ASD CTL
SCZ CTL
BD CTL
C
ΔPSI = -2.41%
ΔPSI = +2.41%
ΔPSI = -1.61%
ΔPSI = -1.88%
ΔPSI = +3.49%
ΔPSI = -2.27%
ΔPSI = +2.27%
ΔPSI = -1.51%
ΔPSI = +1.51%
ΔPSI = -0.87%
ΔPSI = -0.86%
ΔPSI = +1.73%
ΔPSI = -1.32%
ΔPSI = +1.32%
D
CADPS
ASD: 1.39x10-3
SCZ: 1.18x10-5
clu_26295
SCZ: 3.2x10-4
clu_26296
SCZ: 2.6x10-4
clu_26297
62.4 Mb
62.6 Mb
62.7 Mb
62.8 Mb
DUF1041
PH
CaLB
E
CADPS
F
ΔPSI = -0.75%
ΔPSI = -0.65%
ΔPSI = +1.39%
ΔPSI = −0.65%
ΔPSI = +0.65%
ΔPSI = -0.48%
ΔPSI = -0.55
ΔPSI = +1.06%
ΔPSI = -1.06%
ΔPSI = -1.16
ΔPSI = +2.21%

Figure S9 -- Additional differential local splicing examples

A) Top: Significant differentially spliced (DS) intron clusters in *DTNA* for ASD, SCZ and BD. Increased or decreased intron usage in cases compared to controls (CTL) are shown in red and blue, respectively. Bottom: Overview of known isoforms (GENCODE v19) and protein domains for *DTNA*. Locations of significant DS clusters are indicated by dotted lines. Protein domains (purple) are annotationed as EF_hand_2 - EF hand domain 2; EF_hand_3 - EF hand domain 3; ZnF_ZZ - Zinc-binding domain, present in Dystrophin, CREB-binding protein. B) Violin-plots with the distribution of covariate-adjusted percent spliced in (PSI) *per* sample for the intron with the maximum change in PSI for each cluster and disorder. C) Visualization of introns in each significant cluster for each disorder, with their change in PSI (PSI). Covariate-adjusted average PSI levels in disorder *vs* CTL are indicated for each intron. D) Same as A), but for *CADPS*. Protein domains (purple) are annotated as PH - Pleckstrin homology domain; CaLB - C2 domain (Calcium/lipid-binding domain, CaLB) superfamily; DUF1041 - Domain of unknown function. E) Same as B), but for *CADPS*. F) Same as C, but for *CADPS*.

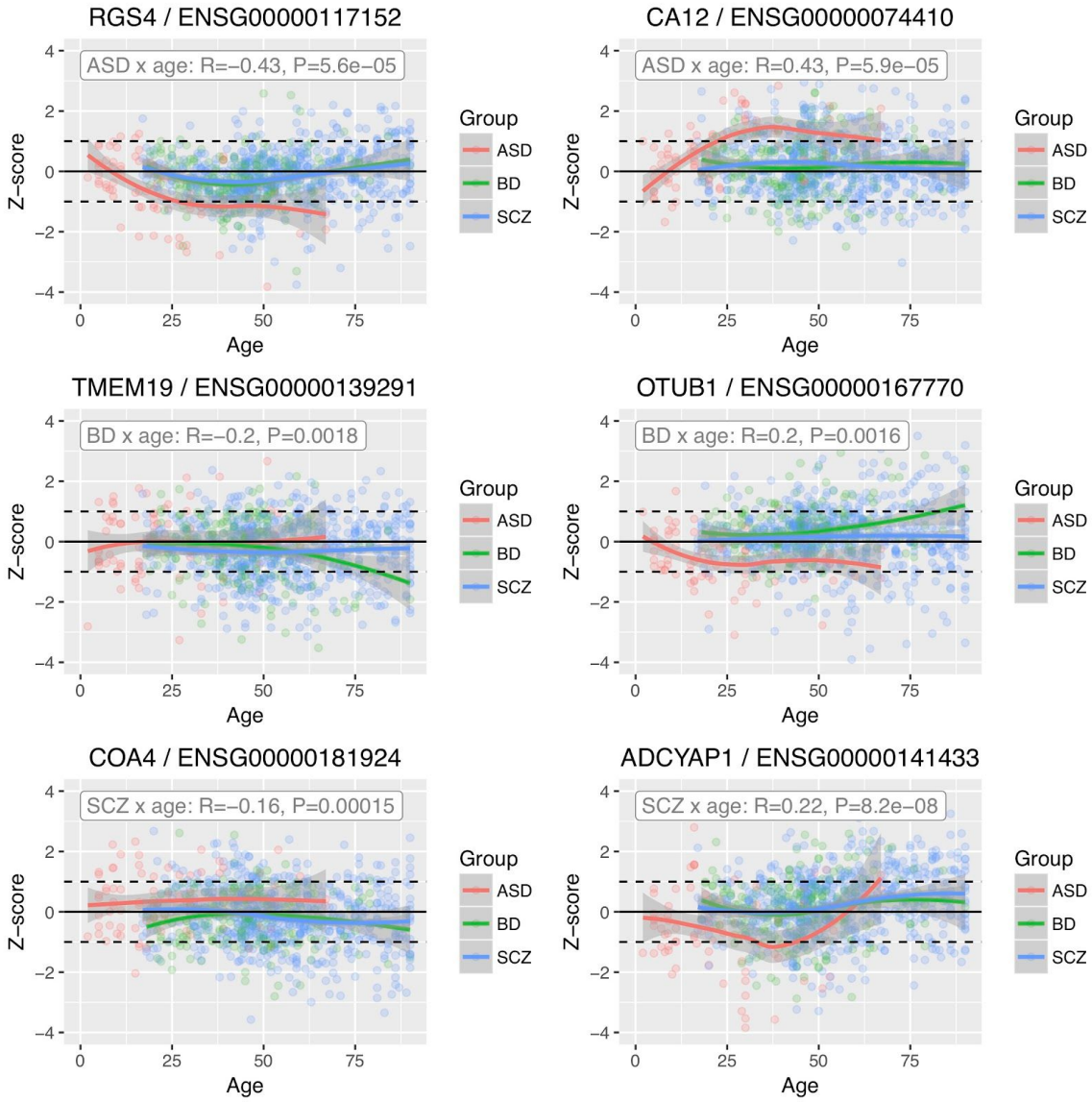


Figure S10 - Age effects on differential gene expression

Examples are shown for genes whose magnitude of differential expression in a given disorder is significantly associated with age. For each gene, a local regression function was fit to model the effect of age on expression in control samples, and expression in cases was then converted to a z-score relative to the local mean in controls. We then assessed the correlation between z-transformed expression and age within each disease group (ASD, SCZ, BD) separately, to identify those genes whose magnitude of differential expression was associated with age. We find that 143 of the 4821 DGE genes in SCZ show a nominal increase in effect size magnitude as a function of age, consistent with a reactive interpretation. In ASD, 85 of 1611 DE genes showed this same pattern and in BD there were 29 of the 1119 DE genes.

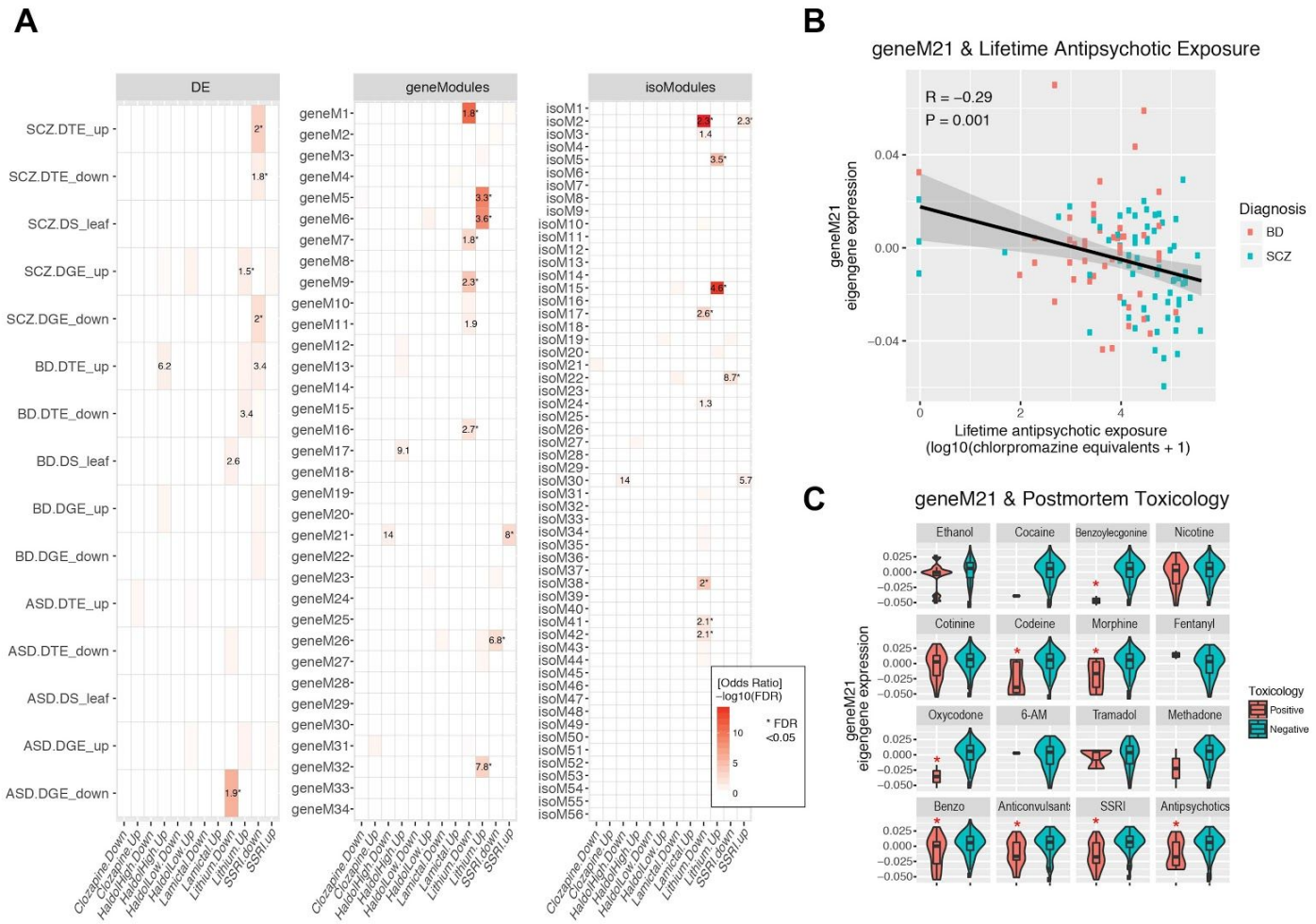


Figure S11 -- Assessment of psychiatric medication effects

A) We investigated whether antipsychotic medications could explain differential gene expression and module associations identified in SCZ, BD, and ASD. We used three experimental datasets: (1) an RNA-Seq dataset from DLPFC of nonhuman primates exposed for 6 months to clozapine, haloperidol (low dose), or haloperidol (high dose) compared to placebo; (2) a microarray dataset from mouse brain following chronic exposure to the SSRI fluoxetine; (3) a microarray dataset from rat cortex following chronic exposure to the mood stabilizers lithium or lamotrigine compared with vehicle (*21*). Overlap of DE genes and modules with genes up or downregulated by medications (at nominal significance thresholds, except for lithium) was assessed by Fisher's exact test. Plot shows odds-ratios of enrichment for P<0.05 significant associations, with * denoting FDR<0.05 associations. With the exception of lithium, medications show minimal overlap with disease-associated transcriptomic changes. The one exception was for the activity dependent module pair, geneM21/isoM30, which did seem to be associated with SSRIs and high dose haloperidol. B) To investigate this relationship further, we compared geneM21 eigengene expression with medication history in those subjects where this information was available. There was a significant negative correlation between geneM21 expression and lifetime antipsychotic exposure (chlorpromazine equivalents, log scale). C) A subset of samples also had results from post-mortem toxicology testing. We found broadly decreased levels of geneM21 eigengene expression in those subjects who tested positive for a host of psychiatric medications, including antipsychotics (*FDR<0.05).

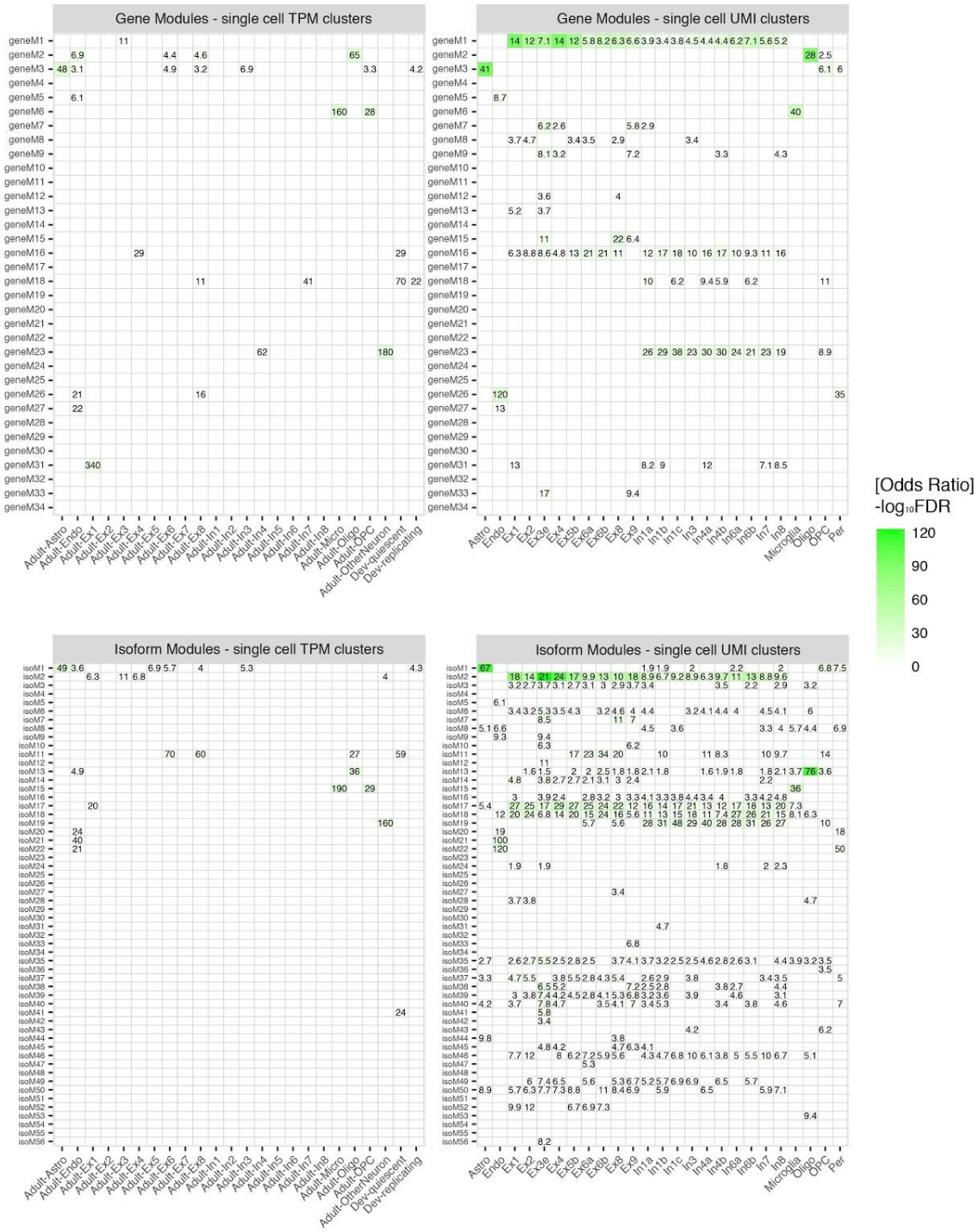


Figure S12 -- Co-expression network cell type enrichments

Plots show enrichment of gene and isoform-level co-expression modules for established markers of CNS cell types from human brain single-cell RNA-Seq clusters, as compiled in the companion manuscript (*18*). Clusters were defined separately for TPM- and UMI- based scRNA-Seq quantifications. Text denotes odds ratios of enrichment for significant associations (FDR<0.05). The UMI dataset is from adult human brain, whereas the TPM dataset includes two fetal cell types. (Ex# - excitatory neuron cluster; In# - interneuron cluster; Per- pericyte; OPC-oligodendrocyte progenitor cell; Oligo - oligodendrocyte; Micro - microglia; End - endothelial; Ast - astrocyte).

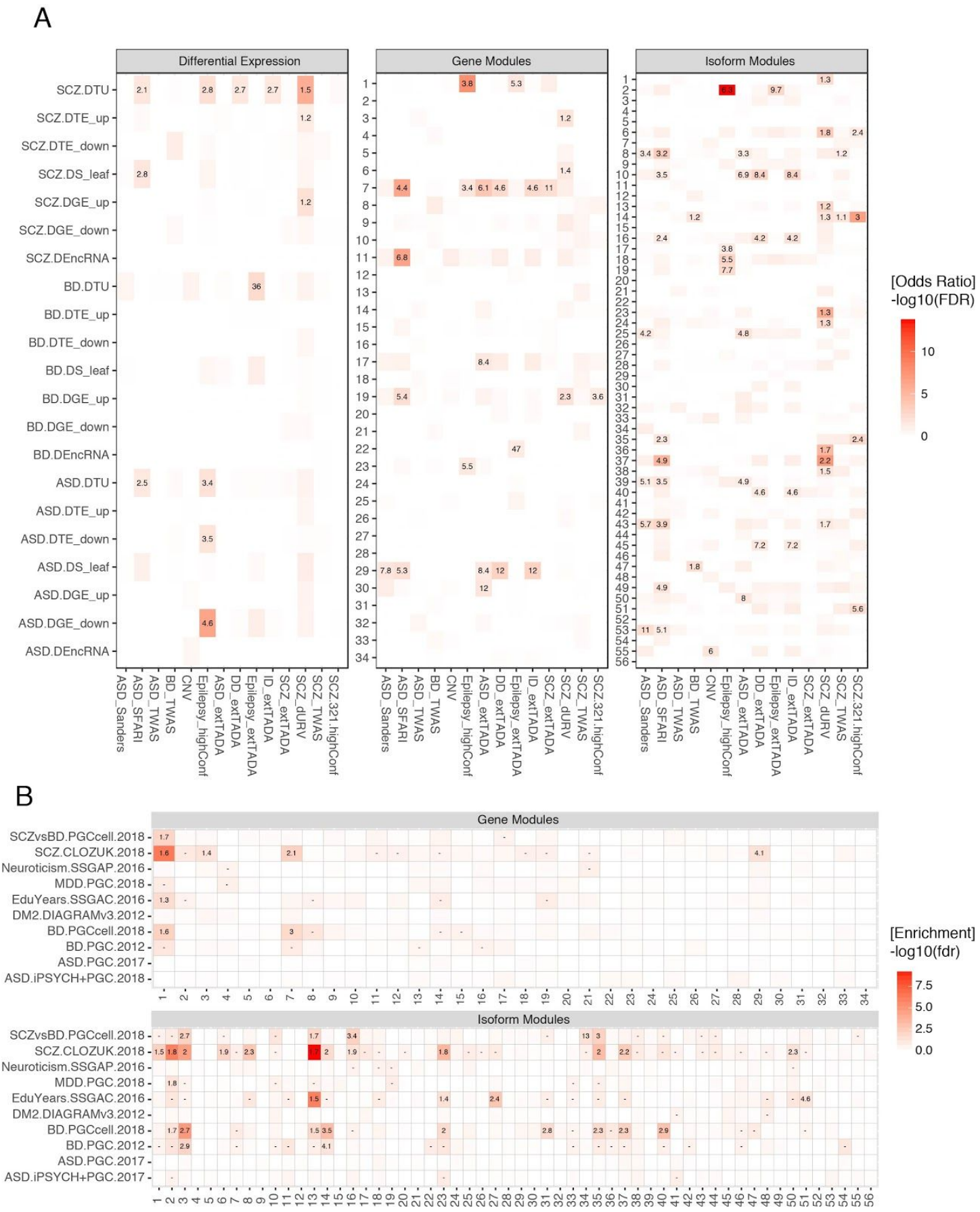

A
Differential Expression
Gene Modules
Isoform Modules
SCZ.DTU
SCZ.DTE_up
SCZ.DTE_down
SCZ.DS_leaf
SCZ.DGE_up
SCZ.DGE_down
SCZ.DEncRNA
BD.DTU
BD.DTE_up
BD.DTE_down
BD.DS_leaf
BD.DGE_up
BD.DGE_down
BD.DEncRNA
ASD.DTU
ASD.DTE_up
ASD.DTE_down
ASD.DS_leaf
ASD.DGE_up
ASD.DGE_down
ASD.DEncRNA
ASD_Sanders
ASD_SFARI
ASD_TWAS
BD_TWAS
CNV
Epilepsy_highConf
ASD_extTADA
DD_extTADA
Epilepsy_extTADA
ID_extTADA
SCZ_extTADA
SCZ_dURV
SCZ_TWAS
SCZ.321.highConf
[Odds Ratio]
-log10(FDR)
10
5
0
B
Gene Modules
Isoform Modules
SCZvsBD.PGCcell.2018
SCZ.CLOZUK.2018
Neuroticism.SSGAP.2016
MDD.PGC.2018
EduYears.SSGAC.2016
DM2.DIAGRAMv3.2012
BD.PGCcell.2018
BD.PGC.2012
ASD.PGC.2017
ASD.iPSYCH+PGC.2018
[Enrichment]
-log10(fdr)
7.5
5.0
2.5
0.0

Figure S13 -- Genetic enrichment analyses

A) Enrichment for several sets of disease risk genes was assessed among DE features, gene, and isoform-coexpression modules, including those harboring rare *de novo* variants identified in each disorder, as well as in related neurodevelopmental and psychiatric traits. TWAS signal for each disorder was also included as was the list of 321 “high confidence” SCZ risk genes identified in the companion manuscript (*18*). Enrichment was calculated using logistic regression, controlling for gene and transcript length as well as GC content (*21*). Risk gene sets include: 71 risk loci harboring rare *de novo* variants associated with ASD through the transmission and *de novo* association test (TADA; “ASD_Sanders”) (*80*); Syndromic and highly ranked (1 and 2) genes from SFARI Gene database (“ASD_SFARI”); genes harboring recurrent *de novo* copy-number variants associated with ASD or SCZ, as defined in (*1*) (“CNV”); genes harboring an excess of rare exonic variants in ASD, SCZ, intellectual disability (ID), developmental delay (DD), and epilepsy as assessment through an extended version of TADA (“extTADA”) (*140*); genes harboring disruptive and damaging ultra-rare variants in SCZ (*54*) (“SCZ_dURVs”); a list of high confidence epilepsy risk genes, compiled from (*141*). B) Enrichment of GWAS signal among gene and isoform co-expression modules, using stratified LD score regression (s-LDSR) with summary statistics from several psychiatric, cognitive, and behavioral traits (*21*). Cells are labeled with GWAS enrichment, for those with FDR < 0.05. Cells labeled with “-” are nominally (P<0.05) significant but do not pass FDR-correction.


geneM1
geneM2
geneM3
geneM4
geneM5
geneM6
geneM7
geneM8
geneM9
geneM10
geneM11
geneM12
geneM13
geneM14
geneM15
geneM16
geneM17
geneM18
geneM19
geneM20
geneM21
geneM22
geneM23
geneM24
geneM25
geneM26
geneM27
geneM28
geneM29
geneM30
geneM31
geneM32
geneM33
geneM34
isoM1
isoM2
isoM3
isoM4
isoM5
isoM6
isoM7
isoM8
isoM9
isoM10
isoM11
isoM12
isoM13
isoM14
isoM15
isoM16
isoM17
isoM18
isoM19
isoM20
isoM21
isoM22
isoM23
isoM24
isoM25
isoM26
isoM27
isoM28
isoM29
isoM30
isoM31
isoM32
isoM33
isoM34
isoM35
isoM36
isoM37
isoM38
isoM39
isoM40
isoM41
isoM42
isoM43
isoM44
isoM45
isoM46
isoM47
isoM48
isoM49
isoM50
isoM51
isoM52
isoM53
isoM54
isoM55
isoM56
0.02
0.01
0.00
-0.01
-0.02
beta
Group
ASD
BD
SCZ
SCZ subsample (BD matched)
SCZ subsample (ASD matched)

Figure S14 -- Module-trait associations after SCZ downsampling

To determine whether differences in module associations observed across disorders was due to the larger sample size of the SCZ dataset, we repeated our module-trait association analyses using a randomly subsampled SCZ dataset to match the sample size of ASD and BD datasets. We repeated this 100 times and reran our module-level associations using these matched sample sizes. Plots show module-trait association β values with standard errors. *P<0.05.

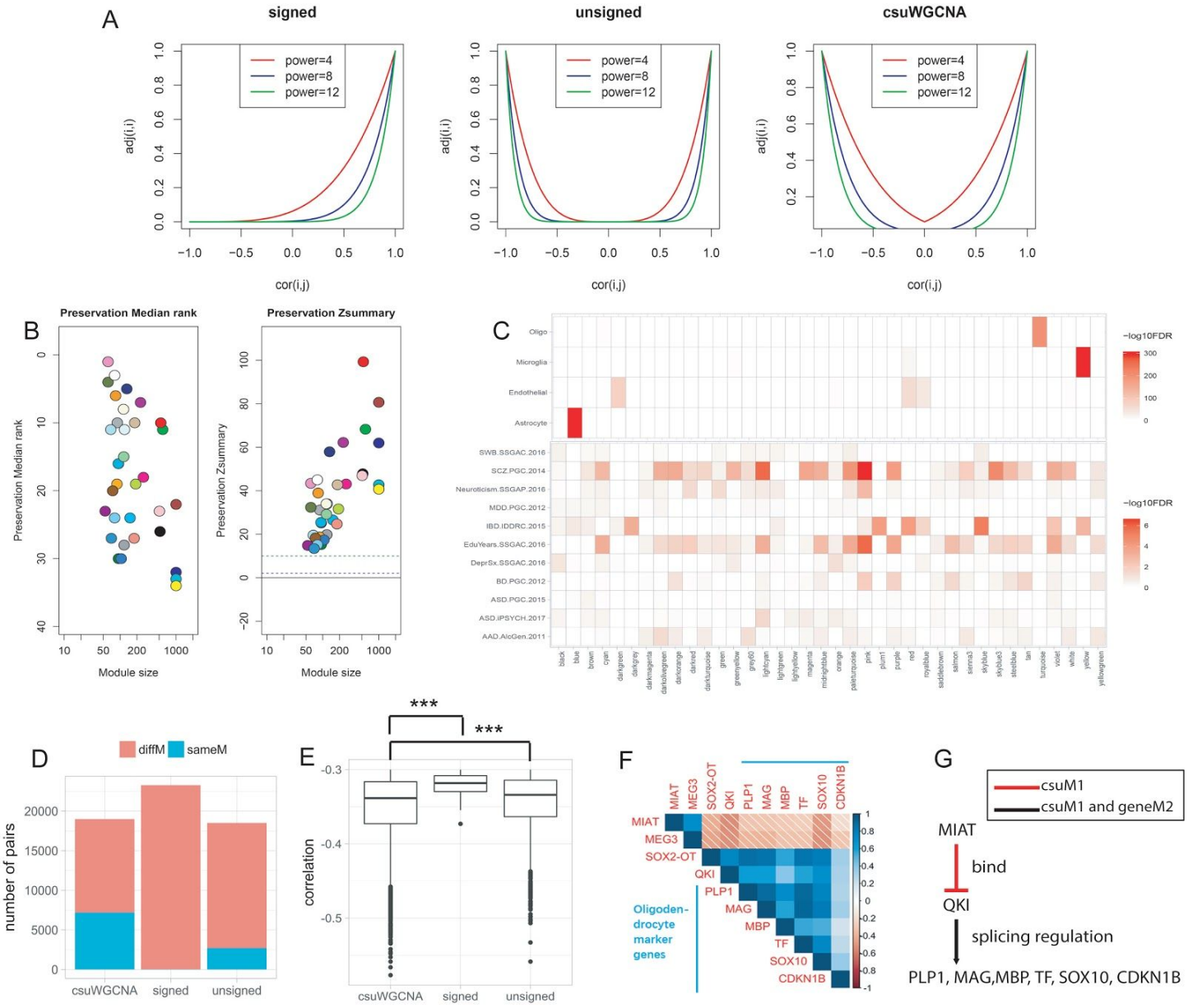


Figure S15 -- csuWGCNA identifies putative lncRNA negative regulatory relationships

A) Network adjacency (y-axis) versus correlation (x-axis) in the signed network, unsigned network, and csuWGCNA network. The color of the line denotes the soft threshold power setting. Note that correlation=-1 leads to adjacency = 0 in the signed network and adjacency =1 in the unsigned and csuWGCNA network. B) All modules detected by csuWGCNA were well preserved in the signed networks (Zsummary > 10 indicates high preservation). C) The enrichment of cell type and GWAS signal in csuWGCNA modules. D) csuWGCNA captures more negative lncRNA-gene pairs (cor<-0.3) in the same module than signed and unsigned WGCNA (csuWGCNA=7186, signed=20, unsigned=2701). E) csuWGCNA captures stronger negative relationships than signed and unsigned network types (Welch two sample t-test, $p<10^{-6}$ and $p<10^{-11}$, respectively). F) The lncRNAs *MIAT* and *MEG3* are negatively correlated with most of the hubs in oligodendrocyte modules (show in **Fig 4H**), including *SOX2-OT* and oligodendrocyte marker genes (*PLP1*, *MAG*, *MBP*, *TF*, *SOX10*, and *CDKN1B*). The blue color indicates negative correlations and the red indicates positive correlations. G) Putative target relationships for the lncRNA *MIAT*. The red line indicates a negative relationship only detected in csuM1, and the black line indicates positive relationships detected in both csuM1 and geneM2.

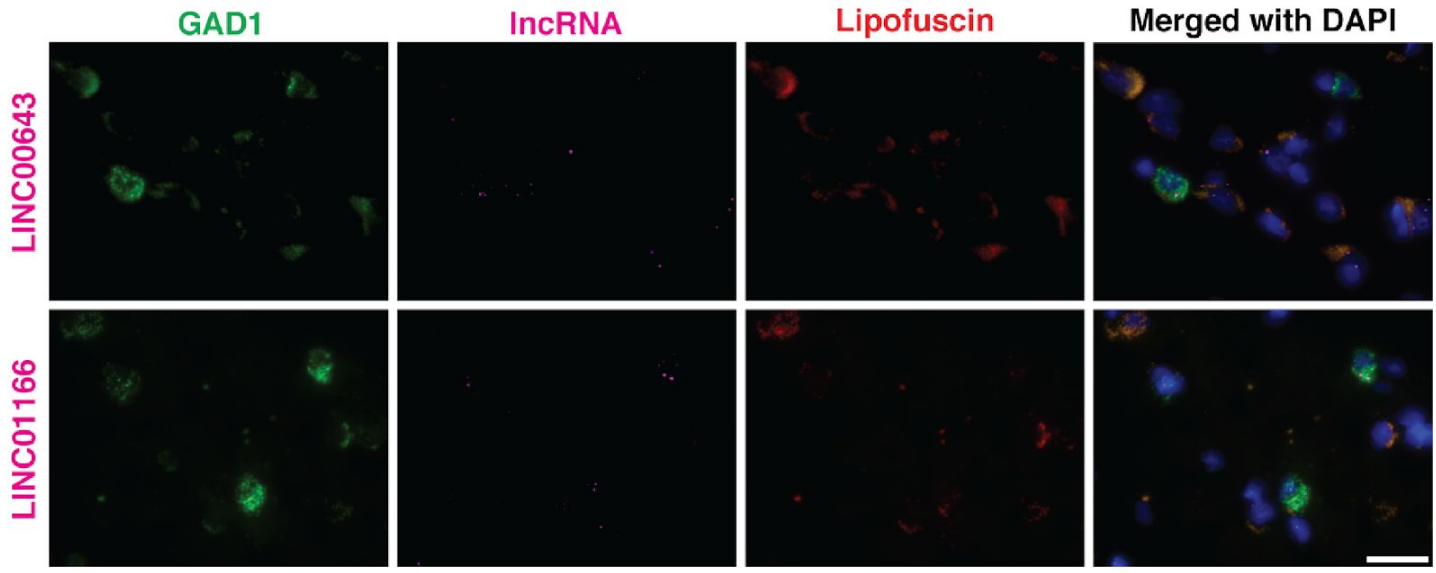


Figure S16 -- *LINC00643* and *LINC01166* expression in human prefrontal cortex

Sections from human prefrontal cortex (area 9) were labeled with *GAD1* probe (green) and lncRNA (magenta) probes for *LINC00643* (upper panel) or for *LINC01166* (lower panel). All sections were counterstained with DAPI (blue) to reveal cell nuclei. Lipofuscin autofluorescence is visible in both the green and red channels and appears yellow/orange in the merged image. The lncRNAs are present both in GABAergic interneurons and cells without *GAD1* signal. Scale bar, 25 µm.

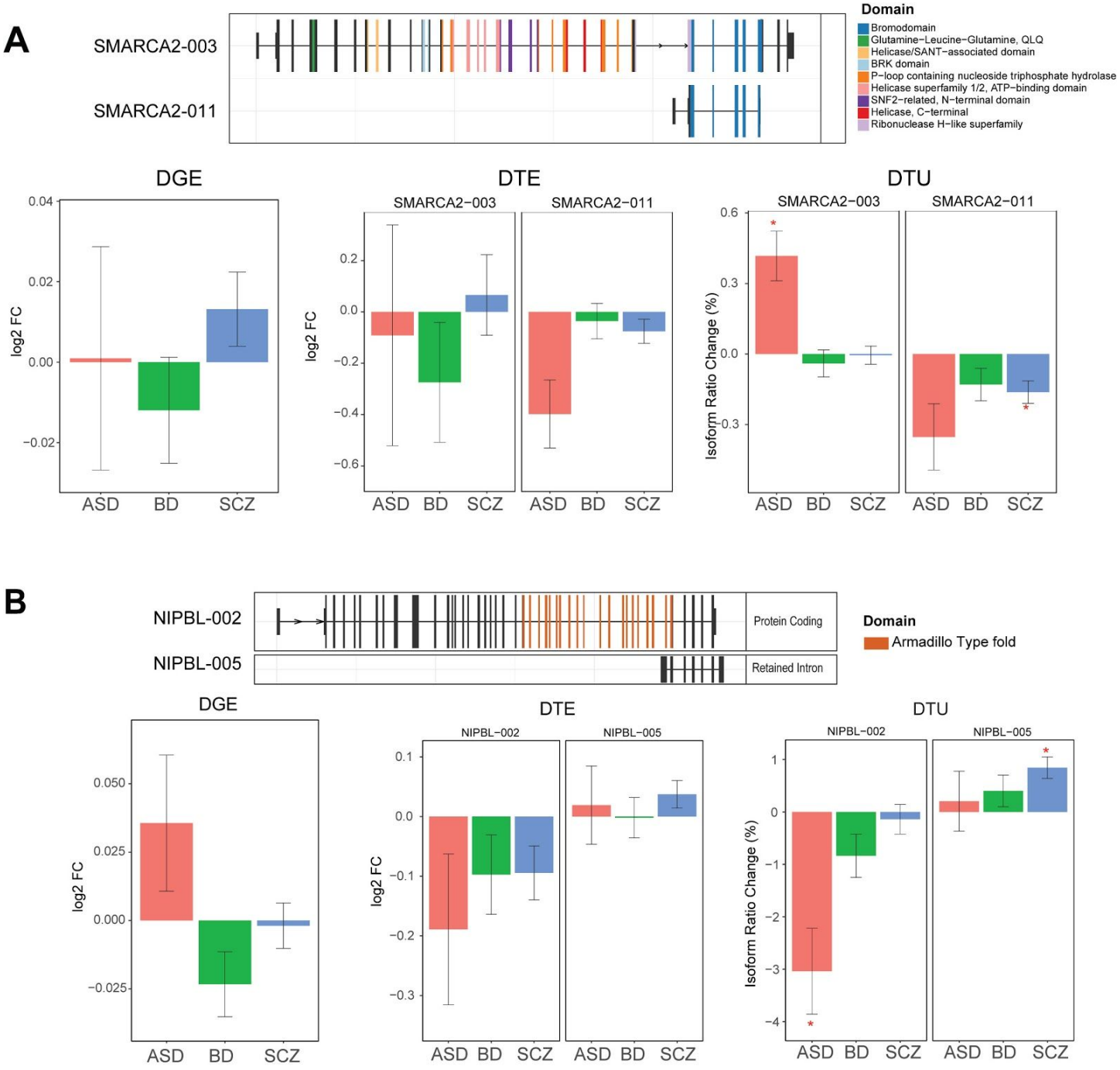


Figure S17 -- Additional switch isoforms

A) The isoform ratio of two *SMARCA2* isoforms, *SMARCA2-003* and *SMARCA2-011*, are up and downregulated in ASD and SCZ, respectively. B) The isoform ratio of two *NIPBL* isoforms, *NIPBL-002* and *NIPBL-005*, are down and upregulated in ASD and SCZ, respectively. *FDR < 0.05

Tables S1 to S9

Table S1 -- Differential gene and isoform expression summary statistics and DE enrichment analyses

Table S2 -- Annotation of psychiatric ncRNAs

Table S3 -- Differential splicing summary statistics, annotation and disease overlaps

Table S4 -- TWAS summary statistics and PRS associations with gene and isoform expression

Table S5 -- Gene and isoform co-expression module annotation

Table S6 -- csuWGCNA network annotation and putative lncRNA-mRNA targets

Table S7 -- Switch isoform and microexon characterization

Table S8 -- Splicing and isoform validation primers and samples

Table S9 -- RNAscope - Tissue samples and RNA FISH probes