



Published in final edited form as:

Mol Psychiatry. 2015 May ; 20(5): 563–572. doi:10.1038/mp.2014.82.

Transcriptome Sequencing and Genome-wide Association Analyses Reveal Lysosomal Function and Actin Cytoskeleton Remodeling in Schizophrenia and Bipolar Disorder

Zhongming Zhao^{#1}, Jiabao Xu^{#2}, Jingchun Chen^{#3}, Sanghyeon Kim⁴, Mark Reimers³, Silviu-Alin Bacanu³, Hui Yu¹, Chunyu Liu⁵, Jingchun Sun¹, Quan Wang¹, Peilin Jia¹, Fengping Xu², Yong Zhang², Kenneth S. Kendler³, Zhiyu Peng², and Xiangning Chen^{3,*}

¹ Departments of Biomedical Informatics and Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

² Beijing Genomics Institute (BGI), Shenzhen, Guangdong, 518083, China

³ Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23298, USA

⁴ Stanley Laboratory of Brain Research, 9800 Medical Center Drive, Rockville, MD 20850, USA.

⁵ Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60637, USA

These authors contributed equally to this work.

Abstract

Schizophrenia (SCZ) and bipolar disorder (BPD) are severe mental disorders with high heritability. Clinicians have long noticed the similarities of clinic symptoms between these disorders. In recent years, accumulating evidence indicates some shared genetic liabilities. However, what is shared remains elusive. In this study, we conducted whole transcriptome analysis of postmortem brain tissues (cingulate cortex) from SCZ, BPD and control subjects, and identified differentially expressed genes in these disorders. We found 105 and 153 genes differentially expressed in SCZ and BPD, respectively. By comparing the t-test scores, we found that many of the genes differentially expressed in SCZ and BPD are concordant in their expression level ($q < 0.01$, 53 genes; $q < 0.05$, 213 genes; $q < 0.1$, 885 genes). Using genome-wide association data from the Psychiatric Genomics Consortium, we found that these differentially and concordantly expressed genes were enriched in association signals for both SCZ ($p < 10^{-7}$) and BPD ($p = 0.029$). To our knowledge, this is the first time that a substantially large number of genes shows concordant expression and association for both SCZ and BPD. Pathway analyses of these genes indicated that they are involved in the lysosome, Fc gamma receptor mediated phagocytosis, regulation of actin skeleton pathways, along with several cancer pathways. Functional analyses of these genes revealed an interconnected pathway network centered on lysosomal function and the regulation of actin cytoskeleton. These pathways and their interacting

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

* Corresponding author: tel. 804 828 8124, fax 804 628 1035, xchen@vcu.edu.

The authors have declared there is **NO** conflict of interest to disclose

network were principally confirmed by an independent transcriptome sequencing dataset of hippocampus. Dysregulation of lysosomal function and cytoskeleton remodeling has direct impacts on endocytosis, phagocytosis, exocytosis, vesicle trafficking, neuronal maturation and migration, neurite outgrowth, and synaptic density and plasticity, and different aspects of these processes have been implicated in SCZ and BPD.

2 Introduction

It is widely known amongst psychiatrists that schizophrenia (SCZ) and bipolar disorder (BPD) share some clinic presentations. Psychotic symptoms such as delusion and hallucination are commonly seen in patients of SCZ and BPD. Additionally, both disorders share affective deficits (manic and depressive symptoms). Family studies have shown that both disorders have relatively high heritability, and genetic factors play a critical role in the manifestation of the disorders. It is also known that environmental factors have significant impacts on the development of the disorders. In recent years, there are many reports that the two disorders share some risk genes¹⁻³. More recently, polygenic analyses have demonstrated that the two disorders share aggregated genetic liability across the genome⁴. However, the extent and identity of these shared genetic risks remain largely unknown.

Rapid advance in next generation sequencing technologies has made it feasible to conduct whole transcriptome (i.e., RNA-Seq) analysis of a large number of samples. In this study, we applied comparative RNA sequencing to postmortem brain tissues from 31 SCZ patients, 25 BPD patients, and 26 healthy controls. We performed expression analyses to identify differentially expressed genes for the two disorders, and examined if these differentially expressed genes for SCZ and BPD were correlated. Based on the correlation of these differentially expressed genes, we tested whether these genes were enriched in association signals using data from the Psychiatric Genomics Consortium (PGC) stage I genome wide association studies (GWAS) of both SCZ and BPD (<https://www.nimhgenetics.org/>). In these analyses, we found evidence that genes differentially and concordantly expressed between SCZ and BPD were enriched in association signals for both diseases. We conducted further analyses to investigate these concordantly and differentially expressed genes for enrichment of rare variants and biological pathways. From these analyses, we obtained independent evidence that SCZ and BPD shared multiple pathways in genetic liability.

3 Materials and Methods

3.1 RNA sequencing

In this study, we used the array collection samples from Stanley Medical Research Institute (SMRI, <http://www.stanleyresearch.org/dnn/>) for transcriptome sequencing. The array collection, consisted of postmortem brain samples from 35 SCZ patients, 35 BPD patients and 35 healthy controls. The brain region used was anterior cingulate cortex (Brodmann region 24), a region known to be involved in learning and executive functions, which are deficit in SCZ patients. Image studies of SCZ patients also showed abnormalities in this region^{5,6}. The brain regions were dissected by the staff at SMRI, and total RNA was isolated by SMRI staff using the trizol method. Total RNA samples were delivered by dry ice to

Beijing Genomics Institute (BGI), China for whole transcriptome sequencing. Some RNA samples were degraded during the transportation, and several batches of delivery were made in a period of 6 months. RNA samples were quality-controlled by BGI staff using Agilent 2100 Bioanalyzer with RNA6000 test kit. The concentration of total RNA, RNA integrity number value and the ratio of 28S and 18S ribosomal RNA were measured. Samples with total RNA amount of 5 µg, concentration of 80 µg/µl, RNA integrity number 7, and 28S/18S ratio of 1.0 were selected for sequencing. Messenger RNAs were extracted from the total RNA by oligo(dT) beads. Selected mRNAs were then sheared randomly into 200 bp fragments, and these mRNA fragments were reverse-transcribed into cDNA using random hexamers. Synthesized cDNA was purified using QiaQuick PCR extraction kit, and ligated with sequencing adaptors, which contained sample ID information when multiple subjects were pooled. After limited amplification with PCR (15 cycles), the cDNA were size-selected and purified by agarose gel electrophoresis. cDNA with sizes between 200 and 300 bp were selected for library construction. Libraries were sequenced by Genome Analyzer II and HiSeq 2000 instruments using paired-end chemistry. The average reading length is about 90 bp. Due to RNA delivery and quality issues, library construction and sequencing were done in several batches. Batches 1-5 were sequenced with Genome Analyzer II and batch 6 was sequenced with HiSeq 2000. In batch 5, five samples were pooled and sequenced in a single lane.

3.2 Base-calling, read quality control and mapping

RNA-Seq was conducted over a period of 6 months using different Illumina sequencers. In order to ensure the consistency of the data, we redid base-calling of raw reads with HiSeq 2000 pipeline (Hiseq version HCS1.0.69). Sequence reads with adaptor contamination, or with more than 5% unscorable bases, or with 50% of the bases with low quality score (PHRED score 5) were removed. The cleaned reads were then aligned and mapped to the human genome reference sequence (UCSC hg18) using the SOAP software.^{7,8} For the alignment and mapping, the maximum of allowable mismatch was set to 3 for each read. Only reads uniquely mapped to genes were used to calculate the gene expression level. Reads per kilo-base per million reads (RPKM)⁹ were used as an index for gene expression. Single nucleotide polymorphisms (SNPs) were identified using SOAP SNP program.^{8,10}

3.3 Differential gene expression

SMRI provided information on many covariates for both primary and replication samples (detailed below), which had been shown to affect expression levels. We selected genes whose expression levels were high enough to be modeled as continuous data, and used a standard linear model for our analysis. It is generally considered that a continuous approximation fits adequately to count data if the mean count is at least five, and fewer than 20% of sample values are zero. Therefore we first removed genes with a mean count under five or over 20% zeros. We computed RPKM values for those genes. We further removed genes whose median or median absolute deviation (MAD) was low (< 0.5 RPKM or < 0.25 RPKM, respectively), leaving 14,454 genes for our differential expression analysis.

For purposes of multivariate analysis we used a log-transform of the RPKM measure, with a positive offset of the maximum of the 10th percentile of values or 0.1 to ensure the logarithm

was well-defined. We then fitted a linear model to estimate the effects of the technical and sample covariates believed to affect expression levels and subtracted these effects from the logarithmic values.

We did all differential expression calculations in R. We first noticed that there were many outliers (values at least 3 MAD from the median) in the gene expression data, and therefore we winsorized the data before fitting a linear model. We fitted a linear model for main effects of diagnosis on gene expression level, including the following nuisance sample covariates: age, sex, cumulative anti-psychotic use (which was square root transformed), and these technical covariates: cDNA concentration, RNA integrity number, brain pH, postmortem interval, and batch number. All of these covariates contributed significantly to several hundred or thousand genes. We did not include other potentially relevant covariates such as illegal drug use or alcohol, after testing that they seemed to contribute little or nothing to expression variation. We had insufficient data on smoking to include it as a covariate.

We fitted the linear model with a step-down procedure: for each gene we fitted the full model with all covariates and then retained in the model for that gene only those variables whose significance was moderate to high as assessed by p-value less than 0.16, which corresponds to a chi-square greater than 2; that in turn is equivalent to the widely used Akaike Information Criterion (AIC) for variable inclusion. Then we re-fitted the model using only the selected variables, including diagnosis, to obtain the coefficients, t-statistics, and p-values, which were reported in this article.

3.4 Correlation between differentially expressed genes between SCZ and BPD

To evaluate whether differentially expressed genes in SCZ correlated with those differentially expressed in BPD, we refined our differential gene expression analyses by using independent controls for SCZ and BPD. We divided our full set of controls randomly into 2 independent groups and fitted the full model for the SCZ cases with one group of the controls and fitted separately for BPD cases with the other group of controls. We repeated this process 100 times. For each random division we obtained a full set of t-scores for differential gene expression in SCZ and in BPD. We then computed the correlations between each of these statistically independent t-score sets.

In order to identify a set of genes, whose t-scores for SCZ and BPD were both concordant and larger than expected by chance, we developed a simple test based on the products of the t-scores. The product would be large only when both t-scores were fairly large and of the same sign. However, as already noted, we would expect some correlations between t-scores through having a common set of controls. Therefore, we generated a null distribution of products of t-scores with the correlation expected from having a common set of controls. We then applied a false discovery rate threshold of 0.1, which means that the number of genes expected to have products exceeding that threshold by chance was less than 10% of the number observed in the data. The set of genes exceeding this threshold was defined as differentially and concordantly expressed genes (DCEGs) between SCZ and BPD, and they would be used in further functional and enrichment analyses.

3.5 Association enrichment test for DCEGs

The significantly expressed genes in SCZ and BPD, as well as their DCEGs, were tested for enrichment of association signals from GWAS data in PGC SCZ¹¹ and BPD¹² cohorts. To perform the analysis we first obtained the gene statistics and p-values using the VEGAS sum of squares approach¹³. Based on VEGAS summaries, the enrichment of association signals was assessed using a Simes test for p-values and non-parametric relative rank p-value of genes significantly co-expressed at q-values of 0.1, 0.05 and 0.01, respectively.

3.6 Rare variants analysis

To determine if the frequency of transcribed rare variants (RVs, minor allele frequency in the sequenced population < 1 %) in DCEGs was elevated when compared to controls, for each individual, we first assessed the RV heterozygous status for SNPs in genes shown to be significantly co-expressed in SCZ and BPD patients. Subsequently, we tested for increased heterozygote rates in cases of SCZ and BPD when compared to controls. The test was carried out for genes found to be significantly co-expressed in patients at three levels of q-value:¹⁴ 0.1, 0.05 and 0.01. In our statistical tests the number of RVs for an individual was modeled as a dependent variable in a Poisson regression having the case-control status as a covariate. Because of the large differences in coverage between the different diagnostic groups, we performed the statistical analysis on a reduced data set obtained by fully matching cases and controls for coverage (as measured by RPKM) using MatchIt package in R.¹⁵

3.7 Pathway and pathway crosstalk analysis

We conducted pathway enrichment analysis of three gene sets: (1) differentially expressed genes in SCZ (105 genes); (2) differentially expressed genes in BPD (153 genes); and (3) the differentially and coordinately expressed genes between SCZ and BPD (q value = 0.05, 213 genes). In these enrichment analyses, we used hypergeometric test implemented in the tool WebGestalt (version 2)¹⁶ and the canonical pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. To avoid too many or too few genes to be considered in each pathway analysis, we only included the pathways whose sizes were between 5 and 250 genes.¹⁷ The p-values from hypergeometric tests were further adjusted by Benjamini-Hochberg method.¹⁸ Only those pathways with adjusted p-values < 0.05 were considered statistically significantly enriched.

We further examined how those enriched pathways interact with each other in terms of their function and regulation. Specifically, we applied the Character Sub-Pathway Network (CSPN) algorithm¹⁹ to sift significantly interacting pathway pairs. CSPN was designed to prioritize pathway pairs having a large number of pathway-bridging Protein-Protein Interactions (PPIs) that would be unlikely to exist in randomly permuted PPI networks (details are available in reference¹⁹). We used the human PPI data from the Protein Interaction Network Analysis (PINA) platform²⁰ (September 14, 2012) as the reference network in this pathway crosstalk analysis. Our working PPI network included a total of 11,318 nodes (protein-coding genes) and 67,936 interactions. We restricted the analysis specifically to the DCEGs and the pathways enriched by them. When running CSPN, a mode “OR” was selected, meaning that we considered all PPIs formed by the DCEGs as

well as their one-step extension. In the final step, we selected the significant pathway interaction pairs as having permutation p-values less than 0.05.

4 Results

4.1 Descriptive summary

We obtained total RNAs for all 105 subjects from the Stanley Foundation for transcriptome sequencing. Due to quantity and quality issues, only 83 subjects were sequenced (31 SCZ patients, 26 BPD patients, and 26 controls; one BPD subject, AC-36, was dropped from analyses per Stanley's recommendation). A summary of descriptive statistics of the data was provided in Table 1. Overall, we had more sequencing reads for the SCZ samples, and the coverage of genes and the number of SNPs identified were approximately proportional to the number of sequence reads.

4.2 Differentially expressed genes in SCZ and BPD

After mapping the reads to the human reference genome (hg18), we calculated RPKM values for all known transcripts for each of the sequenced subjects. After removing the genes with low counts, we treated the RPKM as continuous data, and modeled the main effect (diagnostic groups) by linear regression with age, sex, psychotic use, RNA integrity number, brain pH, postmortem interval, and batch number. In these analyses, we identified 105 and 153 genes differentially expressed in SCZ and BPD, respectively, by controlling false discovery rate using the Benjamini-Hochberg procedure¹⁸ (q value = 0.1) (Supplementary Table S1). Nine genes (*FER*, *FGF13*, *IDH1*, *DYNLT3*, *TRPV1*, *CBFA2T2*, *MTA2*, *IGSF11*, and *NAP1L5*) were shared in both lists. Functional analysis using KEGG pathway annotations revealed that the regulation of cytoskeleton, Huntington's disease and axon guidance pathways were enriched for genes differentially expressed in SCZ (Supplementary Table S2). A similar analysis was conducted with those genes differentially expressed in BPD, which resulted in a total of 26 pathways that were enriched with BPD genes (Supplementary Table S3). The regulation of cytoskeleton and Huntington's disease pathways were enriched in differentially expressed genes for both disorders.

4.3 Correlation in gene expression between SCZ and BPD

Based on polygenic analyses of SCZ and BPD, there was substantial overlap of genetic risks between the two disorders. We were interested in examining the correlation of gene expression between these two disorders. Since the control group was shared between SCZ and BPD, to avoid the bias caused by the shared controls, we used the bootstrap technique to randomly sample half (13 subjects) of the controls and compared gene expression of these control subjects with SCZ patients, and used the other half of the controls to calculate the t-scores for BPD. We then plot the median t-scores for the two disorders (Figure 1). Based on the 100 bootstraps, the median correlation between SCZ and BPD was 0.285 with a standard deviation of 0.10, which was statistically significant ($p < 0.002$). To facilitate further analyses of those concordantly expressed genes between SCZ and BPD, we took the product of t-scores of differential expression analyses between SCZ/control and BPD/control, and defined those genes with a t-product q value = 0.1 as differentially and concordantly expressed genes (DCEGs). In the following analyses, we used three cutoff q-values: 0.1,

0.05, and 0.01. There were 885 DCEGs by $q < 0.1$, 213 DCEGs by $q < 0.05$, and 53 DCEGs by $q < 0.01$ (Supplementary Table S4). These results indicated that there were substantial genes differentially expressed in both SCZ and BPD, and the expression of these genes was correlated.

4.4 DCEGs enrichment in association signals

To evaluate whether DCEGs were enriched in association signals for the disorders, we conducted enrichment test using the PGC stage I GWAS datasets from NIMH Genetics Repository (<https://www.nimhgenetics.org/>). We conducted 2 separate analyses. One was the sum of squares test, with which we evaluated whether the DCEGs had more small p-values than that in the null distribution. The other was the Simes test, by which we evaluated if the DCEGs had improved ranks relative to the genes ranked by GWASs. The results were summarized in Table 2. Genes with cut-offs at q-values of 0.01, 0.05 and 0.1 were all significantly enriched in association signals for both SCZ and BPD. However, the enrichment did not surpass that of the GWASs, as indicated by the Simes test of relative ranks. Of the 885 DCEGs, 758 genes were included in gene based association analyses of the GWAS datasets, and 105 genes showed nominally significant association with SCZ and 82 showed nominally significant association with BPD. There were 24 genes showed association ($p \leq 0.05$) with both diseases (Table 3). An examination of the functions of these genes showed that many of them were involved in membrane and intracellular vesicle trafficking and transport directly (SUFU, AKTIP and COG7) and indirectly (DDR1, MAN1A1, NTRK3, PPM1A, TGFA, C10ORG26 and LBH). These were consistent with the results of KEGG pathway crosstalk analyses of the DCEGs that showed interconnected regulation network centered on the regulation of cytoskeleton, phagocytosis and axon guidance (see below). Furthermore, other genes (*TGFA*, *JAM3*) were involved in cell-cell junction pathway that was implicated in SCZ. *DDR1*^{21,22}, *NTRK3*^{23,24} and *C10ORF26*^{25,26} had been reported to be associated with SCZ and BPD.

4.5 DCEGs enrichment in rare variants

There was evidence that rare variants contributed significantly to the genetic risks of both SCZ and BPD. We were interested in testing if DCEGs between SCZ and BPD were enriched in rare variants. Towards this end, we defined the rare variants as those observed only once in our RNA-Seq dataset. We first identified all SNPs in the 82 sequenced subjects. On average, there were about 1.5, 0.7 and 2.1 million SNPs identified for each subject in the controls, BPD and SCZ patients respectively (Table 1). The difference in the number of SNPs identified among the 3 groups was proportional to the depth of sequencing, as expected. In this analysis, we included only those SNPs with one observed heterozygote. For each observed rare variant, we matched it to individuals from other groups with similar sequencing coverage using a matching procedure implemented in MatchIt package in R. We found that subjects diagnosed with SCZ had more rare variants compared to healthy controls. However, we did not find more rare variants in BPD patients (Table 4). Since the power to discover rare variants was proportional to sequence coverage, the negative results for BPD could be due to the lower sequencing coverage of the BPD subjects.

4.6 Pathway enrichment analyses of the DCEGs

We conducted pathway enrichment analyses for the DCEGs between SCZ and BPD using the KEGG pathway annotations. Since there were too many genes ($n = 885$) with t-product q value > 0.1 , we selected the genes with q value cutoff at 0.05 (213 genes) for these analyses. We found 18 pathways that were statistically significantly enriched in the DCEGs (Table 5). Among the most significant pathways were lysosome, Huntington's disease, Fc gamma R-mediated phagocytosis and regulation of actin cytoskeleton. Of the 110 genes in the KEGG lysosome pathway, 19 genes had differential expression p values < 0.05 for SCZ, and 17 genes had p values < 0.05 for BPD. Moreover, we found that amongst the 104 genes with PGC GWAS data, 17 and 16 genes had gene-based p values < 0.05 in SCZ and BPD, respectively. Of these genes, 5 (*GNPTG*, *PPT2*, *AP3B2*, *ATP6V0D1* and *GGA2*) had association p values < 0.05 in both SCZ and BPD. The Simes test indicated that genes in lysosome pathway were enriched for association signals in both SCZ and BPD ($p = 0.02$ and 0.05 , respectively). For Huntington's disease pathway, there were 155 genes included in the PGC GWAS dataset. We found 26 and 9 genes showing nominal association with SCZ and BPD, respectively, and 2 genes (*NDUFAB1* and *CREB1*) were associated with both diseases. The Simes test indicated that Huntington's disease pathway was also enriched with association signals in both SCZ and BPD ($p = 0.005$ and 0.06 respectively).

To better explore the functional relationship of genes involved in SCZ and BPD, we performed pathway crosstalk analyses. Among the 18 pathways shown in Table 5 above, we found 13 pairs of pathways were statistically significantly interlinked (Figure 2). Three pathways had the strongest pair-wise interactions: Fc gamma R-mediated phagocytosis, Axon guidance, and regulation of actin cytoskeleton. The interactions of these three pathways were primarily contributed by interaction pairs of DCEGs: GSN, LIMK1, PAK3, and RDX (see Supplementary Figure S1). There are 10 genes (*CDC42*, *CFL1*, *CFL2*, *LIMK1*, *LIMK2*, *MAPK1*, *MAPK3*, *PAK1*, *RAC1*, and *RAC2*) shared amongst these 3 pathways. Based on these shared genes, a regulatory network shared by SCZ and BPD is emerging: small GTPases (*RAC1*, *RAC2* and *CDC42*) take upstream signals to activate *PAK1*, which in turn phosphorylates *LIMK1* and *LIMK2*, leading to the phosphorylation of *CFL1* and *CFL2*. Phosphorylated *CFL1* and *CFL2* directly regulate the assembling of actin filaments, resulting in changes of cytoskeleton in response to signals received at cell surface and intracellular stimuli (Figure 3). Additionally, *RAC1/RAC2* can also direct the signal to endocytosis,²⁷ regulating endosome/membrane trafficking.

4.7 Replication of enriched pathways and regulatory network

To verify the shared pathways and regulatory networks amongst the DCEGs, we obtained an independent RNA-Seq dataset. The subjects used in this replication data were from SMRI neuropathology collection, including 14 SCZ patients, 14 BPD patients and 15 healthy controls. The brain region used in this dataset was hippocampus, and the preparation of RNA and the sequencing were conducted researchers at SMIR. Standard procedures for base-calling and quality controls were applied. On average, the replication samples had higher sequencing reads (54.6 million reads/subject, and 912 reads/gene) compared to our primary samples (Table 1), but the RPKM was comparable (26.65 RPKM/gene). Due to

higher sequencing coverage, slightly more genes were covered by the sequencing (16,845 genes/subject).

We conducted similar analyses to identify differentially expressed genes in SCZ, and BPD (see Supplementary methods for details). In the differential expression analyses, we found that the p value distributions were largely indistinguishable from the null for both SCZ and BPD. When FDR techniques were applied, the best q value for SCZ was 0.2291, and the best q value for BPD was 0.7267. Based on these results, we failed to replicate the differentially expressed genes obtained from the primary samples for both SCZ and BPD. While disappointing, it was not unexpected since the replication sample size was just half of that of the primary sample. We were underpowered for replication.

Although we could not verify individual genes whether they were differentially expressed in SCZ and BPD, we reasoned that the top ranked genes (by p values) in the differential analyses were still likely to be enriched for those biological pathways involved in these diseases. Thus, we selected the same number of genes differentially expressed in the primary sample for SCZ (105) and BPD (151), and treated them as the most likely candidates for pathway and network analyses. The analyses identified 16 pathways enriched in the SCZ candidates and 22 pathways enriched in the BPD candidates (Supplementary Tables S6 and S7). The regulation of actin cytoskeleton pathway identified by the differentially expressed genes in SCZ from the primary sample was enriched ($p = 0.0268$) in the top-ranked SCZ genes from the replication sample. The lysosome pathway ($p = 0.0442$) was also enriched in these top-ranked SCZ genes. As described above, lysosome pathway was one of the most significant pathways identified by the DCEGs from the primary sample. No pathways identified from the top-ranked BPD candidates overlapped with that identified from the primary samples. The Fc gamma R-mediated phagocytosis pathway, another pathway identified by the DCEGs of primary sample, was statistically significantly enriched ($p = 0.0105$) among these top-ranked BPD candidates.

Using the same rationale, we selected 213 genes ranked by the t-product of the replication sample as our candidates of DCEGs, and performed pathway and pathway cross-talk analyses as with the primary samples. Pathway analyses revealed 13 pathways enriched in these top ranked DCEG candidates (Supplementary Table S8), the small cell lung cancer pathway ($p = 0.0066$) was a direct replication of the pathways found in the primary sample. Taking the 18 pathways from the primary sample as a whole set, we found no enrichment of these 18 pathways in the top-ranked DCEG candidates ($p\text{-value}=0.44$), but we did find significant enrichment in an extended set of these candidates together with their immediate PPI neighbors ($p\text{-value}=6.78 \times 10^{-10}$). This suggested that while these DCEG candidates were not reliably selected due to the sample size, they did include the immediate interacting proteins of the 18 pathways. To evaluate how these 18 pathways interacted with each other in the replication dataset, a cross-talk analysis was performed for the top-ranked candidate DCEGs in the PPI network background. This analysis showed that the interactions among the regulation of actin cytoskeleton, axon guidance, Fc gamma R-mediated phagocytosis and pancreatic cancer pathways had similar network between the primary and replication samples (comparing Figure 2A and 2B), suggesting that these interactions are likely to be shared between SCZ and BPD.

5 Conclusion and Discussion

SCZ and BPD share many clinic symptoms, and in recent years there is evidence that they share genetic etiology. Using GWAS data, Purcell et al.⁴ demonstrated that genetic liability scores of SCZ could reliably predict the diagnosis of BPD, indicating the shared liability between these two diseases. Analyses of gene expression also suggested the correlation of gene expression between the two disorders²⁸. However, there has been no report yet that integrates both GWAS and expression data to systematically examine the relationship between the two diseases. In this study, we conducted whole transcriptome sequencing analyses of 82 subjects of SCZ, BPD and controls, tested the expression correlation between the diseases for all differentially expressed genes, and compared the association of these genes with the diseases using GWAS data. In these analyses, we found many genes differentially expressed in either SCZ or BPD patients. Many of these genes have been reported in the literature with association evidence to either SCZ, or BPD, or both. Of particular interest is that we found many of these differentially expressed genes in SCZ and BPD were also correlated in expression between SCZ and BPD. Based on this finding, we sought to test if these DCEGs were enriched in genetic association signals, and furthermore, to identify biological pathways shared by these disorders. By using GWAS datasets for SCZ and BPD, we found that DCEGs were enriched in association signals for both SCZ and BPD, and many of these genes were involved in interconnected pathways. Using an independent transcriptome sequencing dataset from the hippocampus of SCZ and BPD patients, we largely replicated the core interactions of this interconnected network. These results suggested that the shared clinic presentation of symptoms may have common genetic etiology, and further study of these shared biological pathways and processes can provide valuable information to the treatment and diagnosis of these two disorders.

5.1 Correlation in gene expression and association between SCZ and BPD

It has long been speculated that similar clinic presentation of symptoms between SCZ and BPD might originate from shared pathophysiology and etiology. In recent years, accumulating evidence suggested that they share some genetic liabilities. Many candidate genes have been tested for association with SCZ and BPD, and some of them showed association with both diseases. Under polygenic framework, it is shown that aggregated risk factors for SCZ can reliably predict BPD diagnosis,⁴ confirming the shared genetic risks between these disorders. Expression analyses have also shown correlation between SCZ and BPD²⁸. However, no genetic associations are provided to support the observed correlation in gene expression. In this study, we showed that many genes differentially expressed in the two diseases were also expressed concordantly between the two diseases. Using GWAS data, we further demonstrated that many of these same genes, i.e. DCEGs, were associated with both diseases in gene-based testing (Table 3). To our knowledge, this present study is the first to show that for a significant number of genes, genetic association and gene expression are correlated between SCZ and BPD. When the functions of these genes are examined, they are involved in a limited number of biological pathways and processes that are interconnected (see below for more discussions). If this observation is proved to be true, it would have significant clinic implications. Drugs targeting these pathways have the

potential to treat both SCZ and BPD. As matter of fact, clinicians have used same drugs to treat both diseases and observed similar efficacies.

We also noticed that while these DCGEs were enriched in association signals, the degree of enrichment did not surpass that of GWASs since the Simes rank test was not significant for SCZ or BPD. This suggested that using data from gene expression alone is not very likely to identify those genes with small effects. In order to accomplish this goal, it is necessary to combine information from multiple different sources. In our analyses, we also found that the DCGEs were enriched with rare variants for SCZ. In contrast, we did not observe an enrichment of rare variants in BPD samples. We believe that this is likely due to the relatively lower sequencing coverage of the BPD samples. Another possible factor is that the sample size is too small to have enough power in rare variant comparison. Future studies using more samples will be needed to validate this finding.

5.2 SCZ and BPD share interconnected pathways

Analyses of DCEGs indicated that these genes were enriched in a variety of pathways (Table 5). Using an independent RNA-Seq data from the hippocampus, we were able to confirm one of the pathways (small cell lung cancer), and the core of pathway interaction network. Despite the smaller sample size and of different brain region of the replication sample, therefore, likely underpowered, the replication of the core interactions between the regulation of actin cytoskeleton, axon guidance and Fc gamma R-mediated phagocytosis implied its importance in both SCZ and BPD as well as their shared genetic risk between these interactions. Of note, in both the primary and replication samples, we observed that the pancreatic cancer pathway interacted significantly with the regulation of actin cytoskeleton, axon guidance and Fc gamma R-mediated phagocytosis. The implication of these interactions and the small cell lung cancer pathway was not immediately clear. One of the possibilities is that genes involved in pancreatic cancer and small cell lung cancer pathways have some novel functions in the brain tissues that have not been well studied so these genes remain associated primarily with cancers.

The pathways discovered in the DCEGs are biologically interconnected. The lysosome pathway was one of the most significant pathways implicated in our DCEG pathway analyses. While this pathway was not replicated by the top-ranked DCEG candidates of the replication sample, it was replicated by the top-ranked differentially expressed genes in SCZ. The lysosome pathway influences many cellular processes, such as vesicle trafficking, protein maturation and modification and degradation, endocytosis, phagocytosis, metabolic pathway and exocytosis, by changing the dynamics of available endosomes, which in turn are involved in a variety of processes regulating cell-cell communication, cell growth and differentiation, apoptosis and immune responses. Endocytosis can remove active membrane proteins such as receptors for neurotransmitters and growth factors from cytoplasm membrane, sequestering their signal transduction.²⁹⁻³¹ These proteins can then be recycled back to plasma membrane, or ubiquitinated and delivered to lysosome for degradation. Lysosomal enzymes are synthesized in the endoplasmic reticulum, and transported through the Golgi apparatus to endosomes, which then mature to fully functional lysosomes. Other membrane enzymes/receptors undergo post-translational modifications (glycosylation,

palmitoylation and other modifications) in the endoplasmic reticulum and then delivered to their destination via endosomes. Phagocytosis is a cellular response to clear dead cells or exogenous pathogens, in this process, dead cells or pathogens are engulfed and delivered to lysosome for digestion. The Fc gamma R-mediated phagocytosis pathway was identified by the top-ranked candidate genes in BPD as well. There is ample evidence that lysosomal dysregulation is linked to neuronal cell survival, leading to neurodegenerative diseases. In fact, there are suggestions that the Huntington's disease and Alzheimer's disease are caused in part by lysosomal dysfunction.³²⁻³⁵ Intriguingly, DCEGs were enriched for genes involved in both pathways of Huntington's and Alzheimer's diseases as well (Table 5).

The regulation of actin cytoskeleton is another pathway identified by the DCEGs, and confirmed by the top-ranked differentially expressed candidate genes in SCZ. It is involved in many of the same pathways described above. All the processes discussed above require remodeling of cytoskeleton. Normally, the initial events for cytoskeleton remodeling come from cell membrane. This could be the binding of G-protein coupled receptors (GPCR) to its ligands such as neurotransmitters, growth factors and other signaling molecules. Depending on the nature of the GPCRs, small Rho GTPases (*RAC1/RAC2*, *CDC42*) are activated. Activated small GTPases funnel the signal to PAK (p21 protein (*Cdc42/Rac*)-activated kinase), LIMK (LIM domain kinase) and CFL (cofilin), leading to the stabilization of actin filaments, therefore, remodeling the cytoskeleton.³⁶⁻³⁸ The backbone of this signaling process, i.e. from *RAC/CDC42* to CFL, is implicated in our analyses of the DCEGs. The three interacting pathways (regulation of actin cytoskeleton, Fc gamma receptor mediated phagocytosis, and axon guidance) share 10 genes (*CDC42*, *CFL1*, *CFL2*, *LIMK1*, *LIMK2*, *MAPK1*, *MAPK3*, *PAK1*, *RAC1*, and *RAC2*), forming the backbone described above. *PAK1* and *LIMK1* are amongst the DCEGs discovered in this study. In addition, several other genes involved in the same pathways, *CDC42EP1*, *CDC42EP2*, *GSN*, *PAK3*, *PLXNA4*, *PLXNB3*, *RDX*, and *SEMA6A* have differential and concordant expression between SCZ and BPD. There is ample evidence that rearrangement of cytoskeleton is essential for neuronal cell maturation and migration, neurite outgrowth, and maintenance of synaptic density and plasticity.^{36,37,39-41} In fact, there are reports that the *CDC42* signaling, especially the *PAK1* gene, is involved in SCZ.^{38,42,43} The semaphorin-plexin signaling is known to be involved in axon guidance, and is implicated in SCZ as well.⁴⁴⁻⁴⁶ Integrin signaling and WNT signaling, which are reported to be associated with SCZ and BPD,⁴⁷⁻⁵¹ also use the small GTPase pathways. Dysregulation and disturbance of these pathways in the central nervous system could have severe consequences and manifest as SCZ and BPD.

The axon guidance pathway interacts with the regulation of actin cytoskeleton and Fc gamma R-mediated phagocytosis pathways, constitutes the core interaction network (Figure 3A). In the replication sample we did not directly replicate this pathway. However, when all the top-ranked DCEG candidates from the replication sample were pooled for pathway interaction analyses, this pathway was found interacting with both Fc gamma R-mediated phagocytosis and regulation of actin cytoskeleton pathways (Figure 3B). Furthermore, the top-ranked DCEG candidates also identified the focal adhesion pathway (Supplementary Table S8), which was also identified by the top-ranked differentially expressed gene candidates in both SCZ and BPD (Supplementary Tables S6 and S7), a pathway that

regulates the cues for axon guidance and growth.⁵²⁻⁵⁴ The ECM (extracellular matrix) receptor interaction pathway, identified by the top ranked DCEG candidates of replication sample, is also involved in the regulation of axon guidance.^{55,56} All these suggest that the replication sample does provide supporting evidence that the axon guidance pathway is involved in both SCZ and BPD.

The notion that these interconnected pathways are involved in both SCZ and BPD is supported by several lines of independent evidence. Molecular studies have indicated that the lysosomal functions are involved in SCZ and BPD. *DTNBP1* gene was first identified as a risk gene for SCZ,⁵⁷ and was later found to play a key role in the BLOC1 (biogenesis of lysosome-related organelles complex-1) functions relevant to multiple psychiatric disorders.⁵⁸ *AP3B2* encodes an adaptor protein involved in cargo selection of clathrin mediated endocytosis, and it had been implicated in synaptic vesicle formation and recycling.^{59,60} Adaptor protein complex was found physically interacting with BLOC1, linking genetic risk of SCZ to lysosomal functions and endosome trafficking.⁶¹⁻⁶³ In fact, there is a hypothesis articulating that endocytosis is the common pathophysiology underlying SCZ and BPD.⁶⁴ This hypothesis is supported by the finding of involvement of clathrin-mediated endocytosis in both SCZ and BPD through a systematic examination of protein expression of mid-hippocampus samples of well-matched groups of 20 SCZ subjects, 20 BPD subjects, and 20 control subjects.⁶⁵ The subjects used in this proteomic study were the same as our primary sample. Although different brain regions were investigated, and more specific comparison is needed, the finding based on GWAS and transcriptomic data in this study generally converges on that from the proteomic data. Additionally, other known candidate genes for the two diseases are either a part of these pathways or modulated by these pathways. *DTNBP1* interacts with the BLOC1 complex, and is part of the lysosomal pathway.⁶⁶ *DISC1* is involved in the ERBB signaling, the ERB receptor, FGF receptor, and other neurotransmitter receptors that are all modulated by endocytosis and exocytosis, which in turn are hinged on the efficiency of lysosomes. *DISC1* is another risk gene to both SCZ and BPD⁶⁷ that has been found to be involved in the regulation of cytoskeleton.⁶⁸ More recently, analyses of copy number variations and *de novo* mutations also implicated cytoskeleton regulation in SCZ.^{69,70} *CDC42* is a key regulator of cytoskeleton and focal adhesion, and this pathway is reported to be involved in SCZ.^{42,43,71} *PLXNA2*, *PLXNB3* and *SEMA3A*, component of axon guidance pathway, are also reported to be associated with SCZ.⁴⁴⁻⁴⁶ Furthermore, dysfunction of these processes have been shown to influence immune responses and trigger apoptosis; both of which have been speculated to be involved in SCZ and BPD.⁷²⁻⁷⁶

Our study has some limitations. While the sample size of the primary sample is reasonable for expression study, it may be under powered for rare variant analyses. For the replication sample, the sample size may not have sufficient power to verify the findings of the primary sample. This may be the main reason why we could not clearly define the differentially expressed genes in SCZ and BPD, and the DCEGs in replication sample. The use of hippocampus to replicate results from cingulate cortex adds complexity in the interpretation of the results, although it is well established that both the cortex and hippocampus are involved in SCZ and BPD.^{65,77,78} While we have provided evidence from gene expression

and genetic association that these pathways and processes are shared between SCZ and BPD, we could not determine to what extent that a particular pathway influences SCZ or BPD, neither could we quantify the relative influences between the two diseases. Further research with other information is warranted to answer these questions.

In summary, by combining data from transcriptome-sequencing and GWAS, we provide corroborative evidence that an interconnected pathway network of remodeling of actin cytoskeleton and regulation of lysosomal function underlies the common causes of SCZ and BPD. This inference is consistent with a large body of literature of SCZ and BPD studies. We believe that further study of these pathways will provide information to clarify the commonality and difference between the two diseases, and help clinic diagnosis and treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This study was supported by an Independent Investigator Award from NARSAD to X.C., and by National Institutes of Health grant (R01LM011177). We thank the Stanley Medical Institute for providing the brain specimens for transcriptome sequencing. We are grateful to the PGC investigators for providing their data to the NIMH Genetics Repository. We thank three reviewers for their valuable comments, which improved the quality of this manuscript.

References

1. Sheng G, Demers M, Subburaju S, Benes FM. Differences in the circuitry-based association of copy numbers and gene expression between the hippocampi of patients with schizophrenia and the hippocampi of patients with bipolar disorder. *Arch Gen Psychiatry*. 2012; 69:550–561. [PubMed: 22309971]
2. Karege F, Meary A, Perroud N, Jamain S, Leboyer M, Ballmann E, et al. Genetic overlap between schizophrenia and bipolar disorder: a study with AKT1 gene variants and clinical phenotypes. *Schizophr Res*. 2012; 135:8–14. [PubMed: 22277669]
3. Williams HJ, Craddock N, Russo G, Hamshere ML, Moskvin V, Dwyer S, et al. Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet*. 2011; 20:387–391. [PubMed: 21037240]
4. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
5. Kubicki M, Westin CF, Nestor PG, Wible CG, Frumin M, Maier SE, et al. Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Biol Psychiatry*. 2003; 54:1171–1180. [PubMed: 14643084]
6. Yucel M, Pantelis C, Stuart GW, Wood SJ, Maruff P, Velakoulis D, et al. Anterior cingulate activation during Stroop task performance: a PET to MRI coregistration study of individual patients with schizophrenia. *Am J Psychiatry*. 2002; 159:251–254. [PubMed: 11823267]
7. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. 2008; 24:713–714. [PubMed: 18227114]
8. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009; 25:1966–1967. [PubMed: 19497933]
9. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–628. [PubMed: 18516045]

10. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010; 329:75–78. [PubMed: 20595611]
11. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011; 43:969–976. [PubMed: 21926974]
12. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011; 43:977–983. [PubMed: 21926972]
13. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010; 87:139–145. [PubMed: 20598278]
14. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003; 100:9440–9445. [PubMed: 12883005]
15. Ho D, Imai K, King G, Stuart E. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal*. 2007; 15:199–236.
16. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res*. 2005; 33:W741, W748. [PubMed: 15980575]
17. Jia P, Liu Y, Zhao Z. Integrative pathway analysis of genome-wide association studies and gene expression data in prostate cancer. *BMC Syst Biol*. 2012; 6(Suppl 3):S13. [PubMed: 23281744]
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995; 57:289–300.
19. Huang Y, Li S. Detection of characteristic sub pathway network for angiogenesis based on the comprehensive pathway network. *BMC Bioinformatics*. 2010; 11(Suppl 1):S32. [PubMed: 20122205]
20. Wu J, Vallenius T, Ovaska K, Westermarck J, Makela TP, Hautaniemi S. Integrated network analysis platform for protein-protein interactions. *Nat Methods*. 2009; 6:75–77. [PubMed: 19079255]
21. Roig B, Franco-Pons N, Martorell L, Tomas J, Vogel WF, Vilella E. Expression of the tyrosine kinase discoidin domain receptor 1 (DDR1) in human central nervous system myelin. *Brain Res*. 2010; 1336:22–29. [PubMed: 20380825]
22. Roig B, Virgos C, Franco N, Martorell L, Valero J, Costas J, et al. The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Mol Psychiatry*. 2007; 12:833–841. [PubMed: 17440435]
23. Athanasiu L, Mattingsdal M, Melle I, Inderhaug E, Lien T, Agartz I, et al. Intron 12 in NTRK3 is associated with bipolar disorder. *Psychiatry Res*. 2011; 185:358–362. [PubMed: 20554328]
24. Otnaess MK, Djurovic S, Rimol LM, Kulle B, Kahler AK, Jonsson EG, et al. Evidence for a possible association of neurotrophin receptor (NTRK-3) gene polymorphisms with hippocampal function and schizophrenia. *Neurobiol Dis*. 2009; 34:518–524. [PubMed: 19344762]
25. Kwon E, Wang W, Tsai LH. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Mol Psychiatry*. 2013; 18:11–12. [PubMed: 22182936]
26. Kim AH, Parker EK, Williamson V, McMichael GO, Fanous AH, Vladimirov VI. Experimental validation of candidate schizophrenia gene ZNF804A as target for hsa-miR-137. *Schizophr Res*. 2012; 141:60–64. [PubMed: 22883350]
27. Qualmann B, Kessels MM. Endocytosis and the cytoskeleton. *Int Rev Cytol*. 2002; 220:93–144. [PubMed: 12224553]
28. Kim S, Webster MJ. Correlation analysis between genome-wide expression profiles and cytoarchitectural abnormalities in the prefrontal cortex of psychiatric disorders. *Mol Psychiatry*. 2010; 15:326–336. [PubMed: 18762803]
29. Niehrs C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol*. 2012; 13:767–779. [PubMed: 23151663]
30. van der Sluijs P, Hoogenraad CC. New insights in endosomal dynamics and AMPA receptor trafficking. *Semin Cell Dev Biol*. 2011; 22:499–505. [PubMed: 21843653]
31. Scita G, Di Fiore PP. The endocytic matrix. *Nature*. 2010; 463:464–473. [PubMed: 20110990]

32. Lipinski MM, Zheng B, Lu T, Yan Z, Py BF, Ng A, et al. Genome-wide analysis reveals mechanisms modulating autophagy in normal brain aging and in Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2010; 107:14164–14169. [PubMed: 20660724]
33. Pimplikar SW, Nixon RA, Robakis NK, Shen J, Tsai LH. Amyloid-independent mechanisms in Alzheimer's disease pathogenesis. *J Neurosci*. 2010; 30:14946–14954. [PubMed: 21068297]
34. Koga H, Martinez-Vicente M, Arias E, Kaushik S, Sulzer D, Cuervo AM. Constitutive upregulation of chaperone-mediated autophagy in Huntington's disease. *J Neurosci*. 2011; 31:18492–18505. [PubMed: 22171050]
35. Martinez-Vicente M, Tallozy Z, Wong E, Tang G, Koga H, Kaushik S, et al. Cargo recognition failure is responsible for inefficient autophagy in Huntington's disease. *Nat Neurosci*. 2010; 13:567–576. [PubMed: 20383138]
36. Auer M, Hausott B, Klimaschewski L. Rho GTPases as regulators of morphological neuroplasticity. *Ann Anat*. 2011; 193:259–266. [PubMed: 21459565]
37. de C.I. Functions of Rac GTPases during neuronal development. *Dev Neurosci*. 2008; 30:47–58. [PubMed: 18075254]
38. Deo AJ, Goldszer IM, Li S, Dibitto JV, Henteleff R, Sampson A, et al. PAK1 Protein Expression in the Auditory Cortex of Schizophrenia Subjects. *PLoS ONE*. 2013; 8:e59458. [PubMed: 23613712]
39. Bellenchi GC, Gurniak CB, Perlas E, Middei S, Ammassari-Teule M, Witke W. N cofilin is associated with neuronal migration disorders and cell cycle control in the cerebral cortex. *Genes Dev*. 2007; 21:2347–2357. [PubMed: 17875668]
40. Kreis P, Barnier JV. PAK signalling in neuronal physiology. *Cell Signal*. 2009; 21:384–393. [PubMed: 19036346]
41. Lundquist EA. Rac proteins and the control of axon development. *Curr Opin Neurobiol*. 2003; 13:384–390. [PubMed: 12850224]
42. Rubio MD, Haroutunian V, Meador-Woodruff JH. Abnormalities of the Duo/Ras-related C3 botulinum toxin substrate 1/p21-activated kinase 1 pathway drive myosin light chain phosphorylation in frontal cortex in schizophrenia. *Biol Psychiatry*. 2012; 71:906–914. [PubMed: 22458949]
43. Ide M, Lewis DA. Altered cortical CDC42 signaling pathways in schizophrenia: implications for dendritic spine deficits. *Biol Psychiatry*. 2010; 68:25–32. [PubMed: 20385374]
44. Mah S, Nelson MR, DeLisi LE, Reneland RH, Markward N, James MR, et al. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry*. 2006; 11:471–478. [PubMed: 16402134]
45. Rujescu D, Meisenzahl EM, Krejcova S, Giegling I, Zetzsche T, Reiser M, et al. Plexin B3 is genetically associated with verbal performance and white matter volume in human brain. *Mol Psychiatry*. 2007; 12:190–4. [PubMed: 17033634]
46. Eastwood SL, Law AJ, Everall IP, Harrison PJ. The axonal chemorepellant semaphorin 3A is increased in the cerebellum in schizophrenia and may contribute to its synaptic pathology. *Mol Psychiatry*. 2003; 8:148–155. [PubMed: 12610647]
47. Hattori T, Shimizu S, Koyama Y, Yamada K, Kuwahara R, Kumamoto N, et al. DISC1 regulates cell-cell adhesion, cell-matrix adhesion and neurite outgrowth. *Mol Psychiatry*. 2010; 15:798–809.
48. Walsh MT, Ryan M, Hillmann A, Condren R, Kenny D, Dinan T, et al. Elevated expression of integrin alpha(IIb) beta(IIIa) in drug-naive, first-episode schizophrenic patients. *Biol Psychiatry*. 2002; 52:874–879. [PubMed: 12399140]
49. Singh KK, De RG, Drane L, Mao Y, Flood Z, Madison J, et al. Common DISC1 polymorphisms disrupt Wnt/GSK3beta signaling and brain development. *Neuron*. 2011; 72:545–558. [PubMed: 22099458]
50. Brennan KJ, Simone A, Jou J, Gelboin-Burkhardt C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. *Nature*. 2011; 473:221–225. [PubMed: 21490598]
51. Freyberg Z, Ferrando SJ, Javitch JA. Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. *Am J Psychiatry*. 2010; 167:388–396. [PubMed: 19917593]

52. Woo S, Rowan DJ, Gomez TM. Retinotopic mapping requires focal adhesion kinase-mediated regulation of growth cone adhesion. *J Neurosci.* 2009; 29:13981–13991. [PubMed: 19890008]
53. Bechara A, Nawabi H, Moret F, Yaron A, Weaver E, Bozon M, et al. FAK-MAPK-dependent adhesion disassembly downstream of L1 contributes to semaphorin3A-induced collapse. *EMBO J.* 2008; 27:1549–1562. [PubMed: 18464795]
54. Li W, Lee J, Vikis HG, Lee SH, Liu G, Aurandt J, et al. Activation of FAK and Src are receptor-proximal events required for netrin signaling. *Nat Neurosci.* 2004; 7:1213–1221. [PubMed: 15494734]
55. Cho JY, Chak K, Andreone BJ, Wooley JR, Kolodkin AL. The extracellular matrix proteoglycan perlecan facilitates transmembrane semaphorin-mediated repulsive guidance. *Genes Dev.* 2012; 26:2222–2235. [PubMed: 23028146]
56. Myers JP, Santiago-Medina M, Gomez TM. Regulation of axonal outgrowth and pathfinding by integrin-ECM interactions. *Dev Neurobiol.* 2011; 71:901, 923. [PubMed: 21714101]
57. Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet.* 2002; 71:337–348. [PubMed: 12098102]
58. Ryder PV, Faundez V. Schizophrenia: the “BLOC” may be in the endosomes. *Sci Signal.* 2009; 2:e66.
59. Faundez V, Horng JT, Kelly RB. A function for the AP3 coat complex in synaptic vesicle formation from endosomes. *Cell.* 1998; 93:423–432. [PubMed: 9590176]
60. Newell-Litwa K, Salazar G, Smith Y, Faundez V. Roles of BLOC-1 and adaptor protein-3 complexes in cargo sorting to synaptic vesicles. *Mol Biol Cell.* 2009; 20:1441–1453. [PubMed: 19144828]
61. Di Pietro SM, Falcon-Perez JM, Tenza D, Setty SR, Marks MS, Raposo G, et al. BLOC-1 interacts with BLOC-2 and the AP-3 complex to facilitate protein trafficking on endosomes. *Mol Biol Cell.* 2006; 17:4027–4038. [PubMed: 16837549]
62. Larimore J, Tornieri K, Ryder PV, Gokhale A, Zlatic SA, Craig B, et al. The schizophrenia susceptibility factor dysbindin and its associated complex sort cargoes from cell bodies to the synapse. *Mol Biol Cell.* 2011; 22:4854–4867. [PubMed: 21998198]
63. Gokhale A, Larimore J, Werner E, So L, Moreno-De-Luca A, Lese-Martin C, et al. Quantitative proteomic and genetic analyses of the schizophrenia susceptibility factor dysbindin identify novel roles of the biogenesis of lysosome-related organelles complex 1. *J Neurosci.* 2012; 32:3697–3711. [PubMed: 22423091]
64. Schubert KO, Focking M, Prehn JH, Cotter DR. Hypothesis review: are clathrin-mediated endocytosis and clathrin-dependent membrane and protein trafficking core pathophysiological processes in schizophrenia and bipolar disorder? *Mol Psychiatry.* 2012; 17:669–681. [PubMed: 21986877]
65. Focking M, Dicker P, English JA, Schubert KO, Dunn MJ, Cotter DR. Common proteomic changes in the hippocampus in schizophrenia and bipolar disorder and particular evidence for involvement of cornu ammonis regions 2 and 3. *Arch Gen Psychiatry.* 2011; 68:477–488. [PubMed: 21536977]
66. Guo AY, Sun J, Riley BP, Thiselton DL, Kendler KS, Zhao Z. The dystrobrevin-binding protein 1 gene: features and networks. *Mol Psychiatry.* 2009; 14:18–29. [PubMed: 18663367]
67. Hodgkinson CA, Goldman D, Jaeger J, Persaud S, Kane JM, Lipsky RH, et al. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet.* 2004; 75:862–872. [PubMed: 15386212]
68. Wang Q, Brandon NJ. Regulation of the cytoskeleton by Disrupted-in-schizophrenia 1 (DISC1). *Mol Cell Neurosci.* 2011; 48:359–364. [PubMed: 21757008]
69. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry.* 2012; 17:142–153. [PubMed: 22083728]
70. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature.* 2014; 506:179–184. [PubMed: 24463507]

71. Myers JP, Robles E, Ducharme-Smith A, Gomez TM. Focal adhesion kinase modulates Cdc42 activity downstream of positive and negative axon guidance cues. *J Cell Sci.* 2012; 125:2918–2929. [PubMed: 22393238]
72. Chen X, Sun C, Chen Q, O'Neill FA, Walsh D, Fanous AH, et al. Apoptotic engulfment pathway and schizophrenia. *PLoS ONE.* 2009; 4:e6875. [PubMed: 19721717]
73. Jarskog LF, Glantz LA, Gilmore JH, Lieberman JA. Apoptotic mechanisms in the pathophysiology of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry.* 2005; 29:846–858. [PubMed: 15908096]
74. Miuller N, Schwarz MJ. The immunological basis of glutamatergic disturbance in schizophrenia: towards an integrated view. *J Neural Transm Suppl.* 2007:269–280. [PubMed: 17982903]
75. Nunes SO, Matsuo T, Kaminami MS, Watanabe MA, Reiche EM, Itano EN. An autoimmune or an inflammatory process in patients with schizophrenia, schizoaffective disorder, and in their biological relatives. *Schizophr Res.* 2006; 84:180–182. [PubMed: 16530388]
76. Xu J, Sun J, Chen J, Wang L, Li A, Helm M, et al. RNA-Seq analysis implicates dysregulation of the immune system in schizophrenia. *BMC Genomics.* 2012; 13(Suppl 8):S2. [PubMed: 23282246]
77. Sheng G, Demers M, Subburaju S, Benes FM. Differences in the circuitry-based association of copy numbers and gene expression between the hippocampi of patients with schizophrenia and the hippocampi of patients with bipolar disorder. *Arch Gen Psychiatry.* 2012; 69:550–561. [PubMed: 22309971]
78. Erk S, Meyer-Lindenberg A, Schmierer P, Mohnke S, Grimm O, Garbusow M, et al. Hippocampal and Frontolimbic Function as Intermediate Phenotype for Psychosis: Evidence from Healthy Relatives and a Common Risk Variant in CACNA1C. *Biol Psychiatry.* 2013

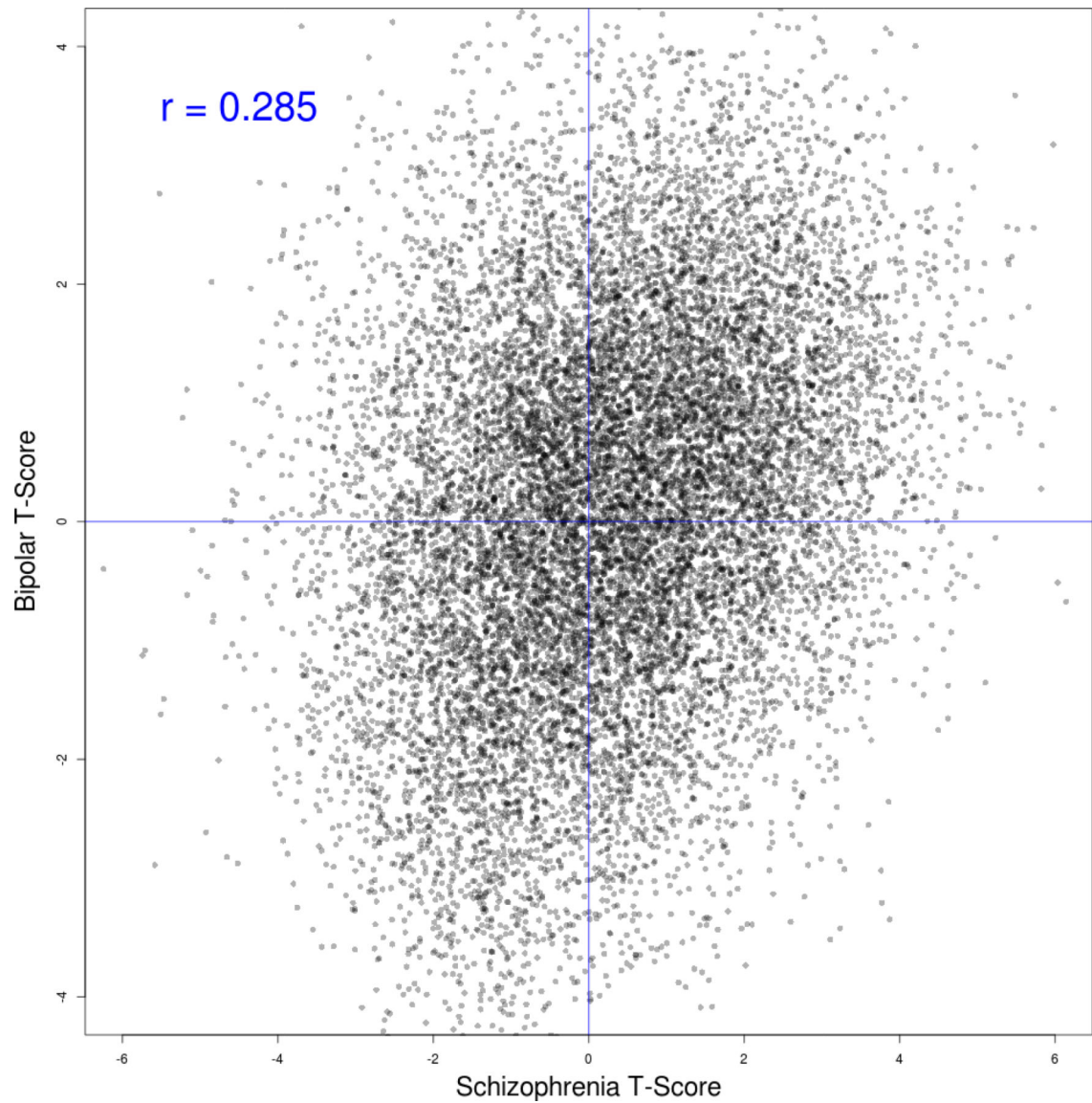


Figure 1.

Correlation of gene expression between schizophrenia and bipolar disorder. T-test scores from differential expression analyses of schizophrenia and bipolar disorder were plotted. This plot is a typical among the 100 trials in which half of the control samples (n=26) were randomly selected for controls for schizophrenia and the other half for bipolar disorder. The correlation coefficient ($r = 0.285 \pm 0.10$) was the median of these 100 trials.

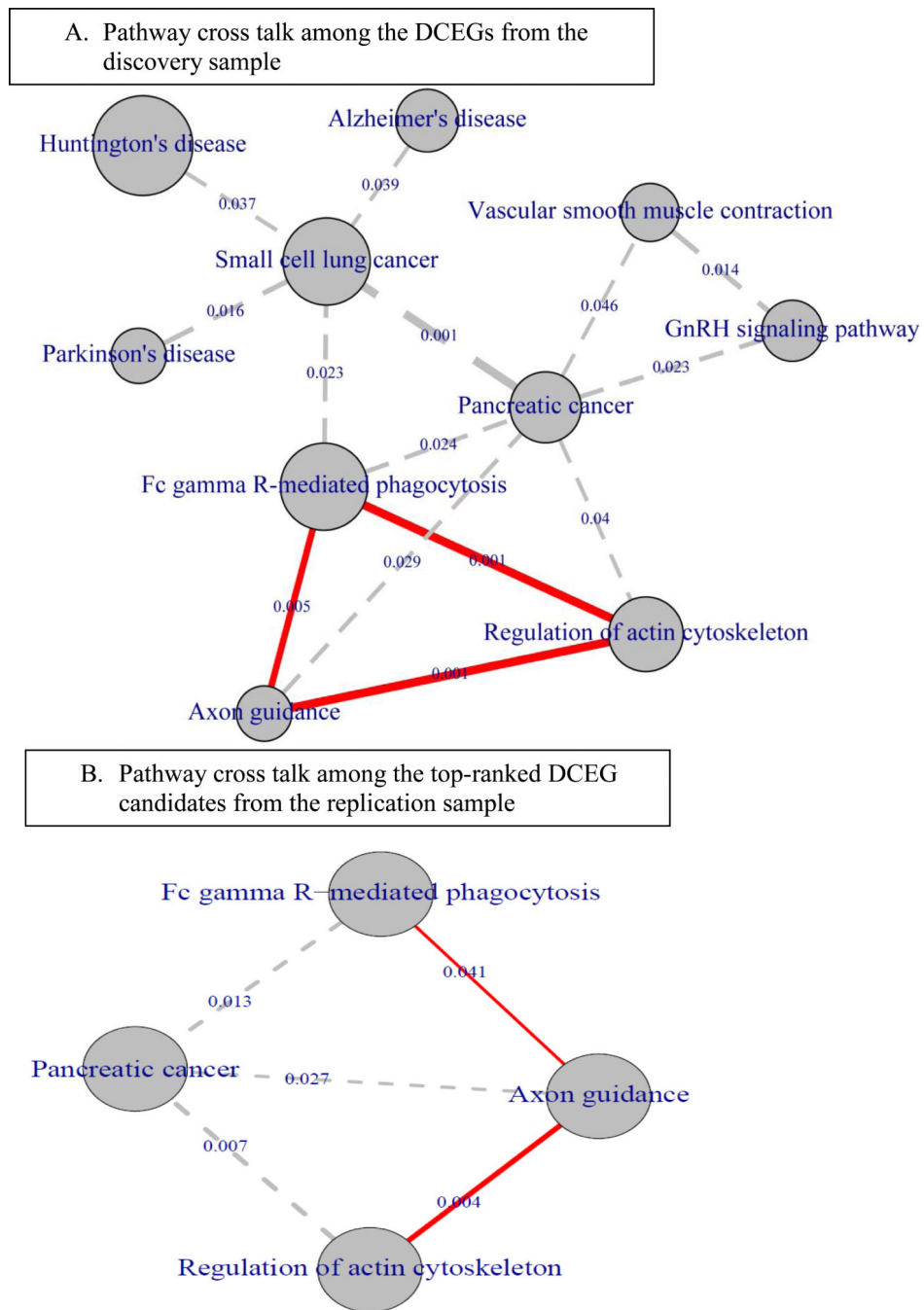


Figure 2. Pathway interaction networks for differentially and concordantly expressed genes (DCEG) and top-ranked candidates between schizophrenia and bipolar disorder. A. Interaction network constructed from the pathways significantly enriched with DCEGs from the discovery sample. Node size and edge width are proportional to pathway enrichment significance level and pathway interaction significance level, respectively. Fc gamma receptor mediated phagocytosis, regulation of actin cytoskeleton, and axon guidance pathways had the strongest mutual interactions, and the triangle formed by these three

pathways were highlighted by solid, red edges. B. Interaction network constructed from the top-ranked DCEG candidates from the replication sample. The legends were the same as in A. The interactions between the regulation of actin cytoskeleton, axon guidance, Fc gamma R-mediated phagocytosis and pancreatic cancer were replicated between the discovery and replication samples.

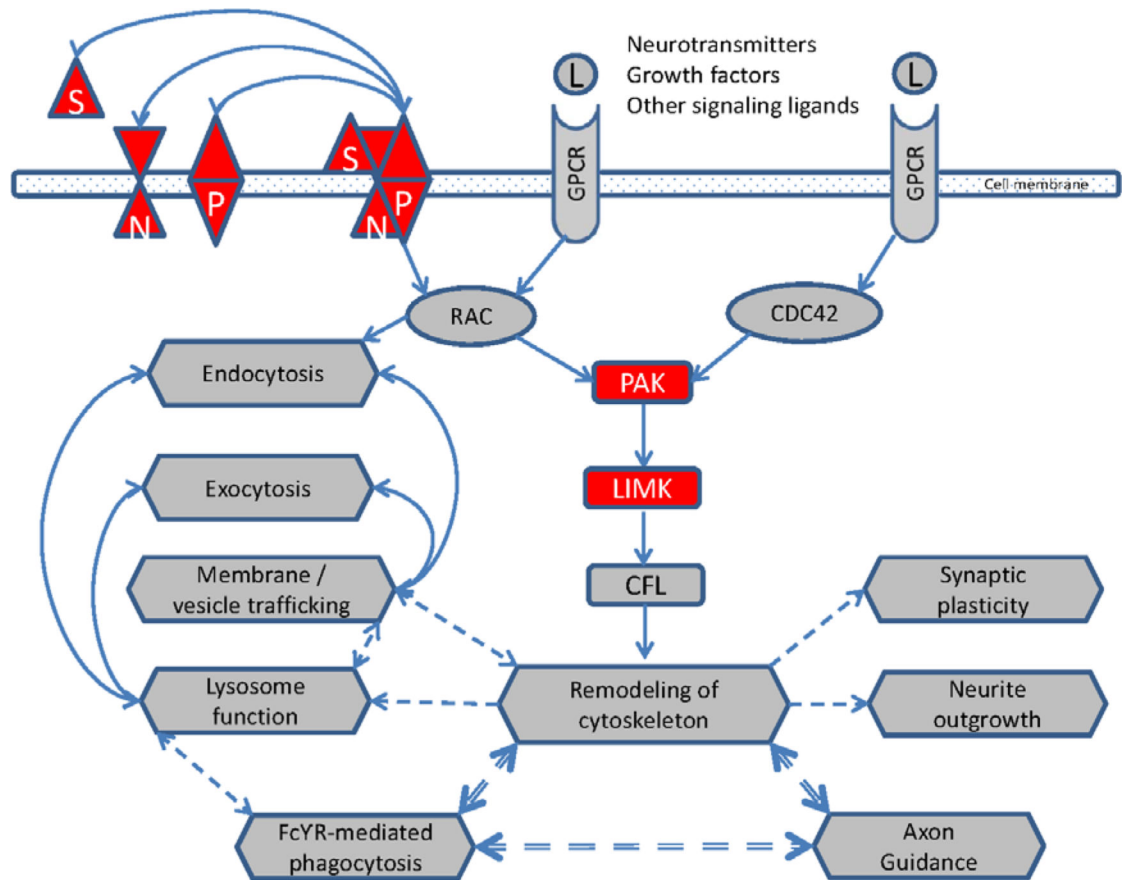


Figure 3. Biological processes implicated from the pathways enriched with differentially and coordinately expressed genes in schizophrenia and bipolar disorder. The solid arrows indicate known regulatory relationships. The dashed arrows denote relationships with some unknown links. Double-lined arrows denote the interacting relationships identified in this study. Red color denotes differentially and coordinately expressed genes in schizophrenia and bipolar disorder. L, ligand; N, neurophilin; P, plexin; and S, semaphorin.

Table 1

Sequencing summary

Group	N	Total Reads /Subject	Reads Mapped to Genome (%)	Reads Mapped to Gene (%)	Reads /Gene	RPKM /Gene	Coverage /Gene (%)	# Genes Sequenced	# SNPs Identified
CTRL	26	11,426,942	9999730 (87.5)	5320966 (46.5)	315	24.69	0.55	15,879	1,517,598
BPD	26	5,272,478	4665049 (88.5)	2314132 (43.8)	144	25.92	0.47	15,294	695,509
SCZ	31	14,733,231	12894782 (87.5)	6418990 (43.5)	378	23.59	0.62	16,640	2,128,528

RPKM: reads per kilo-base-pair per million reads.

Table 2

Differentially and concordantly expressed genes (DCEGs) between schizophrenia and bipolar disorder are enriched with association signals

Q-value cut-off (Number of genes)	SCZ		BPD	
	Sum of Squares Enrichment p-value	Simes Rank p-value	Sum of Squares Enrichment p-value_	Simes Rank p-value
0.10 (885)	10^{-7}	0.95	0.029	0.56
0.05 (213)	0.006	0.94	0.007	0.17
0.01 (53)	0.012	0.71	0.039	0.40

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

List of differentially and concordantly expressed genes (DCEGs) that are associated with both SCZ and BPD

DCEG	SCZ Expression p-value	BPD Expression p-value	SCZ GWAS p-value	BPD GWAS p-value	Function
DDR1	0.0377	0.0099	0.0001	0.0175	Collagen activated receptor tyrosine kinase, neurite outgrowth
LSS	0.0036	0.0255	0.0021	0.0065	Sterol biosynthesis, cholesterol metabolism, alcohol metabolism
MAN1A1	0.0461	0.0012	0.0020	0.0145	Mannosidase activity, hydrolase activity, protein glycosylation
NTRK3	0.0091	0.0042	0.0000	0.0039	Neurotrophin signaling, ErbB signaling
PPM1A	0.0749	0.0055	0.0120	0.0327	Protein phosphatase, negative regulator of stress response pathways
TGFA	0.0261	0.0274	0.0146	0.0367	Growth factor, involved in cancers, regulation of actin cytoskeleton, adherens junction
PRPF3	0.0260	0.0013	0.0095	0.0034	Pre-mRNA processing factor, mRNA splicing
MRPL33	0.0176	0.0070	0.0131	0.0025	Mitochondrial ribosome protein
MATR3	0.0281	0.0190	0.0034	0.0474	Nuclear matrix protein
CNNM3	0.0000	0.0208	0.0329	0.0028	Cyclin M3, possible magnesium transporter
FAM49B	0.0369	0.0022	0.0067	0.0126	-
SUFU	0.0127	0.0211	0.0136	0.0205	Negative regulator of hedgehog signaling
C10orf26	0.0449	0.0074	0.0001	0.0344	Regulated by miR-137, glycosylation
C2orf18	0.0136	0.0043	0.0184	0.0034	-
OTUD7B	0.0249	0.0016	0.0221	0.0112	-
SLC12A9	0.0351	0.0032	0.0392	0.0234	Potassium/chloride transporter
SLC44A2	0.0196	0.0012	0.0116	0.0345	-
AS3MT	0.0205	0.0190	0.0000	0.0394	Arsenic metabolism
AKTIP	0.0059	0.0235	0.0009	0.0075	Endosome to lysosome transport, lysosome organization, vascular transport
FNDC4	0.0049	0.0378	0.0000	0.0166	-
ZNF34	0.1002	0.0016	0.0421	0.0177	-
LBH	0.0167	0.0003	0.0185	0.0273	Transcription factor involved in limb bud and heart development, WNT signaling pathway
JAM3	0.0023	0.0268	0.0009	0.0064	Tight junction, cell-cell junction, adherens junction
COG7	0.0504	0.0019	0.0090	0.0013	Golgi vesicle transport, intracellular transport, protein glycosylation

Table 4

DCEGs enriched with rare variants in schizophrenia subjects

Q-value cut-off (Number of genes)	SCZ		BPD	
	Coefficient	p-value	Coefficient	p-value
0.10 (885)	0.36	6×10^{-7}	-0.24	0.850
0.05 (213)	0.40	0.007	0.15	0.730
0.01 (53)	0.39	0.200	-0.06	0.720

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Pathways enriched with DCEGs (q<0.05)

Pathway Name	# Genes	Contributing Genes	Adjusted P-Value ^a	DCEG Candidates from Replication Dataset
Lysosome	5	<i>AP3B1, TPP1, CTNS, CTSK, AP1S2</i>	0.0041	<i>SLC11A1</i>
Huntington's disease	6	<i>NDUFA4, CYCS, NDUFA5, CREB5, REST, ITPR1</i>	0.0041	<i>NDUFV3</i>
Small cell lung cancer	4	<i>CYCS, TRAF4, CDK4, LAMB3</i>	0.0082	<i>COL4A1, COL4A2, LAMC2, CCNE2</i>
Fc gamma R-mediated phagocytosis	4	<i>MARCKSL1, PLA2G4A, GSN, LIMK1</i>	0.0082	<i>RAC2</i>
Glyoxylate and dicarboxylate metabolism	2	<i>GRHPR, AFMID</i>	0.0123	<i>MCEE</i>
Regulation of actin cytoskeleton	5	<i>GSN, RDX, PAK3, FGF13, LIMK1</i>	0.0164	<i>ITGA5, RAC2, SOS2</i>
Pancreatic cancer	3	<i>RALB, SMAD4, CDK4</i>	0.0200	<i>RAC2</i>
Alzheimer's disease	4	<i>NDUFA4, CYCS, NDUFA5, ITPR1</i>	0.0316	<i>NDUFV3</i>
Glycine, serine and threonine metabolism	2	<i>CHDH, PSAT1</i>	0.0328	<i>PSPH</i>
Fructose and mannose metabolism	2	<i>PFKFB4, FUK</i>	0.0346	<i>FPGT</i>
GnRH signaling pathway	3	<i>ADCY6, PLA2G4A, ITPR1</i>	0.0346	<i>SOS2</i>
Basal transcription factors	2	<i>GTF2IRD1, GTF2E1</i>	0.0346	
Pyruvate metabolism	2	<i>GRHPR, LDHD</i>	0.0394	
Vascular smooth muscle contraction	3	<i>ADCY6, PLA2G4A, ITPR1</i>	0.0407	<i>ACTG2</i>
Amino sugar and nucleotide sugar metabolism	2	<i>FUK, UGP2</i>	0.0417	<i>FPGT</i>
Glutathione metabolism	2	<i>IDH1, MGST3</i>	0.0475	
Axon guidance	3	<i>PAK3, LIMK1, SEMA6A</i>	0.0475	<i>EFNB1, RAC2</i>
Parkinson's disease	3	<i>NDUFA4, CYCS, NDUFA5</i>	0.0478	<i>NDUFV3</i>

^a p-value was calculated by hypergeometric test and adjusted by Benjamini-Hochberg method.