De novo mutations involved in post-transcriptional dysregulation contribute to six neuropsychiatric disorders

Fengbiao Mao[1,2¶], Lu Wang[3¶], Qi Liu[4], Luoyuan Xiao[5], Xianfeng Li[6], Xin He[7], Rajesh C. Rao[2], Jinchen Li[1], Yali Dou[2*] and Zhong Sheng Sun[1*]

[1] Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China

[2] Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA.

[3] Institute of Life Science, Southeast University, Nanjing 210096, China

[4] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

[5] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

[6] State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, 410078, China.

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

¶These authors contributed equally to this work

[*] Corresponding authors

E-mail: sunzs@mail.biols.ac.cn (ZSS), yalid@umich.edu (YLD)

Abstract

While *de novo* mutations (DNMs) conferring risk in neuropsychiatric disorders have been revealed by next-generation sequencing, the role of DNMs involved in post-transcriptional regulation in pathogenesis of these disorders remains to be elucidated. By curating 42,871 DNMs from 9,772 core families with six kinds of neuropsychiatric disorders and controls, we identified 2,381 post-transcriptionally impaired DNMs (piDNMs) and prioritized 1,923 candidate casual genes in these six disorders by employing workflow RBP-Var. Our results revealed a higher prevalence of piDNM in the probands than that of controls (P = 9.52E-17). Moreover, we identified 214 piDNM-containing genes with shared enriched co-expression modules and intensive protein-protein interactions (P = 7.75E-07) in at least two of neuropsychiatric disorders. Furthermore, these cross-disorder genes carrying piDNMs could form interaction network centered on RNA binding proteins, suggesting a shared post-transcriptional etiology underlying these disorders. Our findings highlight that piDNMs contribute to the pathogenesis of neuropsychiatric disorders.

Introduction

In the past decades, there have been increased efforts to better understand the pathogenesis of psychiatric disorders. Next-generation sequencing, which allows genome-wide detection of rare and *de novo* mutations (DNMs), is transforming the pace of genetics of neuropsychiatric disorder by identifying protein-coding mutations that confer risk. Despite previous studies that have analyzed the biological and clinical implications of protein truncating mutations identified from next-generation sequencing,[1,2] the risk of DNMs involved in transcriptional and post-transcriptional regulation in these disorders remain obscure.

Various computational methods have been developed to predict the effects of amino acid substitutions on protein function and to classify corresponding mutations as deleterious or benign. The majority of these approaches rely on evolutionary conservation or protein structural constraints of amino acid substitutions and focus on direct changes in protein coding genes, especially from nonsynonymous mutations which directly affect the gene product[3]. However, genetic mutations may not only impact the protein structure or catalytic activity but may also influence transcription[4,5] via direct or indirect effects on DNA-protein binding[6], histone modifications[7], enhancer cis-regulation[8] and gene-enhancer interactions[9]. Meanwhile, genetic mutations could impair post-transcriptional processes[10,11] such as interactions with RNA binding proteins (RBPs), mRNA splicing, mRNA transport, mRNA stability and mRNA translation.

Hitherto, previous studies of neuropsychiatric disorders have focused on genetic mutation in coding region[12], cis-regulation[13-15], epigenome[16], transcriptome[17,18], and proteome[19,20], but less is pursued with regard to how these disorders interface with post-transcriptional regulation. Therefore, potential mechanisms of post-transcriptional dysregulation related to pathology and clinical treatment of neuropsychiatric disorders remain largely uninvestigated. Our previous method named RBP-Var[21] and a recent published platform called POSTAR[22] represents a seminal effort to systematically annotate post-transcriptional regulatory maps and explore the putative roles of RBP-mediated SNVs in human diseases.

To explore association between extremely post-transcriptionally impaired DNMs, namely piDNMs, with psychiatric disorders, we collected whole exome and genome sequencing data from 9,772 core family and curated 42,871 *de novo* mutations from six kinds of neuropsychiatric disorders, including autism spectrum disorder (ASD), epileptic encephalopathy (EE), intellectual disability (ID), schizophrenia (SCZ), brain development defect (DD) and neurodevelopmental defect (NDD) as well as unaffected controls including siblings. By employing our previously developed workflow RBP-Var, we investigated the potential implication of these *de novo* mutations involved in post-transcriptional regulation in these six neuropsychiatric disorders and found that a subset of *de novo* mutations could classified as piDNMs. Moreover, we observed a higher prevalence of piDNMs in the probands of neuropsychiatric disorders versus controls. In addition, gene ontology (GO) enrichment analyses of cross-disorder genes containing piDNMs in at least two of these disorders revealed several shared crucial pathways involved in the biological processes of neurodevelopment. Meanwhile, weighted gene co-expression network analysis (WGCNA) and protein-protein interactions analysis showed enriched co-expression modules and intensive protein-protein interactions between these cross-disorder genes, respectively, implying that there was convergent machinery of post-transcriptional regulation in six psychiatric disorders. Furthermore, we established an interaction network which is centered on several

RBP hubs and encompassed with piDNMs-containing genes. In short, DNMs which are deleterious to post-transcriptional regulation highly likely contribute to the pathogenesis of neuropsychiatric disorders.

Results

Our ultimate goal was to identify all piDNMs from DNM in six neuropsychiatric disorders by analyzing a variety of aspects related to post-transcriptional regulation process. To do this, we first developed a comprehensive workflow with a statistic algorithm that links the potentially functional roles in post-transcriptional regulation to DNMs. We then used this workflow to identify functional variants from 9,772 core family with 42,871 *de novo* mutations associated with six neuropsychiatric disorders, including autism spectrum disorder (ASD), epileptic encephalopathy (EE), intellectual disability (ID), schizophrenia (SCZ), brain development defect (DD) and neurodevelopmental defect (NDD) as well as unaffected siblings (Supplementary Table 1). We further presented an online tool which could rapidly annotates and classifies piDNMs, and determines their potential roles in neuropsychiatric disorders.

High occurrence of piDNM in neuropsychiatric disorders

First, we used our previously developed workflow RBP-Var[21] to identify functional piDNMs from 9,772 trios with 42,871 de novo mutations (29,041 DNMs in probands and 13,830 DNMs in controls) across six kinds of neuropsychiatric disorders as well as their unaffected siblings (control). We determined DNMs with 1/2 categories predicted by RBP-Var as piDNMs and identified 2,381 piDNMs in probands (Supplementary Table 2). In detail, RBP-Var identified 1,410 piDNMs in ASD, 356 piDNMs in DD, 281 piDNMs in SCZ, 103 piDNMs in EE, 129 piDNMs in NDD and 102 piDNMs in ID while only 398 piDNMs were identified in controls groups (Supplementary Table 3). Interestingly, the average frequency of piDNMs in each patient in any one of six neuropsychiatric disorders were significantly over-represented compared with the one in controls (*P* = 9.52E-17, Figure 1A). With the combined data from 9,772 trios, we also observed that probands groups have much higher odds ratio (OR) of piDNMs compared with controls in all six kinds of neuropsychiatric disorders (Figure 1B; Supplementary Table 4). Dramatically, synonymous SNVs and unknown SNVs (SNVs which located in UTRs) in piDNMs were more prevalent in neuropsychiatric disorders versus controls (Supplementary Figure 1A-B), suggesting that synonymous SNVs and UTR SNVs which do not directly change the protein composition may also play crucial roles in the pathogenesis of neuropsychiatric disorders by influencing post-transcriptional regulation. Indeed, more than 50 diseases have been reported to be associated with synonymous SNVs in human genes[23]. In short, piDNMs that are linked to post-transcriptional dysregulation may contribute to the pathogenesis of these six neuropsychiatric disorders.

Robust discrimination of risk by piDNMs

Our platform RBP-Var determined the post-transcriptional dysfunction of piDNMs regardless of considering whether the DNMs give rise to protein truncating. To investigate the potential role of DNMs in different regulatory processes, we compared our RBP-Var[21] prediction tool with other three popular tools including SIFT[25,], PolyPhen2 (PPH2)[26] and RegulomeDB[24]. RBP-Var, RegulomeDB and SIFT/PPH2 were designed to evaluate the impact of mutation on post-transcriptional regulation, transcriptional regulation and protein product, respectively. We found that SIFT and PPH2 discriminated nonsynonymous/stop DNMs (*P* = 5.37E-62, *P* = 4.55E-03) and nonsynonymous/synonymous DNMs (*P* = 8.76E-123, *P* < 2.20E-16), respectively (Figure 2A, B). However, RegulomeDB discriminated nonframeshift (*P* < 2.20E-16), nonsynonymous (*P* = 1.12E-39), splicing (*P* = 7.62E-04), stop (*P* < 2.20E-16) and synonymous (*P* = 3.21E-22) DNMs rather than frameshift DNMs (*P* = 4.14E-01) (Figure 2C). Moreover, RBP-Var could discriminate the deleterious DNMs on effect of frameshift (*P* = 1.38E-03), nonsynonymous (*P* = 6.31E-60), splicing (*P* < 2.20E-16) and synonymous (*P*=8.98E-09) (Figure 2D).

Our results indicated that both RegulomeDB and RBP-Var better distinguish deleterious DNMs from benign ones compared to SIFT and PPH2. However, the number of deleterious DNMs detected by RegulomeDB is much less than that of RBP-Var, suggesting that RegulomeDB has a higher false-negative rate than RBP-Var. Therefore, we performed receiver operating characteristic (ROC) analysis to systemically evaluate the sensitivity and specificity of these four prediction methods. We found that area under curve (AUC) value of SIFT, PPH2, RBP-Var and RegulomeDB are 78.27 %, 76.57 %, 82.89 % and 50.77 %, respectively (Supplementary Figure 2). Moreover, the AUC value of SIFT, PPH2, and RBP-Var is significantly more sensitive and specific higher than that of RegulomeDB with *P* value 1.63E-10, 2.40E-08 and 2.51E-60, respectively. In addition, while the AUC value of AUC value of RBP-Var is significantly higher than that of SIFT and PPH2 with *p* value 0.049 and 0.019, respectively. Consequently, we compared the deleterious DNMs detected by these four methods and found that RBP-Var detected more piDNMs that overlapped with predictions of SIFT and PPH2 than that of RegulomeDB. In contrast, RBP-Var could detect an additional 1,234 piDNMs that were regarded as benign DNMs by other three methods, accounting for 37.29% of total 3,309 deleterious DNMs detected by all four tools (Supplementary Figure 3). In terms of genes related to deleterious DNMs, 909 functional genes (Supplementary Figure 4) which were determined by RBP-Var specially and the gene ontology were enriched in some neurodevelopmental pathways such as Notch signaling pathway (*P* = 9.82E-4), Neural Crest Differentiation (*P* = 0.0044), ErbB Signaling Pathway (*P* = 0.0048) and Signaling Pathways in Glioblastoma (*P* = 7.48E-4) (Supplementary Figure 5; Supplementary Table 5). Whereas, the functional genes detected specifically by RBP-Var in controls were not enriched in any gene ontology. Therefore, the piDNMs detected by RBP-Var are distinct and may play significant roles in the post-transcriptional processes of the development of neuropsychiatric disorders.

Distinct features of piDNM between disorders

Our previous study using the NPdenovo database demonstrated that *de novo* mutations predicted as harmful in protein level are shared by four neuropsychiatric disorders[25]. Therefore, we wondered whether there were common piDNMs among six neuropsychiatric disorders. Firstly, we identified 24 genes harboring recurrent piDNMs among 2,381 piDNMs, including nine genes from ASD trios, 10 genes from DD&NDD trios, six genes from ID trios, and one genes from EE trios (Figure 3A; Supplementary Table 6). Only two genes harbored cross-disorder recurrent piDNMs. One was STXBP1 which occurred in both DD and EE, and another was CTNNB1 which occurred in both ASD and DD. However, we identified 312 genes carrying at least two piDNMs among 1,923 candidate genes (accounting for 16.22%), including 252 genes from ASD trios, 38 genes from EE trios, 40 ID trios, 68 SCZ trios and 132 genes from DD&NDD trios by RBP-Var (Figure 3B; Supplementary Table 7).

Subsequently, we compared the genes harboring piDNMs across six disorders and found 214 genes significantly shared by at least two kinds of disorders rather than random overlaps (permutation test, *P* < 1.00E-5 based on random resampling, Figure 3C, D). Furthermore, the genes harboring piDNMs in controls also have significantly fewer overlaps with the 214 shared genes in the six disorders than random overlaps (*P* = 0.007, Supplementary Figure 6A).

In addition, the number of shared genes for any pairwise comparison is significantly higher than random overlaps ($P < 0.05$) except for the comparison of EE versus SCZ ($P = 0.17864$) (Supplementary Figure 7). On the contrary, the genes harboring piDNMs in controls have significantly fewer overlaps with disorders than random overlaps: ASD ($P <$ 1.00E-5, Supplementary Figure 6B), ID ($P <$ 1.00E-5, Supplementary Figure 6C), EE ($P =$ 1.20E-4, Supplementary Figure 6D), SCZ ($P <$ 1.00E-5, Supplementary Figure 6E) and DD&NDD ($P <$ 1.00E-5, Supplementary Figure 6F). Our observation suggests that there are common genes harboring piDNMs in these six neuropsychiatric disorders.

Several neuropsychiatric disorders share many symptoms suggesting common underlying molecular mechanism exist that may implicate important regulators of pathogenesis. Thus, we performed clueGO analysis for these shared genes and found they were enriched in crucial neurodevelopmental pathways (Supplementary Table 8; Supplementary Figure 8), including thyroid hormone signaling pathway (q = 1.10E-5), modulation of synaptic transmission (q = 4.50E-5), motor activity (q = 6.10E-5), dopamine receptor binding (q = 1.80E-4), visual learning (q = 2.00E-4), cognition (q = 3.00E-4), learning or memory (q = 2.00E-3), postsynaptic density (q = 2.40E-03), regulation of GABAergic synaptic transmission (q = 5.90E-03), regulation of synaptic plasticity (q = 1.00E-2), regulation of axonogenesis (q = 2.20E-2), cell cortex (q = 2.20E-2) and regulation of oligodendrocyte differentiation (q = 4.80E-2). Intriguingly, we found some enriched epigenetic pathways, such as epigenetic regulation of gene expression (q = 3.50E-8), histone methylation (q = 4.60E-4), histone lysine methylation (q = 3.7E-3), macromolecule methylation (q = 1.10E-2), chromatin silencing (q = 1.10E-2), chromatin remodeling (q = 1.20E-2), regulation of chromatin organization (q = 1.40E-2), chromatin-mediated maintenance of transcription (q = 1.60E-2). These epigenetic regulating genes are composed of ARID1B (P_{TADA} = 2.29E-8), CTCF (P_{TADA} = 2.29E-8), CTNNB1 (P_{TADA} = 2.29E-8), EP300 (P_{TADA} = 2.29E-8), MECP2 (P_{TADA} = 4.58E-7), DNMT3A (P_{TADA} = 7.33E-7), KMT2D (P_{TADA} = 1.86E-4), MYO1E (P_{TADA} = 8.80E-4), CNOT1 (P_{TADA} = 6.05E-3), TNRC18 (P_{TADA} = 6.80E-3), JARID2 (P_{TADA} = 1.70E-2), PHF19 (P_{TADA} = 7.54E-2), TNRC6A (P_{TADA} = 9.31E-2), DOT1L (P_{TADA} = 1.68E-1), KMT2B (P_{TADA} = 5.64E-1) (Supplementary Table 9). Moreover, most of these epigenetic modification genes, especially for genes with $P_{\mathrm{TADA}} < 0.05$, have been previously linked with neuropsychiatric disorders[26-29]. Interestingly, shared genes in clueGO enrichment analyses are intensive linkages among these significant pathways as some of piDNMs-containing genes could play roles in more than one of these pathways (Supplementary Figure 9).

In contrast, we also observed hundreds of piDNM-containing genes are disorder-specific (Figure 3C). Follow-up GO enrichment analysis revealed that 226, four and eight GO terms were enriched for ASD, SCZ and DD&NDD from these disorder-specific genes (q < 0.05), respectively. Several GO terms are associated with neurodevelopmental defects. For example, GO terms “neurogenesis” (q = 6.70E-03) and “neuron development” (q = 2.70E-02) were enriched for ASD-specific genes (Supplementary Table 10). And GO terms “regulation of autophagy” (3.00E-03), “regulation of macroautophagy” (3.40E-03), “macroautophagy” (1.60E-02) and “regulation of establishment of planar polarity involved in neural tube closure” (1.80E-02) were enriched in SCZ-specific genes (Supplementary Table 11). In addition, GO terms “response to UV” (q = 2.20E-3), “regulation of chromatin organization” (q = 8.50E-3), “helicase activity” (q = 1.00E-2), “histone methyltransferase complex” (q = 1.40E-02), “cellular response to UV” (q = 1.70E-

02), “viral genome replication” (q = 1.90E-02) and “regulation of histone acetylation” (q = 2.30E-02) were enriched in DD&NDD specific genes (Supplementary Table 12). However, there was no enriched GO terms for ID and EE specific genes. These disorder-specific genes may point to disorder-specific phenotypes and distinctive biological processes among the six neuropsychiatric disorders.

Enriched co-expression modules

As co-expression of genes has been used to explore the common and distinct molecular mechanism in neuropsychiatric disorders[30], we performed weighted gene co-expression network analysis (WGCNA)[31] with 214 cross-disorder transcripts by using up-to-date BrainSpan developmental transcriptome (n=524): gene expression in 16 human brain structures across 31 developmental stages[32]. The results of WGCNA deciphered two gene modules with distinct spatiotemporal expression patterns (Figure 4A, B; Supplementary Figure 10). The turquoise module (n=146 genes) is characterized by high expression during early fetal development (8-37 postconceptional weeks) in most of brain structures (Figure 4C). The blue module (n=58 genes) is delineated to low expression in early fetal development (postconceptional week 8) and early childhood (3 year) in most of brain structures (Figure 4D). Only two out of the 214 cross-disorder genes were not clustered in the co-expression networks (Supplementary Table 13). In addition, we constructed the co-expression network by using gene pairs with correlation weight of at least 0.3 (Supplementary Table 14). Interestingly, we found that majority of the co-expression hubs were histone modifiers, such as KAT6B, KDM5B, SETDB1, TRRAP, EP300, and transcriptional regulators, including CTCF, SMARCA4, ARID1B, ADNP, CHAMP1 (Supplementary Figure 11). Consequently, the link among these six highly heterogeneous neuropsychiatric disorders may be represented by biological processes controlled by these hub genes in co-expression networks.

Furthermore, GO enrichment analysis revealed distinct biological processes within the two modules. The turquoise module was enriched for epigenetic regulation of gene expression (q = 3.99E-08), establishment or maintenance of cell polarity (q = 8.24E-04), Notch binding (q = 1.33E-3), Thyroid hormone signaling pathway (q = 1.75E-3), beta-catenin-TCF complex assembly (q = 5.81E-3), chromatin remodeling (q = 1.30E-2), Notch signaling pathway (q = 2.51E-2), motor activity (q = 3.21E-2) and other enriched GO terms (Supplementary Figure 12; Supplementary Table 15). The blue module was enriched visual learning (q = 4.92E-07), regulation of synaptic plasticity (q = 1.91E-06), actin filament-based movement (q = 6.52E-06), adult locomotory behavior (q = 3.86E-04), neuromuscular process (q = 8.61E-04), Long-term depression (q = 1.02E-3), positive regulation of dendrite development (q = 1.35E-3), negative regulation of autophagy (q = 3.74E-3), negative regulation of G-protein coupled receptor protein signaling pathway (q = 5.07E-3) and other neurodevelopment related pathways (Supplementary Figure 13; Supplementary Table 16).

Intensive protein-protein interactions

The co-expression results indicated that these 214 cross-disorder proteins may have intensive protein-protein interactions (PPI). To identify common biological processes that potentially contribute to disease pathogenesis, we established protein-protein interactions within these 214 cross-disorder genes by investigating protein-protein interactions curated in database BioGRID. Surprisingly, 128 out of 214 (59.81%) cross-disorder genes represent an interconnected network on the level of available direct/genetic protein-protein interactions (Figure 5A). In addition, an overwhelming majority (98.54%, 203 out of 206 pairs) of the protein-protein interactions constitute direct interactions while only 3 interaction pairs are genetic interactions (Supplementary Table 17). Furthermore, we determined several crucial hubs of protein-protein interactions, including CUL3, CTNNB1, HECW2, HNRNPU, EP300, SMARCA4, ARRB2, PTPRF, MYH9 and NOTCH1 (Figure 5A), which may control common biological processes among these six neuropsychiatric disorders. Indeed, these 128 cross-disorder proteins are enriched in nervous system phenotypes, including abnormal synaptic transmission, abnormal nervous system development, abnormal neuron morphology and abnormal brain morphology, and behavior/neurological phenotype such as abnormal motor coordination/balance (two sided Fisher's Exact Test, $q < 0.05$, Supplementary Figure 14A). Similarly, these 128 genes in interaction network are enriched in nervous system phenotype including abnormal nervous system development and abnormal brain morphology (two sided Fisher's Exact Test, $q < 0.05$, Supplementary Figure 14B). Dramatically, HNRNPU belongs to the subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs) which are RNA binding proteins and they form complexes with heterogeneous nuclear RNA (hnRNA). These proteins are associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport.

To investigate the expression of these co-interacting genes in the human cortex of ASD patients and normal donators, we downloaded and compared the normalized FPKM value from Gene Expression Omnibus (GEO) with accession number GSE64018 and GSE76852, respectively. We identified 97 significantly differential expression genes (accounting for 75.78% of the PPI genes) between ASD patients and normal controls (Student's t-test, $q < 0.05$, Figure 5B). And 86 of these PPI genes were highly expressed in ASD patients while only 11 genes were down-regulated in ASD patients when compared with normal controls (Supplementary Table 18), implying that most of these PPI genes were abnormally expressed in ASD patients.

Shared networks between piDNM and RBPs

According to previous studies, dysregulation or mutations of RBPs can cause a range of developmental and neurological diseases[33-35]. By contrast, mutations in RNA targets of RBPs, which could disturb the interaction between RBPs and their mRNA targets, would affect mRNA metabolism and protein homeostasis in neurons by impair RNA transport and translation in neuropathological disorders[36-38]. Indeed, mutations that alter sequence context of RBP binding sites commonly affect RBP binding and regulation[39]. For instance, regulation depended on FMRP-related activity during fetal brain development might be particularly vulnerable to genetic perturbations, with severe mutations resulting in disruption of developmental canalization[40]. Hence, we constructed a regulatory network between piDNMs and RBPs to investigate the genetic perturbations of mRNA-RBP interactions in six disorders (Figure 6). We identified several crucial common RBP hubs, including EIF4A3, FMR1, PTBP1, AGO1/2, ELAVL1, IGF2BP1/3, WDR33 and FXR2. Genes with piDNM in different disorders could be regulated by the same RBP hub while one candidate genes may be regulated by different RBP hubs (Figure 6). In addition, all of these RBP hubs were highly expressed in early fetal development stages (8-37 postconceptional weeks) based on BrainSpan developmental transcriptome (Supplementary Figure 15), suggesting that these RBP hubs may play important roles in the early stages of brain development.

The PostOmics database

To make our findings easily accessible to the research community, we have developed RBP-Var2 platform (http://www.rbp-var.biols.ac.cn/) for storage and retrieval of piDNMs, candidate genes, for exploring the genetic etiology of neuropsychiatric disorders in post-transcriptional regulation.

Discussion

In this study, we systematically analyzed the damaging effect of *de novo* mutations at the level of post-transcriptional regulation in six neuropsychiatry disorders. We discovered 2,381 piDNMs that were extremely deleterious on post-transcriptional regulation, of which 1,034 (43.43%) piDNMs were extremely deleterious (predicted by NPdenovo) on the structure and function of protein. Among the total 1,923 genes containing piDNMs, 862 (44.83%) genes were predicted to simultaneously alter the structure and function of correspondingly coded proteins. On the other hand, 1007 out of 2672 (37.69%) genes and 1023 of 3714 (27.54%) extreme DNMs predicted by NPdenovo were also extremely deleterious predicted by RBP-Var. Our observations indicate that over half of piDNMs are not expect to affect the structure and function of protein but are predicted to affect post-transcriptional regulation, and vice versa. Moreover, probands have higher odds ratio of piDNMs than controls in all kinds of neuropsychiatric disorders. Therefore, piDNMs are one of major contributors to etiology for neuropsychiatry disorders.

We applied RBP-Var algorithm to annotate and interpret *de novo* variants in subjects with six neuropsychiatry disorders and control subjects. We observed a strong enrichment of pathogenic or likely pathogenic variants in affected subjects. In comparison with accuracy of RBP-Var (82.89%), other prediction algorithms such as SIFT or PPH2 have comparative accuracy (76~78%) while RegulomeDB has much lower accuracy (50.77 %) to differentiate affected from control subjects. This observation suggests that the potential pathogenicity of genetic variants is mainly due to post-transcriptional dysregulation and malfunction of protein, but not disruption of transcription process. Therefore, deciphering multiple biological layers of deleteriousness, particularly post-transcriptional regulation and structure/function of protein, may improve the accuracy of algorithm's prediction.

To date, it has been a challenge to estimate the deleteriousness of synonymous and UTRs mutations though such mutations have been widely acknowledged alter protein expression, conformation and function[41]. Nevertheless, our data-based algorithm identified 518 synonymous DNMs and 82 UTR's DNMs were extremely deleterious in post-transcriptional regulation. Moreover, synonymous damaging DNMs were significantly prominent in probands compared with that of controls, supported by predictions of both RBP-Var and RegulomeDB. Meanwhile, *de novo* insertions and deletions (InDels), especially frameshift patterns are taken for granted to be deleterious. Indeed, *de novo* frameshift InDels are more frequent in neuropsychiatry disorders compared to non-frameshift InDels[42], which were demonstrated by predictions of RBP-Var but not SIFT and PPH2 (Figure 2). To our knowledge, our research is the first investigation into the post-transcriptional etiology of *de novo* non-frameshift InDels and *de novo* synonymous mutations as well as *de novo* protein-truncating variants in neuropsychiatry disorders.

Most interestingly, we discovered some enriched epigenetic pathways, such as those that regulate gene expression and histone methylation. This finding is consistent with a previous report that demonstrated that more than 68% of ASD cases shared a common acetylome aberrations at >5,000 cis-regulatory regions in prefrontal and temporal cortex[43]. Such common "epimutations" may be induced by either perturbations of epigenetic regulations (including post-transcriptional regulations) due to genetic mutations of substrates or the disruptions of epigenetic modifications due to the genetic mutation of epigenetic genes. Actually, our observations suggest that alterations of "epimutations" were associated with the dysregulation of post-transcription. In fact, we found several recurrent piDNM-containing genes including CHD8, CHD2, SYNGAP1, ADNP, POGZ, ANK2 and DYRK1A, are non-epigenetic genes. Moreover, we also discovered several recurrent epigenetic genes including KMT2A, KMT2B, KMT2C, KAT6B, KDM3B, JARID2, DNMT3A and MECP2, that contain piDNMs which may play important roles in the genome-wide aberrations of epigenetic landscapes through disruption the post-transcription regulation. Furthermore, WGCNA analysis revealed that major hubs of the co-expression network for these 214 piDNM-containing genes were histone modifiers by using BrainSpan developmental transcriptome. These data indicate that piDNM-containing genes are co-expressed with genes frequently involved in epigenetic regulation of common cellular and biological process in neuropsychiatric disorders.

Importantly, these 214 piDNMs-containing genes harbor intensive protein-protein interactions in physics and shared regulatory networks between piDNMs and RBPs in six neuropsychiatric disorders. We identified several RBP hubs of regulatory networks between piDNMs-containing genes and RBP proteins, including EIF4A3, FMRP, PTBP1, AGO1/2, ELAVL1, IGF2BP1/3, WDR33 and FXR2. Taking FMRP for example, it is a well-known pathogenic gene of Fragile X syndrome which co-occurs with autism in many cases and its targets are highly enriched for de novo mutations in ASD[40]. The mutation of any RBP hubs may result in multiple disorders and mutations of regulated genes may disrupt interactions with multiple RBPs. Fortunately, we have constructed the regulatory networks among multiple RBPs and their target RNAs with *de novo* mutations that could significantly disturb the regulatory networks in six neuropsychiatric disorders. Thus, we provided a framework for interpreting network biology research and leveraging big genomics data sets in neurobiology of post-transcriptional regulation.

Alterations in expression or mutations in either RBPs or their binding sites in target transcripts have been reported to be the cause of several human diseases such as muscular atrophies, neurological disorders and cancer[44]. However, it is still a challenge to decipher the effect of genetic mutation on RNA and post-transcriptional regulation and additional clinical and molecular datasets are required to improve the sensitivity and specificity of our data and algorithm based platform. However, our method sheds light on evaluation of post-transcriptional impact of genetic mutations including synonymous mutations which frequently act as driver mutations in human cancers[45]. In addition, small molecules can be rapidly designed to selectively target RNAs and affect RNA-RBP interactions, leading to anti-cancer strategies[46]. Therefore, the discovery of disease-causing mutations in RNAs is yielding a wealth of new therapeutic targets, and the growing understanding of RNA biology and chemistry is providing new RNA-based tools for developing therapeutics[47].

Materials and methods

Data collection.

For this study, 9,772 trios were recruited from previous WES/WGS studies, comprising 7,453 parent-probands trios associated with six neuropsychiatric disorders and 2,319 control trios. After deduplication, a total of 29,041 DNMs in probands and 13,830 DNMs in controls were identified for subsequent analysis.

Identification of deleterious piDNMs and comparison with variants predicted by other methods.

To determine the likelihood of causing a deleterious mutation in post-transcriptional regulation for all SNVs and InDels, our previously developed program RBP-Var[48] was utilized to assign an exclusive rank for each mutation and only those mutations categorized into rank 1 or 2 were considered as piDNMs. In comparison with those mutations involved in the disruption of gene function or transcriptional regulation, several programs such as SIFT, PolyPhen2 and RegulomeDB were used to analyze the same dataset of DNMs as the input for RBP-Var. We only kept the mutations qualified as “damaging” from the result of SIFT and “possibly damaging” or “probably damaging” from PolyPhen2. In the case of RegulomeDB, mutations labelled as category 1 and 2 were retained. Next, we classified the type of mutation (frameshift, nonframeshift, nonsynonymous, synonymous, splicing and stop) and located regions (UTR3, UTR5, exonic, ncRNA exonic and splicing) in order to determine distribution of piDNMs, genetic variants and other regulatory variants. The number of variants in cases versus controls was illustrated by bar chart (***: $P < 0.001$, **: $0.001 < P < 0.01$, *: $0.01 < P < 0.05$, binomial test).

TADA analysis of DNMs in six disorders

The TADA program (Transmission And De novo Association), which predicts risk genes accurately on the basis of allele frequencies, gene-specific penetrance, and mutation rate, was used to calculate the P-value for the likelihood of each gene contributing to the all six disorders with default parameters.

ROC curves and specificity/sensitivity estimation

We screened a positive (non-neutral) test set of likely casual mutations in Mendelian disease from the ClinVar database (v20170130). From a total of 237,308 mutations in ClinVar database, we picked up 145 exonic mutations presenting in our curated DNMs in probands. Our negative (neutral) set of likely non-casual variants was built from DNMs of unaffected siblings in six neuropsychiatric disorders. In order to avoid rare deleterious DNMs, we selected only DNMs in controls with a minor allele frequency of at least 0.01 in 1000 genome (1000g2014oct), resulting in a set of 921 exonic variants. Then, we employed R package pROC to analyze and compare ROC curves.

Overlaps of piDNMs between disorders.

In order to elucidate the overlap of genes among any two of the six disorders as well as each disorder and the control, we shuffled the intersections of genes and repeated this procedure 100,000 times. During each permutation, we randomly selected the same number of genes as the actual situation from the entire genome for each disorder and the control, then *p* values were calculated as the proportion of permutations during which the observed number of genes was above/below the number of genes in simulated situations.

Functional enrichment analysis

A gene harboring piDNMs would be selected into our candidate gene set to conduct functional enrichment analysis if it occurred in at least two of the six disorders. GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichments analysis was implemented by Cytoscape (version 3.4.0) plugin ClueGO (version 2.3.0) and *P* values calculated by hypergeometric test was corrected to be q values by Benjamini–Hochberg procedure for reducing the false discovery rate resulted from multiple hypothesis testing.

Co-expression and spatiotemporal specificity

Normalized gene-expression of 16 human brain regions were determined by RNA sequencing and obtained from BrainSpan (http://www.brainspan.org). Expression for 206 out of 214 extreme damaging cross-disorder genes was extracted and clustered according to their expression pattern by using R-package WGCNA (weighted correlation network analysis). The expression level for each gene and development stage (only stages with expression data for all 16 structures were selected, n = 14) was presented across all brain regions.

Protein-protein interaction and phenotype enrichment

Cross-disorder genes containing piDNMs were subjected to esyN analysis for generating a network of protein interactions and the network of protein interaction was created by using physical and genetic interactions of H. sapiens curated in BioGRID. Cytoscape (version 3.4.0) was used to analyze and visualize protein-protein interaction networks. Overrepresentation of mouse-mutant phenotypes was evaluated using the web tool MamPhea for the genes in the PPI network and for all cross-disorder genes containing piDNMs. Rest of genome was used as background and q values was calculated from *p* values by Benjamini–Hochberg correction.

Gene-RBP interaction network

Cytoscape (version3.4.0) was utilized for visualization of the associations between genes harboring piDNMs in the six neuropsychiatric disorders and the corresponding regulatory RBPs.

URLs

RBP-Var2, http://www.rbp-var.biols.ac.cn/ ; NPdenovo, http://www.wzgenomics.cn/NPdenovo/index.php; BioGRID, https://thebiogrid.org/;

MamPhea , http://evol.nhri.org.tw/phenome/index.jsp?platform=mmus; BrainSpan, http://www.brainspan.org; ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/; 1000Genomes, http://www.internationalgenome.org/; WGCNA, https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/;

esyN, http://www.esyn.org/; Cytoscape, http://www.cytoscape.org/; TADA, http://wpicr.wpic.pitt.edu/WPICCompGen/TADA/TADA_homepage.htm; ClueGO, http://apps.cytoscape.org/apps/cluego; pROC, http://web.expasy.org/pROC/; R, https://www.r-project.org/; Perl, https://www.perl.org/.

Acknowledgments

The project was funded by National High Technology Research and Development Program of China [2012AA02A202]; National Natural Science Foundation of China [31171236/C060503]; Innovation Center of China, AstraZeneca. We have to thank Kun Zhang for his help in TADA analysis.

Conflict of interest

The authors declare that there are no conflicts of interest regarding to publishing this paper.

Figure legends

Figure 1. The abundance of DNMs in different RBP-Var categories. (A) The quantities of DNMs in patients with neuropsychiatric disorders are significantly larger than that in normal controls, shown in several categories classified by RBP-Var. (B) The horizontal bar plot corresponds to the odds ratios indicating the enrichment of piDNMs in patients with any of the five neuropsychiatric disorders. The bar plot of normal controls corresponds to the odds ratio of 1 with no enrichment.

Figure 2. Performance comparison of the ability to distinguish severe DNMs between RBP-Var and three other tools. (A) Different kinds of DNMs affecting protein function predicted by SIFT. The Y-axis corresponds to the proportion of each kind of mutations within the total number of damaging DNMs predicted by SIFT. (B) Different kinds of DNMs that affect protein function predicted by PolyPhen2. The Y-axis corresponds to the proportion of each kind of mutations within the total number of damaging DNMs predicted by PolyPhen2. (C) The DNMs predicted as functional elements involved in transcriptional regulation by RegulomeDB are categorized into different functional types. The Y-axis corresponds to the proportion of each kind of mutations within the total number of damaging DNMs predicted by RegulomeDB. (D) The DNMs classified as either level 1 or 2 (piDNMs) are categorized into different functional types. The Y-axis corresponds to the proportion of each kind of mutations within the total number of damaging piDNMs. The P values were measured by two-sided binomial test.

Figure 3. Candidate genes selected by RBP-Var involved in six neuropsychiatric disorders. (A) Scatter plot of 24 genes harboring recurrent piDNMs among 2381 piDNMs. The Y-axis corresponds to the –log10(P value) calculated by TADA. The X-axis stands for the TADA output of –log10(mutation rate). (B) Scatter plot of 312 recurrent genes among 1,923 candidate genes. The Y-axis corresponds to the –

log10(P value) calculated by TADA. The X-axis stands for the TADA output of –log10(mutation rate). (C) Venn diagram representing the distribution of candidate genes shared among five neuropsychiatric disorders. (D) Permutation test for the validity of the overlap between the candidate genes involved in the five neuropsychiatric disorders. We shuffled the genes of each disorder and calculated the shared genes between the five disorders, repeating this procedure for 100,000 times to get the null distribution. The vertical dash line stands for the observed value corresponding to a P value of the permutation test.

Figure 4. Weighted co-expression analysis of 214 shared genes. (A) Heat map visualization of the co-expression network of 214 shared genes. The more saturated color corresponds to the highly expressed genes. (B) Hierarchical clustering dendrogram of the two color-coded gene modules displayed in (A). (C, D) Heat maps of the expression of the two respective gene modules.

Figure 5. Protein-protein interaction network analysis. (A) The network of interactions between pairs of proteins of 128 out 214 shared genes. (B) Heat map showing significantly differential expression of 97 out of 128 genes involved in the protein-protein interaction network.

Figure 6. Interaction network of RBPs and genes with piDNMs. Different roles of the nodes are reflected by distinguishable geometric shapes and colors. The magenta vertical arrow stands for the RNA binding proteins. Disks with different colors represent the genes with piDNMs involved in different kinds of disorders.

odds ratios of piDNMs in the forms of insertion that involved in the six neuropsychiatric disorders relative to the normal controls.

Supplementary Figure 1. Odds ratio of different functional types of piDNMs. (A-G) Bar plot of the odds ratios of piDNMs in various functional types involved in the six neuropsychiatric disorders relative to the normal controls.

Supplementary Figure 2. ROC curve showing the performance of the predictions of SIFT, PPH2, RBP-Var and RegulomeDB.

Supplementary Figure 3. Venn diagram depicting the overlap between the DNMs predicted by SIFT, PPH2, RBP-Var and RegulomeDB.

Supplementary Figure 4. Venn diagram depicting the overlap between the genes predicted by SIFT, PPH2, RBP-Var and RegulomeDB.

Supplementary Figure 5. Pathway enrichment analysis of the 909 genes unique to the prediction of RBP-Var.

Supplementary Figure 6. Test of the significance of the genes shared between each pair of the five disorders. (A-G) Permutation test for the validity of the gene overlap between each pair of the five disorders. We shuffled the genes of each disorder and calculated the shared genes between each pair, repeating this procedure for 100,000 times to get the null distribution. The vertical dash line stands for the observed value corresponding to a P value of the permutation test.

Supplementary Figure 7. Test of the significance of the 214 candidate genes involved in the five neuropsychiatric disorders. (A-E) Permutation test for the validity of the gene overlap between each disorder and the normal control. (F) Permutation test for the validity of the gene overlap between the 214 shared genes and the normal control.

Supplementary Figure 8. Pie chart of the pathway enrichment analysis for the 214 shared genes.

Supplementary Figure 9. Interaction network of the gene enrichment analysis for the 214 shared genes.

Supplementary Figure 10. Relationship between Co-expression modules. (A) MDS plot of genes in turquoise module and blue module. (B) Relationship between module eigengenes. (C) Clustering tree based of the module eigengenes. (D) heatmap of adjacency Eigengene.

Supplementary Figure 11. The co-expression network of 206 out of the 214 shared genes. Different size of the node is representative of the number of connections between the gene and others.

Supplementary Figure 12. Pie chart showing the enrichment analysis of the genes clustered in the turquoise module from the weighted co-expression analysis.

Supplementary Figure 13. Pie chart showing the enrichment analysis of the genes clustered in the blue module from the weighted co-expression analysis.

Supplementary Figure 14. Mammalian phenotype enrichment analysis of selected genes. (A) Mammalian phenotype enrichment of 214 cross-disorder piDNMs genes. (B) Mammalian phenotype enrichment of 128 genes in interaction network.

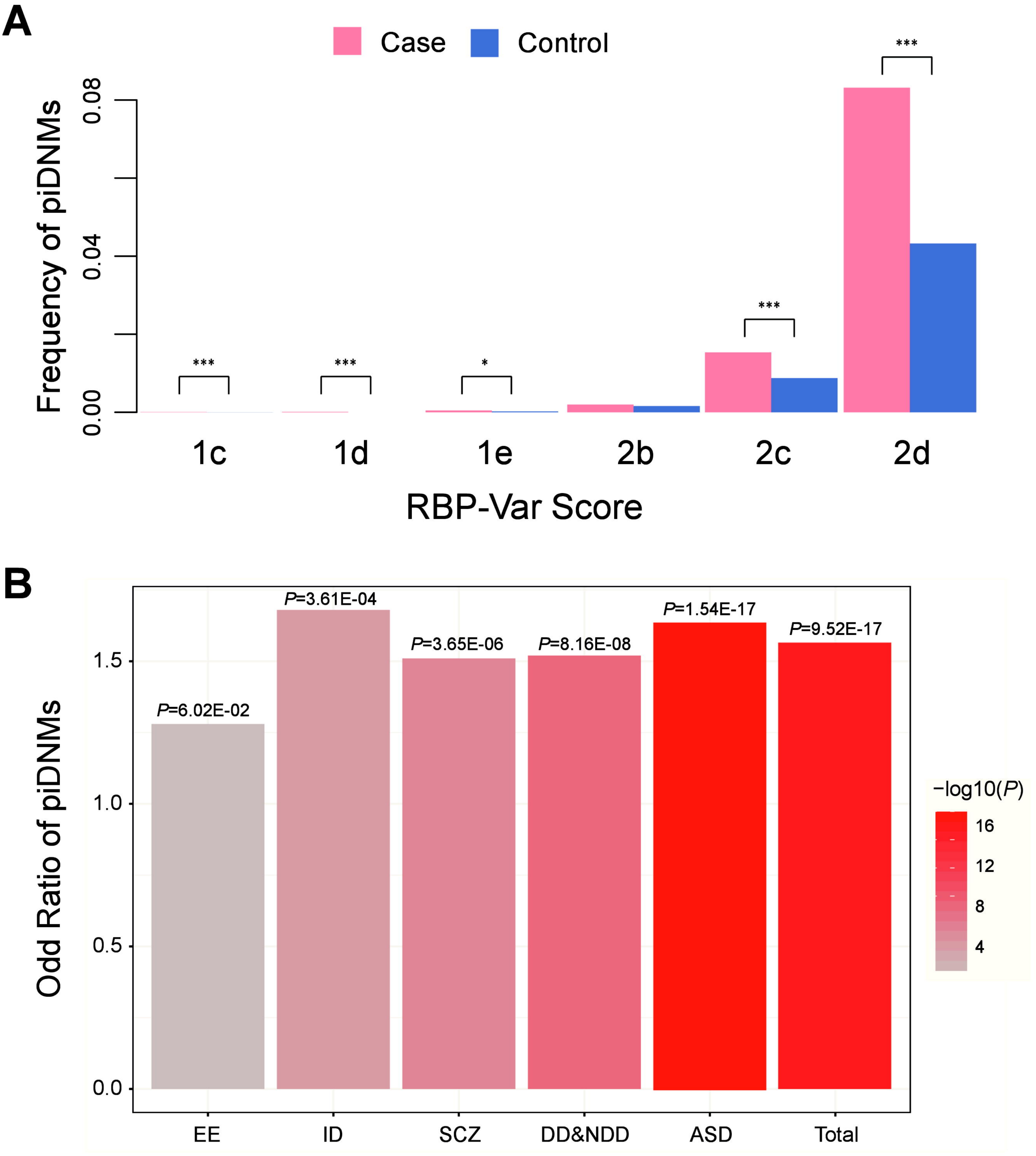
Supplementary Figure 15. Heat map of the expression of the crucial RBP hub genes during the early fetal development stages.

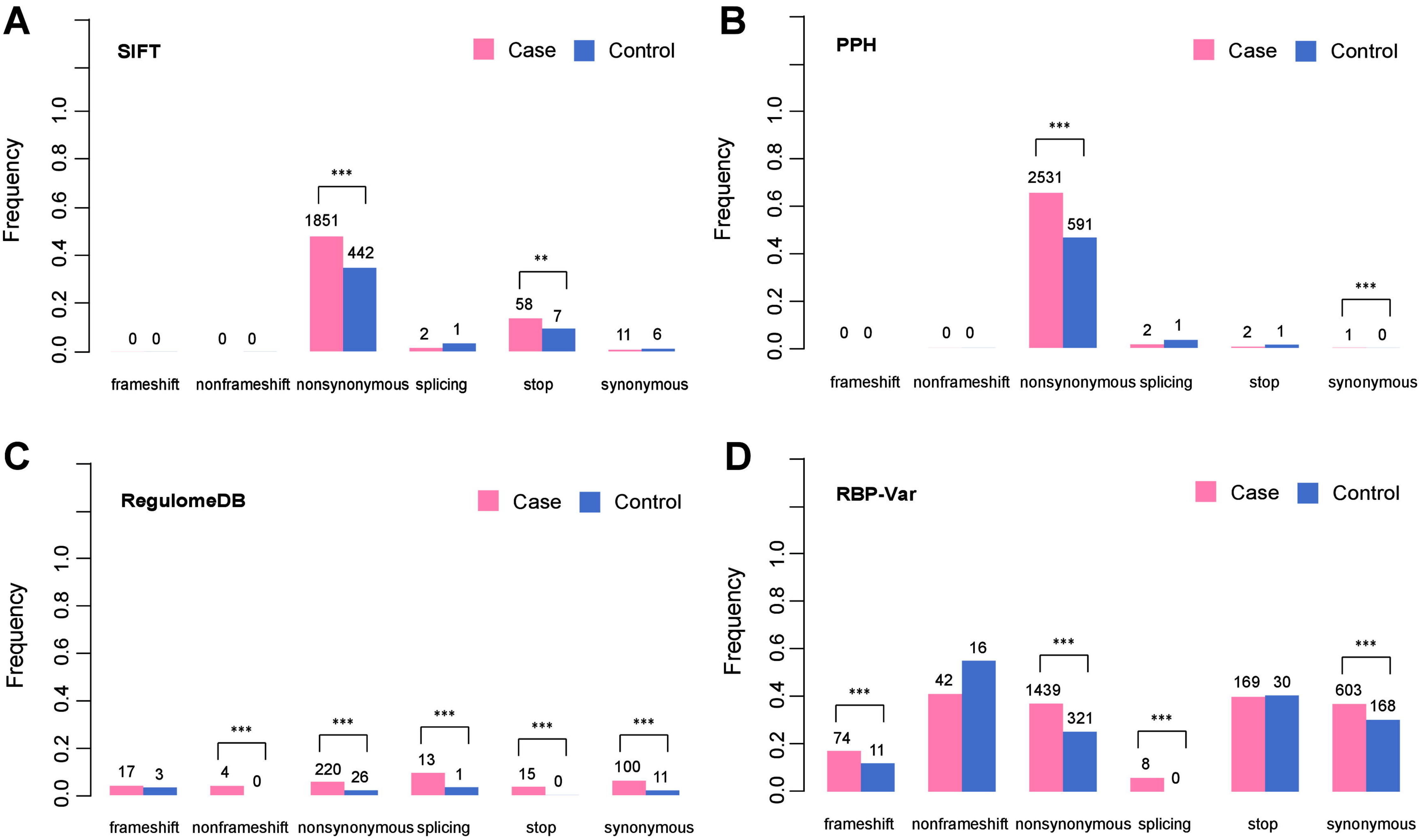
Reference

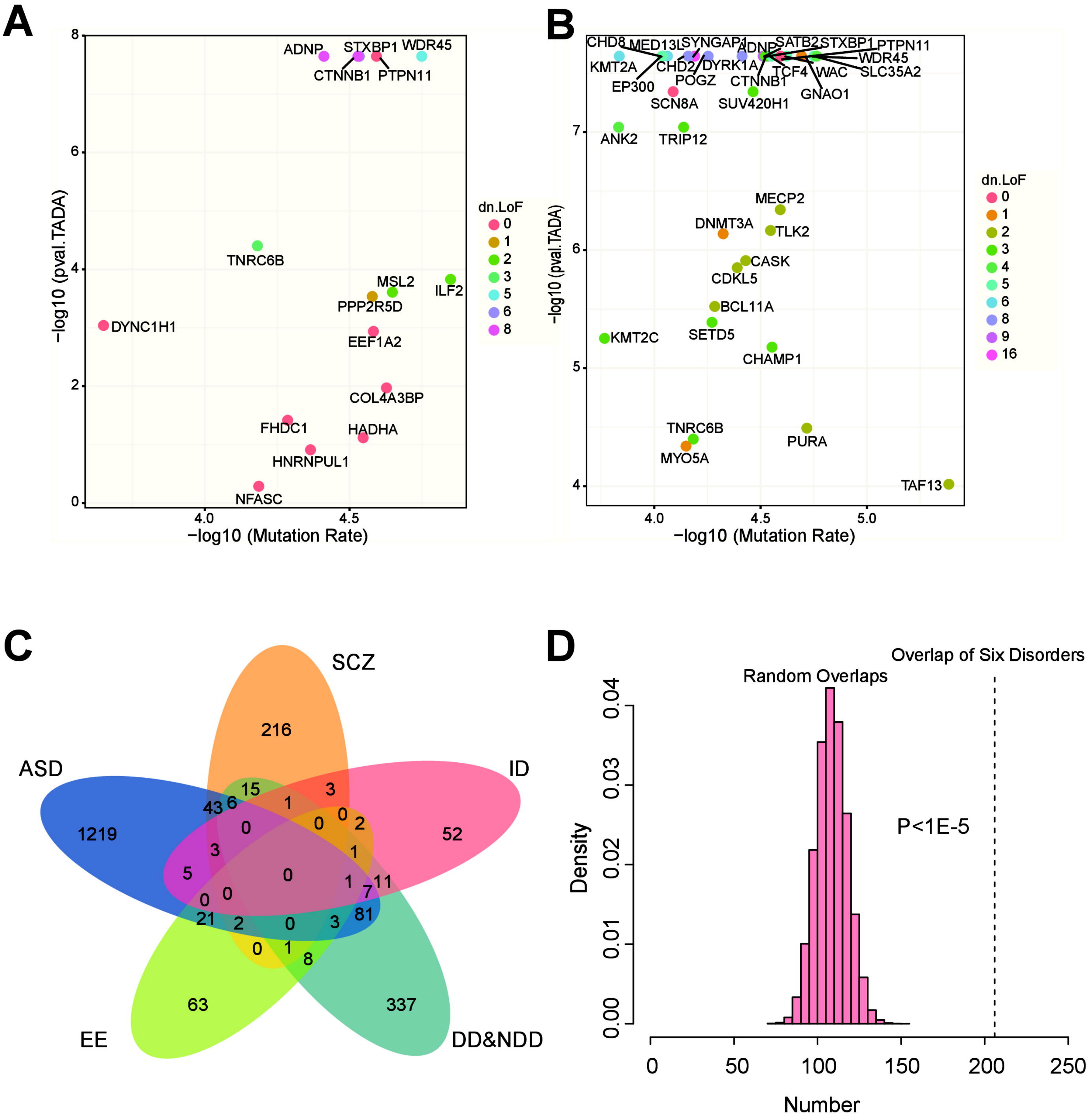
1. Sullivan, P.F., Daly, M.J. & O'Donovan, M. DISEASE MECHANISMS Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Reviews Genetics* **13**, 537-551 (2012).
2. Heinzen, E.L., Neale, B.M., Traynelis, S.F., Allen, A.S. & Goldstein, D.B. The Genetics of Neuropsychiatric Diseases: Looking In and Beyond the Exome. *Annual Review of Neuroscience, Vol 38* **38**, 47-68 (2015).
3. Gnad, F., Baucom, A., Mukhyala, K., Manning, G. & Zhang, Z.M. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *Bmc Genomics* **14**(2013).
4. Chen, X.F., Zhang, Y.W., Xu, H.X. & Bu, G.J. Transcriptional regulation and its misregulation in Alzheimer's disease. *Molecular Brain* **6**(2013).
5. Lee, T.I. & Young, R.A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **152**, 1237-1251 (2013).
6. Mathelier, A. *et al.* Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. *Genome Biology* **16**(2015).
7. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nature Biotechnology* **28**, 1057-1068 (2010).
8. Sakabe, N.J., Savic, D. & Nobrega, M.A. Transcriptional enhancers in development and disease. *Genome Biology* **13**(2012).
9. Lupianez, D.G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012-1025 (2015).
10. Linder, B., Fischer, U. & Gehring, N.H. mRNA metabolism and neuronal disease. *Febs Letters* **589**, 1598-1606 (2015).
11. Halbeisen, R.E., Galgano, A., Scherrer, T. & Gerber, A.P. Post-transcriptional gene regulation: From genome-wide studies to principles. *Cellular and Molecular Life Sciences* **65**, 798-813 (2008).
12. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-21 (2014).
13. Buonocore, F. *et al.* Effects of cis-regulatory variation differ across regions of the adult human brain. *Human Molecular Genetics* **19**, 4490-4496 (2010).
14. Gamazon, E.R. *et al.* Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Molecular Psychiatry* **18**, 340-346 (2013).
15. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* (2016).
16. Loke, Y.J., Hannan, A.J. & Craig, J.M. The Role of Epigenetic Change in Autism Spectrum Disorders. *Front Neurol* **6**, 107 (2015).
17. Gupta, S. *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* **5**, 5748 (2014).
18. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-4 (2011).
19. Patel, S. Role of proteomics in biomarker discovery: prognosis and diagnosis of neuropsychiatric disorders. *Adv Protein Chem Struct Biol* **94**, 39-75 (2014).
20. Corbett, B.A. *et al.* A proteomic study of serum from children with autism showing differential expression of apolipoproteins and complement proteins. *Mol Psychiatry* **12**, 292-306 (2007).

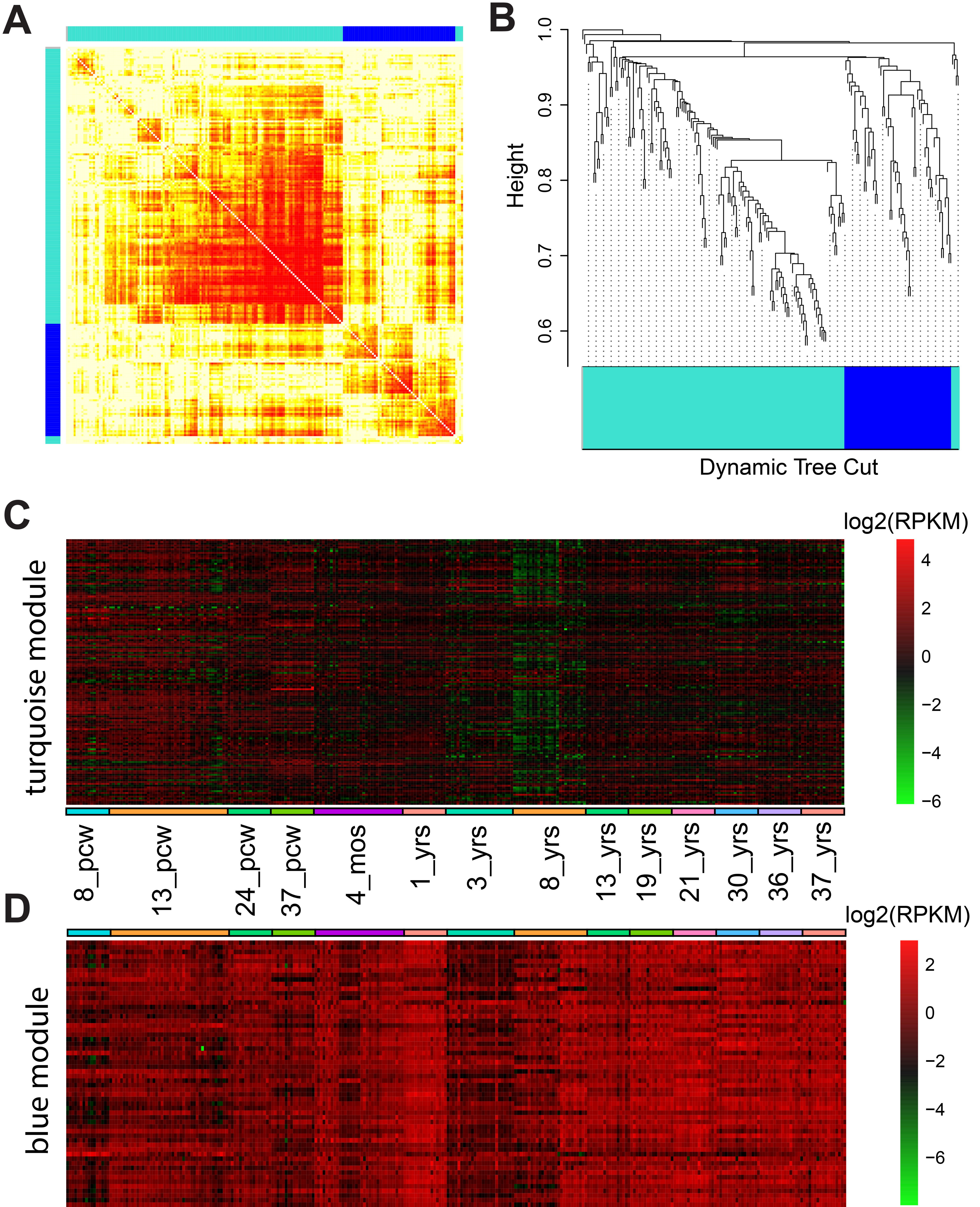
21. Mao, F.B. *et al.* RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Research* **44**, D154-D163 (2016).
22. Hu, B., Yang, Y.-C.T., Huang, Y., Zhu, Y. & Lu, Z.J. POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Research*, gkw888 (2016).
23. Wen, P., Xiao, P. & Xia, J. dbDSM: a manually curated database for deleterious synonymous mutations. *Bioinformatics* **32**, 1914-6 (2016).
24. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790-1797 (2012).
25. Li, J.C. *et al.* Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Molecular Psychiatry* **21**, 290-297 (2016).
26. Hsieh, J. & Eisch, A.J. Epigenetics, hippocampal neurogenesis, and neuropsychiatric disorders: Unraveling the genome to understand the mind. *Neurobiology of Disease* **39**, 73-84 (2010).
27. Brookes, E. & Shi, Y. Diverse Epigenetic Mechanisms of Human Disease. *Annual Review of Genetics, Vol 48* **48**, 237-268 (2014).
28. Jakovcevski, M. & Akbarian, S. Epigenetic mechanisms in neurological disease. *Nature Medicine* **18**, 1194-1204 (2012).
29. Peter, C.J. & Akbarian, S. Balancing histone methylation activities in psychiatric disorders. *Trends in Molecular Medicine* **17**, 372-379 (2011).
30. Lotan, A. *et al.* Neuroinformatic analyses of common and distinct genetic components associated with major neuropsychiatric disorders. *Frontiers in Neuroscience* **8**(2014).
31. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9**(2008).
32. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-9 (2011).
33. Castello, A., Fischer, B., Hentze, M.W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends in Genetics* **29**, 318-327 (2013).
34. Lenzken, S.C., Achsel, T., Carri, M.T. & Barabino, S.M.L. Neuronal RNA-binding proteins in health and disease. *Wiley Interdisciplinary Reviews-Rna* **5**, 565-576 (2014).
35. Klein, M.E., Monday, H. & Jordan, B.A. Proteostasis and RNA Binding Proteins in Synaptic Plasticity and in the Pathogenesis of Neuropsychiatric Disorders. *Neural Plasticity* (2016).
36. Lukong, K.E., Chang, K.W., Khandjian, E.W. & Richard, S. RNA-binding proteins in human genetic disease. *Trends in Genetics* **24**, 416-425 (2008).
37. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nature Reviews Genetics* **15**, 829-845 (2014).
38. Neelamraju, Y., Hashemikhabir, S. & Janga, S.C. The human RBPome: From genes and proteins to human disease. *Journal of Proteomics* **127**, 61-70 (2015).
39. Taliaferro, J.M. *et al.* RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Molecular Cell* **64**, 294-306 (2016).
40. Parikshak, N.N., Gandal, M.J. & Geschwind, D.H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nature Reviews Genetics* **16**, 441-458 (2015).
41. Sauna, Z.E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* **12**, 683-691 (2011).
42. Kosmicki, J.A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet* (2017).
43. Sun, W. *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397 e11 (2016).

44. Kechavarzi, B. & Janga, S.C. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biology* **15**(2014).
45. Supek, F., Minana, B., Valcarcel, J., Gabaldon, T. & Lehner, B. Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. *Cell* **156**, 1324-1335 (2014).
46. Costales, M.G. *et al.* Small Molecule Inhibition of microRNA-210 Reprograms an Oncogenic Hypoxic Circuit. *J Am Chem Soc* **139**, 3446-3455 (2017).
47. Cooper, T.A., Wan, L.L. & Dreyfuss, G. RNA and Disease. *Cell* **136**, 777-793 (2009).
48. Mao, F. *et al.* RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Res* **44**, D154-63 (2016).

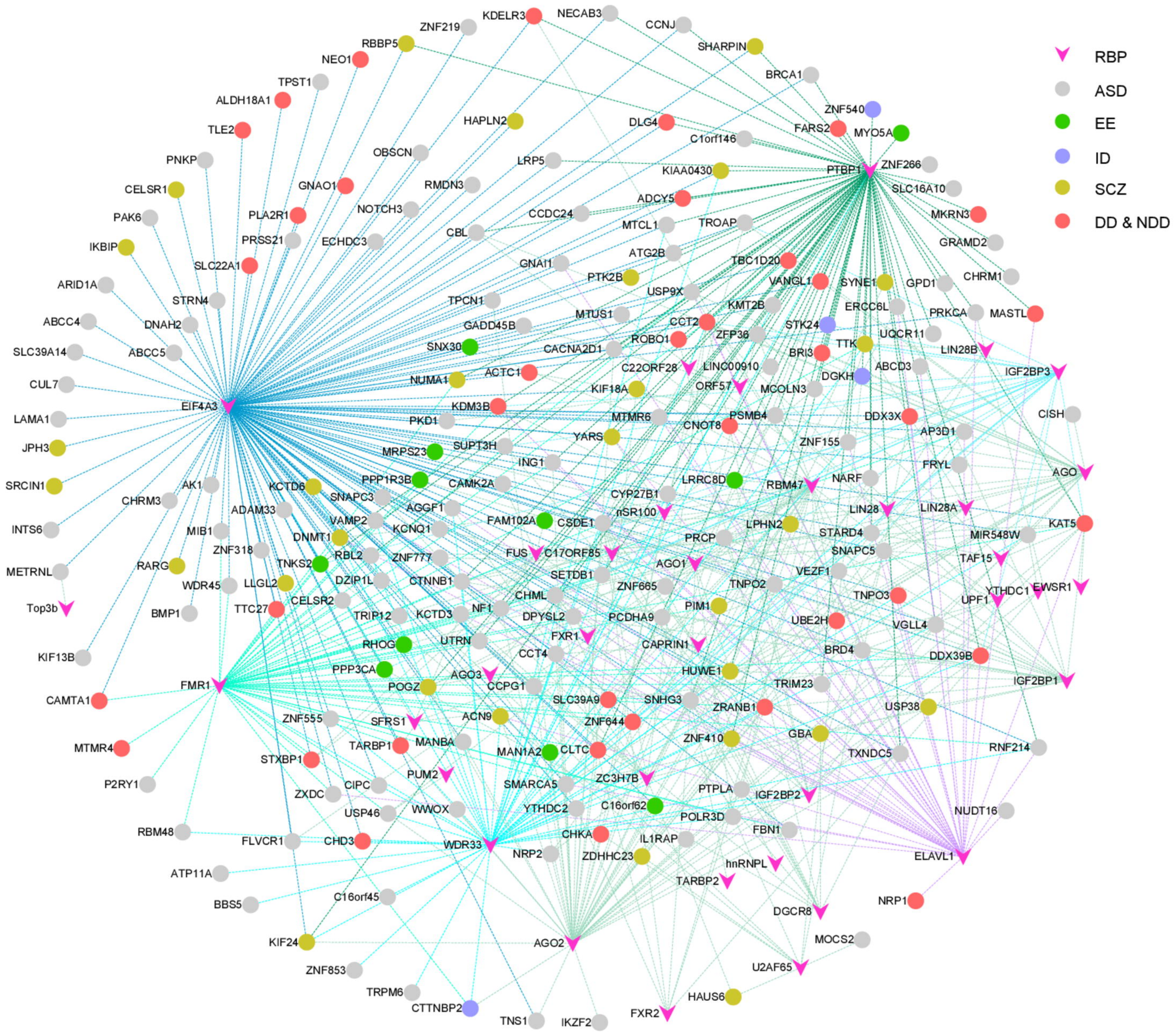

A
Case
Control
Frequency of piDNMs
0.00
0.04
0.08
***
***
*
***
***
1c
1d
1e
2b
2c
2d
RBP-Var Score
B
Odd Ratio of piDNMs
0.0
0.5
1.0
1.5
P=6.02E-02
P=3.61E-04
P=3.65E-06
P=8.16E-08
P=1.54E-17
P=9.52E-17
EE
ID
SCZ
DD&NDD
ASD
Total
−log10(P)
16
12
8
4

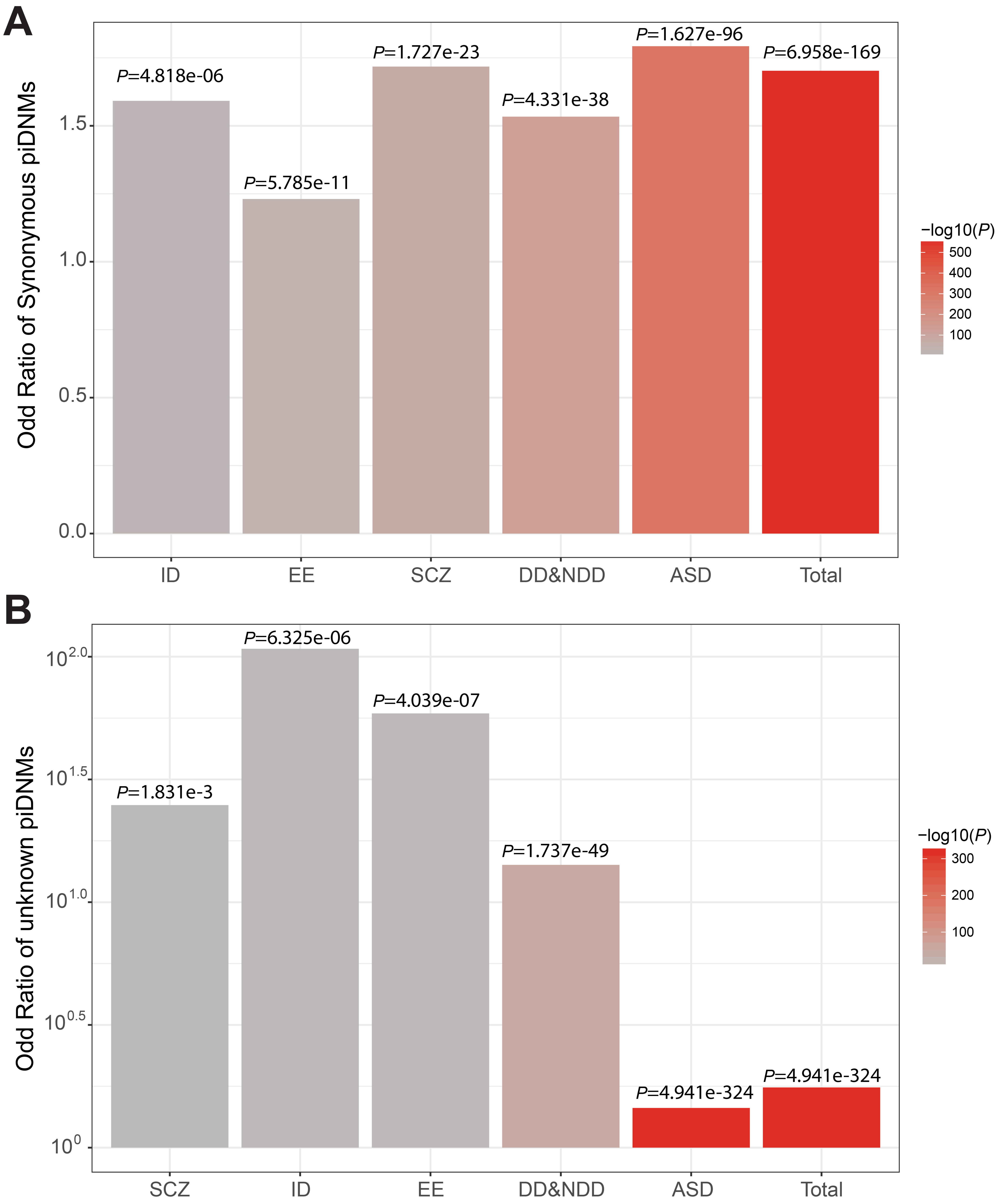

A
SIFT
Case
Control
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0 0
0 0
1851
442
***
2 1
58
7
**
11 6
frameshift
nonframeshift
nonsynonymous
splicing
stop
synonymous
B
PPH
Case
Control
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0 0
0 0
2531
591
***
2 1
2 1
1 0
***
frameshift
nonframeshift
nonsynonymous
splicing
stop
synonymous
C
RegulomeDB
Case
Control
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
17 3
4 0
***
220 26
***
13 1
***
15 0
***
100 11
***
frameshift
nonframeshift
nonsynonymous
splicing
stop
synonymous
D
RBP-Var
Case
Control
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
74 11
***
42 16
1439 321
***
8 0
***
169 30
603 168
***
frameshift
nonframeshift
nonsynonymous
splicing
stop
synonymous

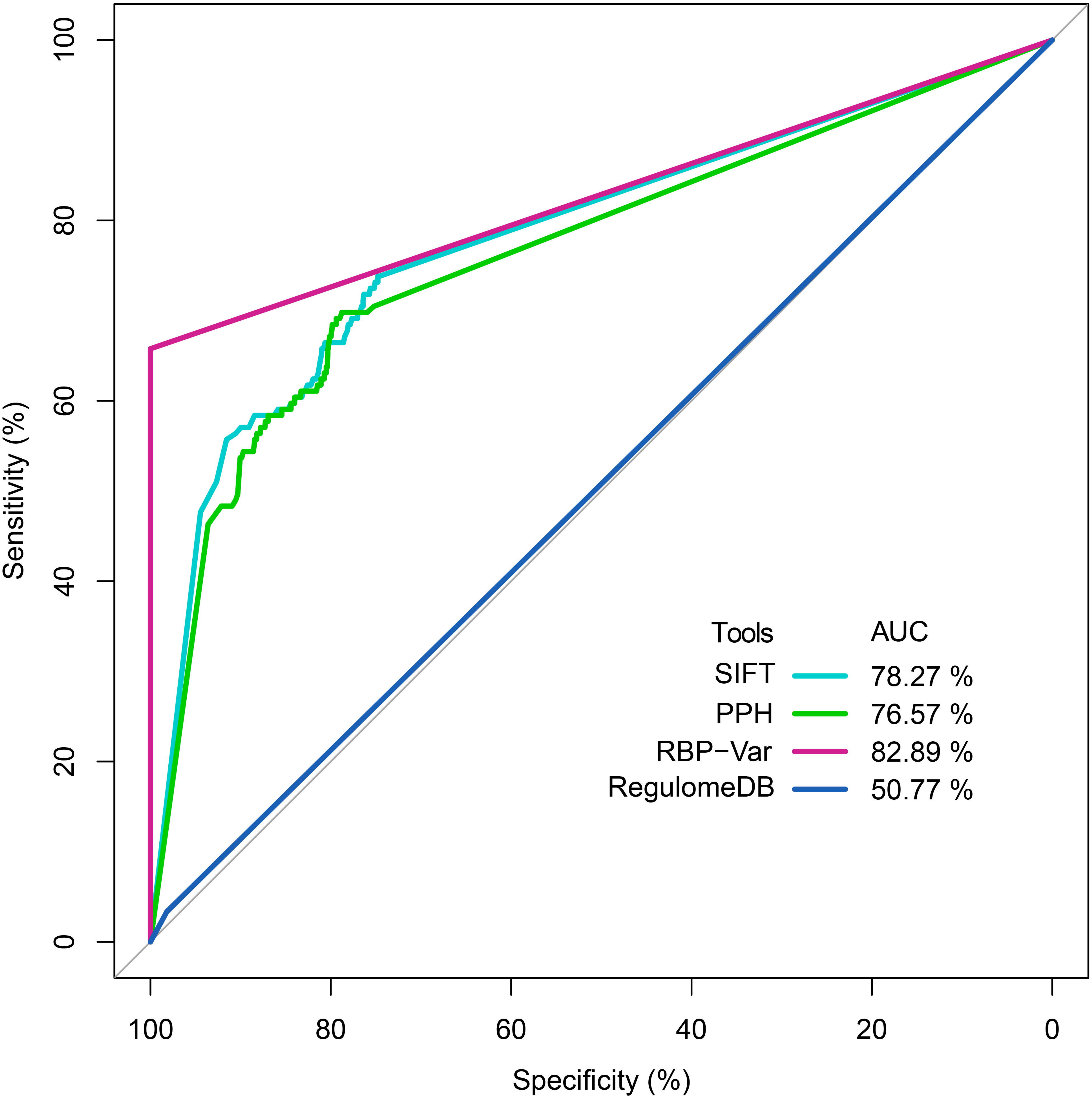

A
8
6
4
2
0
−log10 (pval.TADA)
4.0
4.5
−log10 (Mutation Rate)
ADNP
STXBP1
WDR45
CTNNB1
PTPN11
TNRC6B
MSL2
ILF2
PPP2R5D
DYNC1H1
EEF1A2
COL4A3BP
FHDC1
HADHA
HNRNPUL1
NFASC
dn.LoF
0
1
2
3
5
6
8
B
7
6
5
4
−log10 (pval.TADA)
4.0
4.5
5.0
−log10 (Mutation Rate)
CHD8
MED13L
SYNGAP1
ADNP
SATB2
STXBP1
PTPN11
KMT2A
CHD2
DYRK1A
WDR45
TCF4
WAC
SLC35A2
EP300
POGZ
CTNNB1
GNAO1
SCN8A
SUV420H1
ANK2
TRIP12
MECP2
DNMT3A
TLK2
CASK
CDKL5
BCL11A
SETD5
KMT2C
CHAMP1
TNRC6B
MYO5A
PURA
TAF13
dn.LoF
0
1
2
3
4
5
6
8
9
16
C
SCZ
ASD
ID
EE
DD&NDD
216
1219
52
63
337
15
3
43
6
1
0
0
2
0
1
3
5
0
1
0
7
11
0
21
2
0
3
81
0
1
8
D
Overlap of Six Disorders
Random Overlaps
P<1E-5
Density
0.04
0.03
0.02
0.01
0.00
0
50
100
150
200
250
Number


A
B
Height
1.0
0.9
0.8
0.7
0.6
Dynamic Tree Cut
C
turquoise module
log2(RPKM)
4
2
0
−2
−4
−6
8_pcw
13_pcw
24_pcw
37_pcw
4_mos
1_yrs
3_yrs
8_yrs
13_yrs
19_yrs
21_yrs
30_yrs
36_yrs
37_yrs
D
blue module
log2(RPKM)
2
0
−2
−4
−6

A
CUL3
CTNNB1
HECW2
HNRNPU
SMARCA4
EP300
PTPRF
PPP2R5D
HECTD1
SETDB1
MYH9
ARRB2
EEF1A2
AHCY
PURA
ZC3H3
LLGL2
MTOR
NISCH
IRS1
FOXK2
ADNP
CNOT1
CNOT3
TNRC6A
GAD
ATP1A1
A2M
PONP1
NUMA1
BAIAP2
BRD4
ARID1B
KDM5B
DYNC1LI2
DYNC1H1
RNMT
KPNA1
WIZ
CHAMP1
ZNF644
NOTCH1
NOTCH3
NEDD9
TRRAP
SON
SRRM2
DDX50
MYH10
TCF3
TCF4
GSE1
TCF7L2
DYRK1A
CABIN1
DNM1
NCOR2
KMT2D
DLG5
TCF7L1
USF2
MET
TRIO
ESCO2
JARID2
DNMT3A
CALD1
GNAS
MYO1E
MYO5A
AHNAK
SSH2
SYNE2
ANKRD11
PDE4DIP
NAA10
LARP7
SPAG9
RICTOR
PIK3R3
PXDN
MECP2
ZRANB3
KMT2B
CDKL5
CIC
SATB2
UBR4
NF1
ATP2B4
DOT1L
BRD8
MED13
MAPKBP1
KIF24
TRAPPC9
TBC1D4
POGZ
PPP6R2
PHIP
ILVBL
YLPM1
HBS1L
ATP1A4
LZTR1
LLGL1
BTBD9
IMPDH2
SHKBP1
PGM3
KCTD20
CDC34
B
log2(FPKM)
3
2
1
0
CTL_UMB5168
CTL_UMB5079
CTL_UMB5079
CTL_AN15566
CTL_AN19760
CTL_AN17425
CTL_AN01125
CTL_AN08161
CTL_AN08161
CTL_AN01410
CTL_AN17425
CTL_AN10679
CTL_AN15088
ASD_AN01971
ASD_AN00493
ASD_AN09714
ASD_AN02987
ASD_AN12457
ASD_UMB5278
ASD_AN03632
ASD_AN04682
ASD_AN01570
ASD_AN08166
ASD_AN08792
ASD_AN08043


RBP
ASD
EE
ID
SCZ
DD & NDD
KDELR3
NECAB3
ZNF219
CCNJ
RBBP5
SHARPIN
NEO1
TPST1
BRCA1
ALDH18A1
ZNF540
HAPLN2
DLG4
FARS2
MYO5A
TLE2
C1orf146
OBSCN
PNKP
LRP5
KIAA0430
PTBP1
ZNF266
CELSR1
GNAO1
RMDN3
SLC16A10
ADCY5
PAK6
PLA2R1
NOTCH3
CCDC24
MKRN3
MTCL1
TROAP
PRSS21
ECHDC3
CBL
GRAMD2
IKBIP
ATG2B
SLC22A1
GNAI1
TBC1D20
PTK2B
VANGL1
SYNE1
GPD1
CHRM1
ARID1A
STRN4
TPCN1
USP9X
KMT2B
ERCC6L
PRKCA
MASTL
ABCC4
DNAH2
GADD45B
MTUS1
CCT2
STK24
ZFP36
ROBO1
TTK
UQCR11
SLC39A14
ABCC5
SNX30
CACNA2D1
BRI3
LIN28B
C22ORF28
LINC00910
ABCD3
IGF2BP3
CUL7
NUMA1
ACTC1
ORF57
DGKH
EIF4A3
KIF18A
MCOLN3
KDM3B
CISH
LAMA1
PKD1
MTMR6
PSMB4
DDX3X
CNOT8
AP3D1
JPH3
MRPS23
SUPT3H
YARS
ZNF155
ING1
FRYL
AGO
SRCIN1
PPP1R3B
LRRC8D
NARF
CAMK2A
RBM47
AK1
KCTD6
SNAPC3
CYP27B1
CHRM3
ADAM33
AGGF1
LIN28
LIN28A
nSR100
FAM102A
CSDE1
KAT5
INTS6
VAMP2
LPHN2
MIB1
DNMT1
KCNQ1
STARD4
MIR548W
PRCP
ZNF318
RBL2
ZNF777
FUS
C17ORF85
SNAPC5
TAF15
RARG
TNKS2
AGO1
METRNL
DZIP1L
SETDB1
VEZF1
WDR45
LLGL2
CTNNB1
ZNF665
TNPO2
EWSR1
YTHDC1
CELSR2
CHML
TNPO3
UPF1
Top3b
TTC27
NF1
PIM1
BMP1
TRIP12
KCTD3
DPYSL2
PCDHA9
UBE2H
VGLL4
FXR1
RHOG
UTRN
CAPRIN1
BRD4
DDX39B
CCT4
KIF13B
PPP3CA
AGO3
HUWE1
FMR1
POGZ
CCPG1
TRIM23
IGF2BP1
CAMTA1
SLC39A9
SNHG3
ZRANB1
ZNF555
SFRS1
ACN9
ZNF644
USP38
GBA
ZNF410
MTMR4
TARBP1
MANBA
MAN1A2
CLTC
RNF214
TXNDC5
STXBP1
PUM2
ZC3H7B
P2RY1
CIPC
SMARCA5
PTPLA
IGF2BP2
ZXDC
WWOX
YTHDC2
C16orf62
USP46
POLR3D
NUDT16
FBN1
RBM48
CHKA
IL1RAP
FLVCR1
CHD3
WDR33
ELAVL1
NRP2
ZDHHC23
hnRNPL
ATP11A
TARBP2
NRP1
C16orf45
BBS5
DGCR8
KIF24
AGO2
MOCS2
U2AF65
ZNF853
TRPM6
HAUS6
CTTNBP2
FXR2
TNS1
IKZF2


A
Odd Ratio of Synonymous piDNMs
0.0
0.5
1.0
1.5
ID
EE
SCZ
DD&NDD
ASD
Total
P=4.818e-06
P=5.785e-11
P=1.727e-23
P=4.331e-38
P=1.627e-96
P=6.958e-169
−log10(P)
500
400
300
200
100
B
Odd Ratio of unknown piDNMs
10^{0}
$10^{0.5}$
$10^{1.0}$
$10^{1.5}$
$10^{2.0}$
SCZ
ID
EE
DD&NDD
ASD
Total
P=1.831e-3
P=6.325e-06
P=4.039e-07
P=1.737e-49
P=4.941e-324
P=4.941e-324
−log10(P)
300
200
100

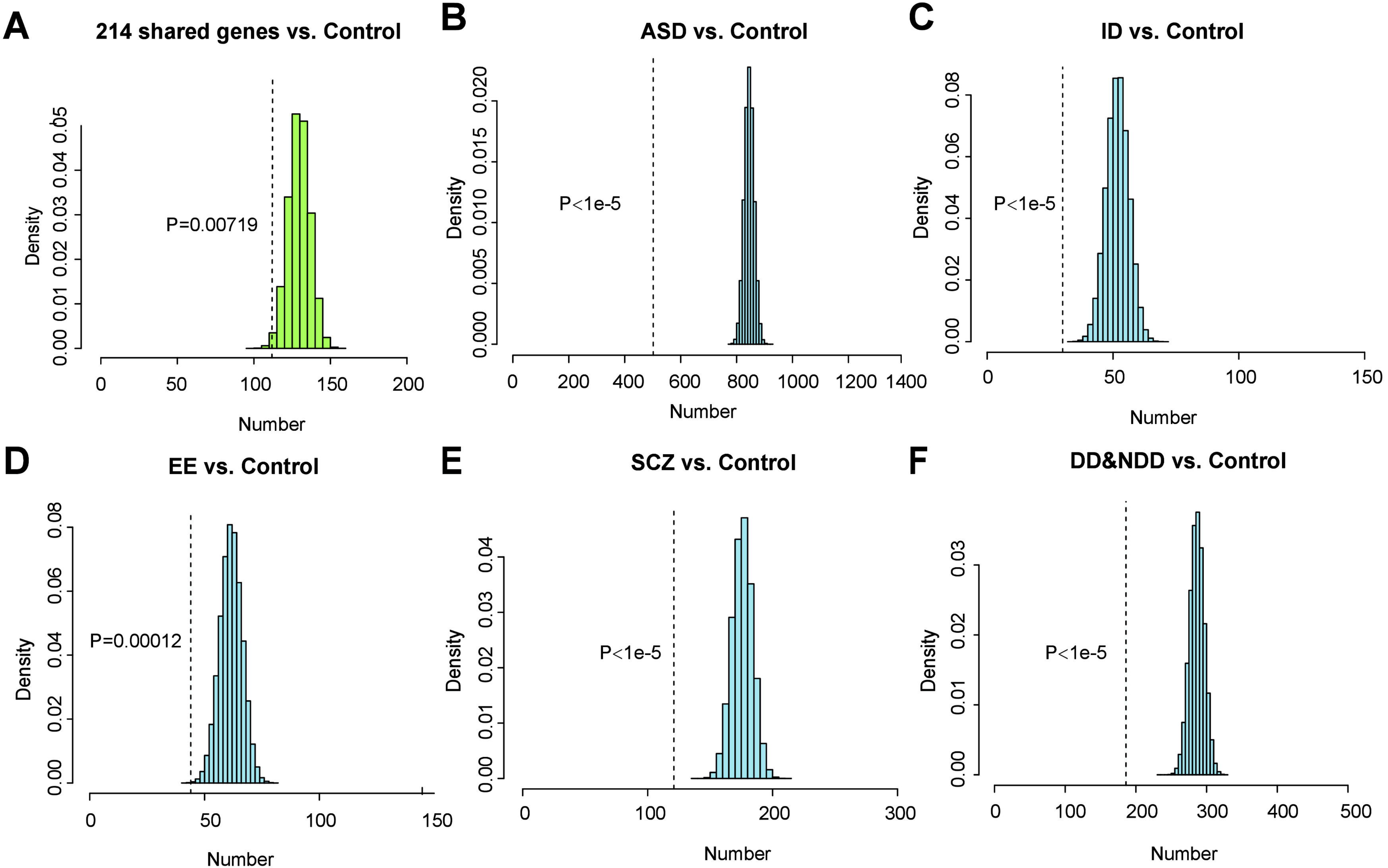


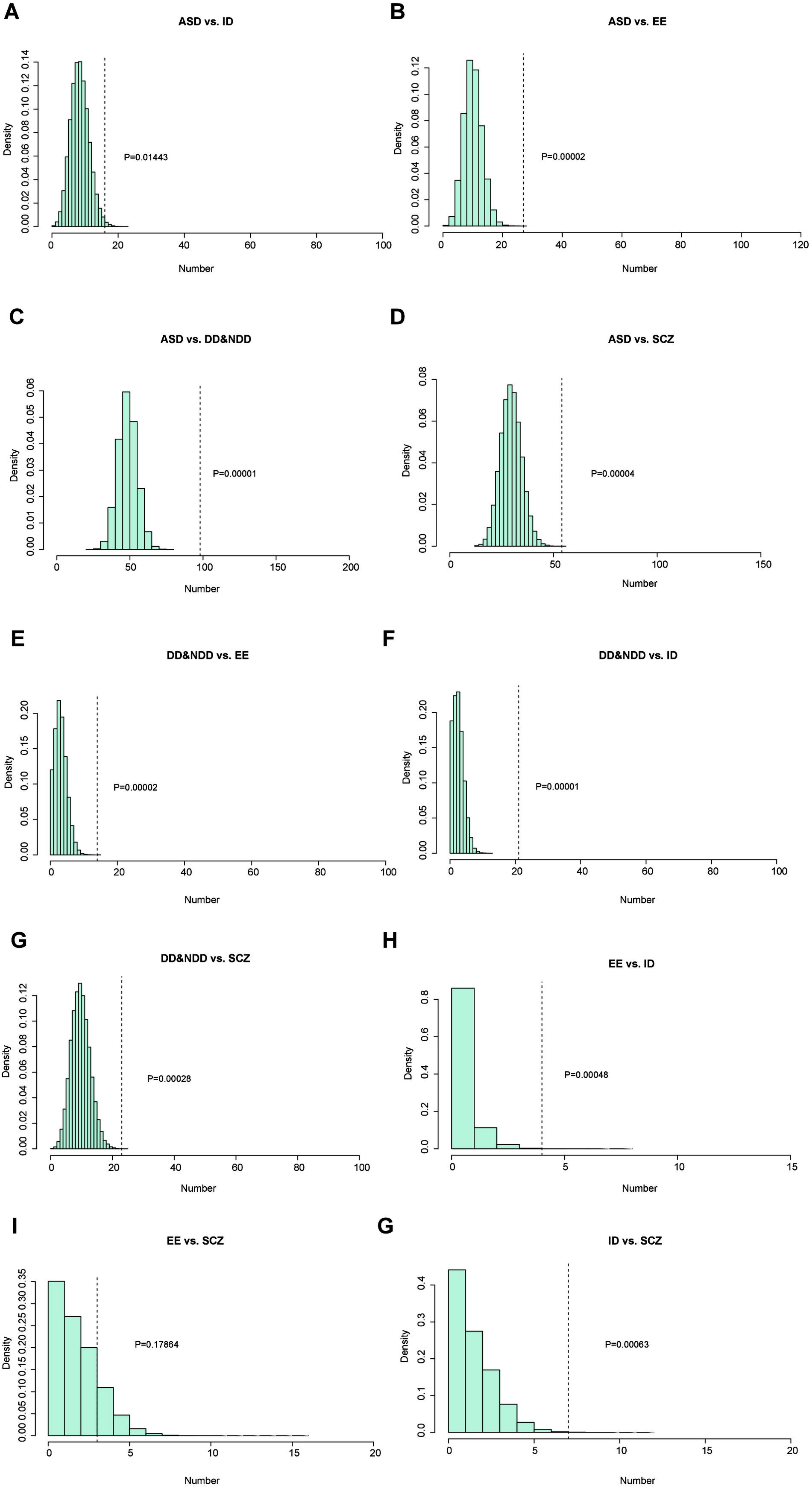
Sensitivity (%)
Specificity (%)
0
20
40
60
80
100
100
80
60
40
20
0
Tools
AUC
SIFT
78.27 %
PPH
76.57 %
RBP−Var
82.89 %
RegulomeDB
50.77 %

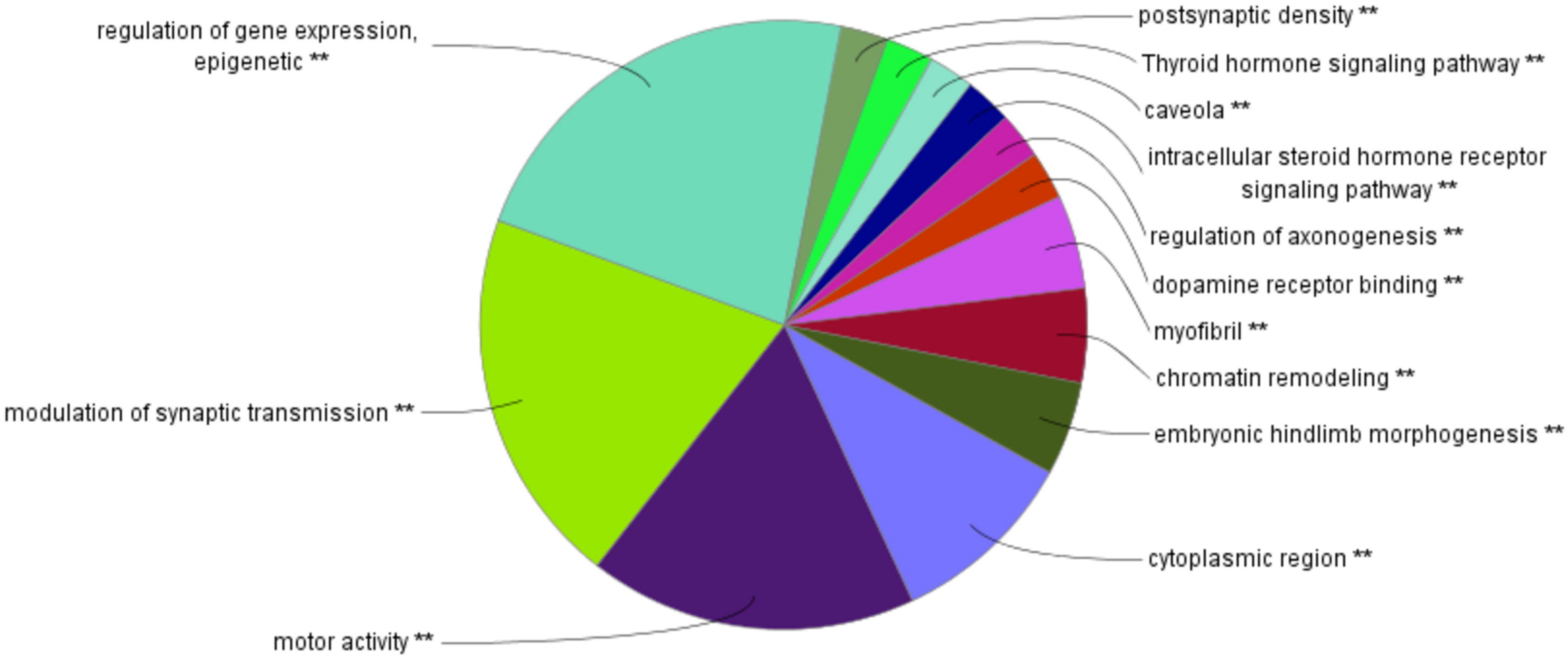
PPH
RegulomeDB
SIFT
RBP-Var
183
260
5
230
69
51
14
22
1234
24
6
433
8
662
108

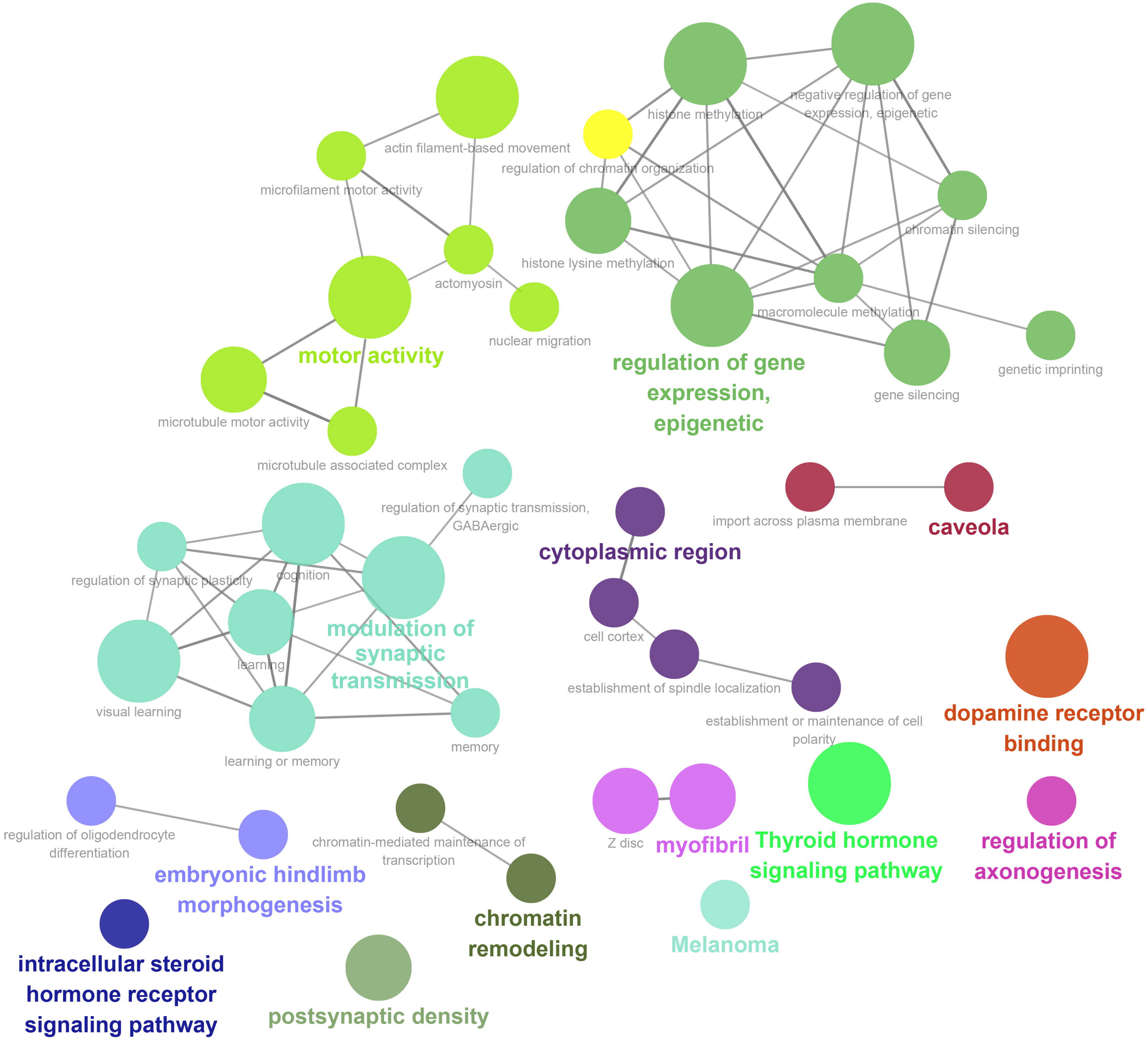
PPH
RegulomeDB
SIFT
RBP-Var
126
187
7
158
65
35
909
14
39
57
7
338
9
602
82

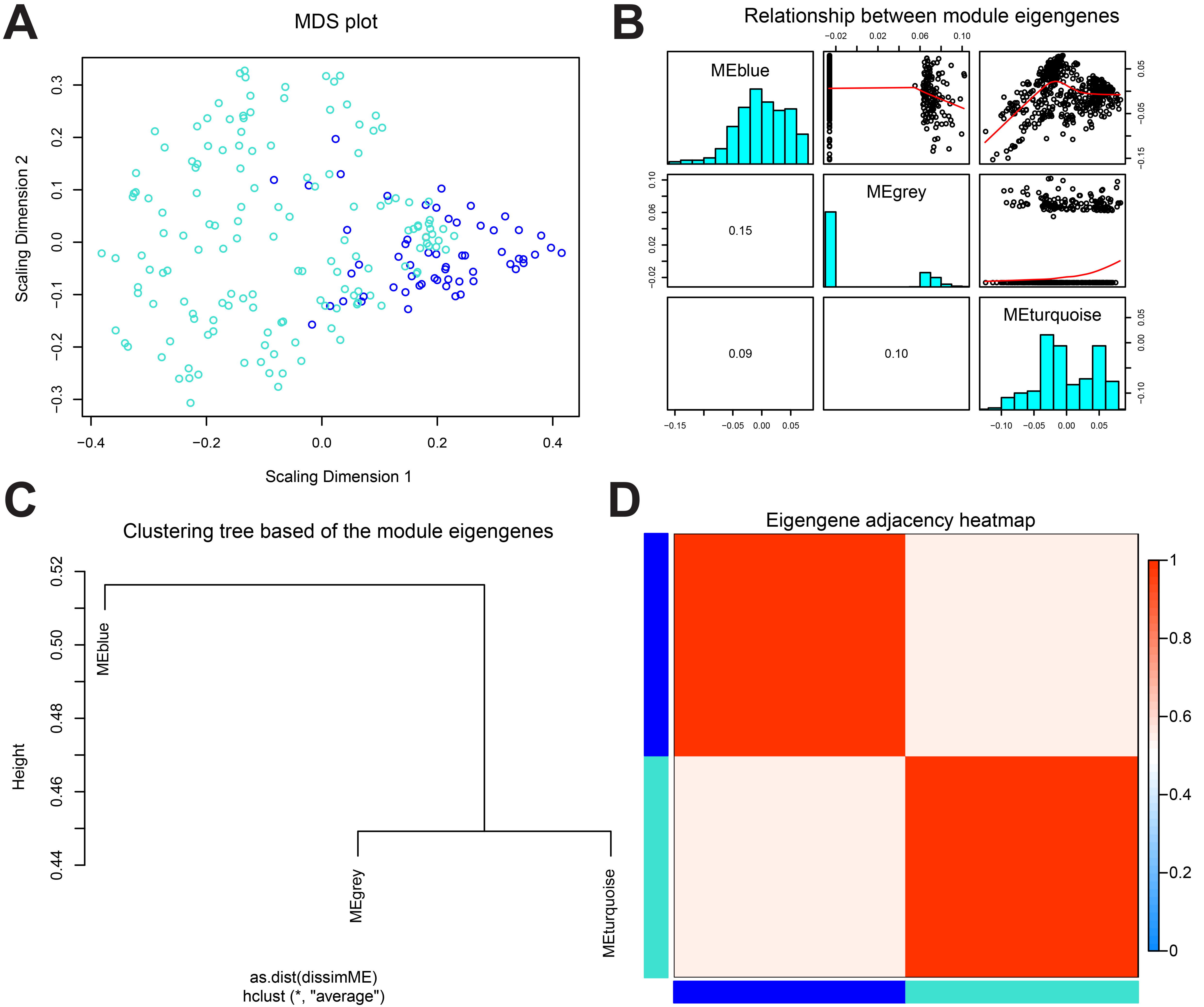
Primary Focal Segmental Glomerulosclerosis FSGS
Canonical and Non-canonical Notch signaling
Integrated Cancer Pathway
miRNA Regulation of DNA Damage Response
Pathways Affected in Adenoid Cystic Carcinoma
ErbB Signaling Pathway
Neural Crest Differentiation
Extracellular vesicle-mediated signaling in recipient cells
Integrated Pancreatic Cancer Pathway
Notch Signaling Pathway
Signaling Pathways in Glioblastoma
DNA Damage Response
Bladder Cancer
Integrated Breast Cancer Pathway
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
-log10 (p vlaue)

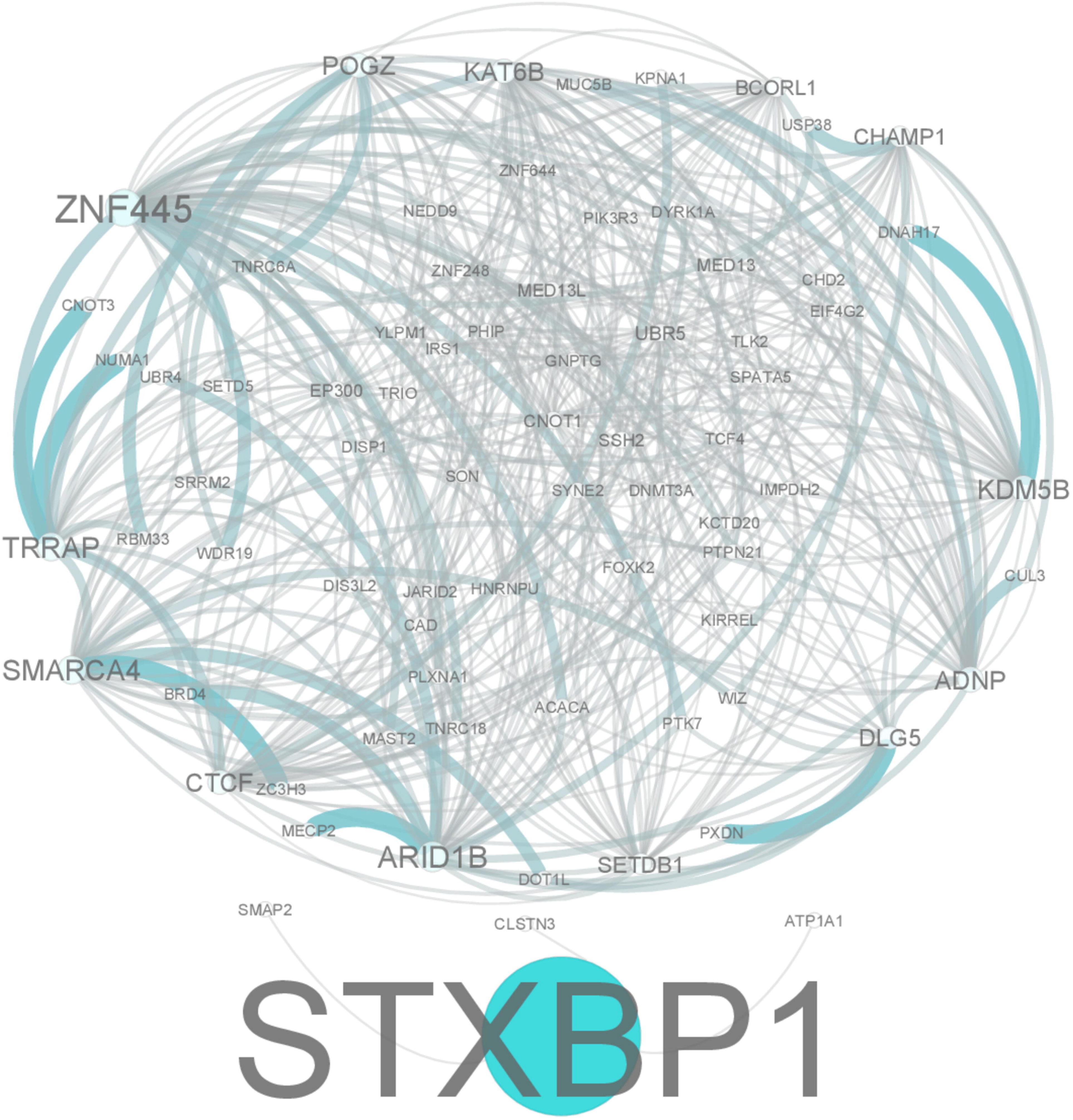

A
214 shared genes vs. Control
P=0.00719
Density
0.00 0.01 0.02 0.03 0.04 0.05
Number
0 50 100 150 200
B
ASD vs. Control
P<1e-5
Density
0.000 0.005 0.010 0.015 0.020
Number
0 200 400 600 800 1000 1200 1400
C
ID vs. Control
P<1e-5
Density
0.00 0.02 0.04 0.06 0.08
Number
0 50 100 150
D
EE vs. Control
P=0.00012
Density
0.00 0.02 0.04 0.06 0.08
Number
0 50 100 150
E
SCZ vs. Control
P<1e-5
Density
0.00 0.01 0.02 0.03 0.04
Number
0 100 200 300
F
DD&NDD vs. Control
P<1e-5
Density
0.00 0.01 0.02 0.03
Number
0 100 200 300 400 500

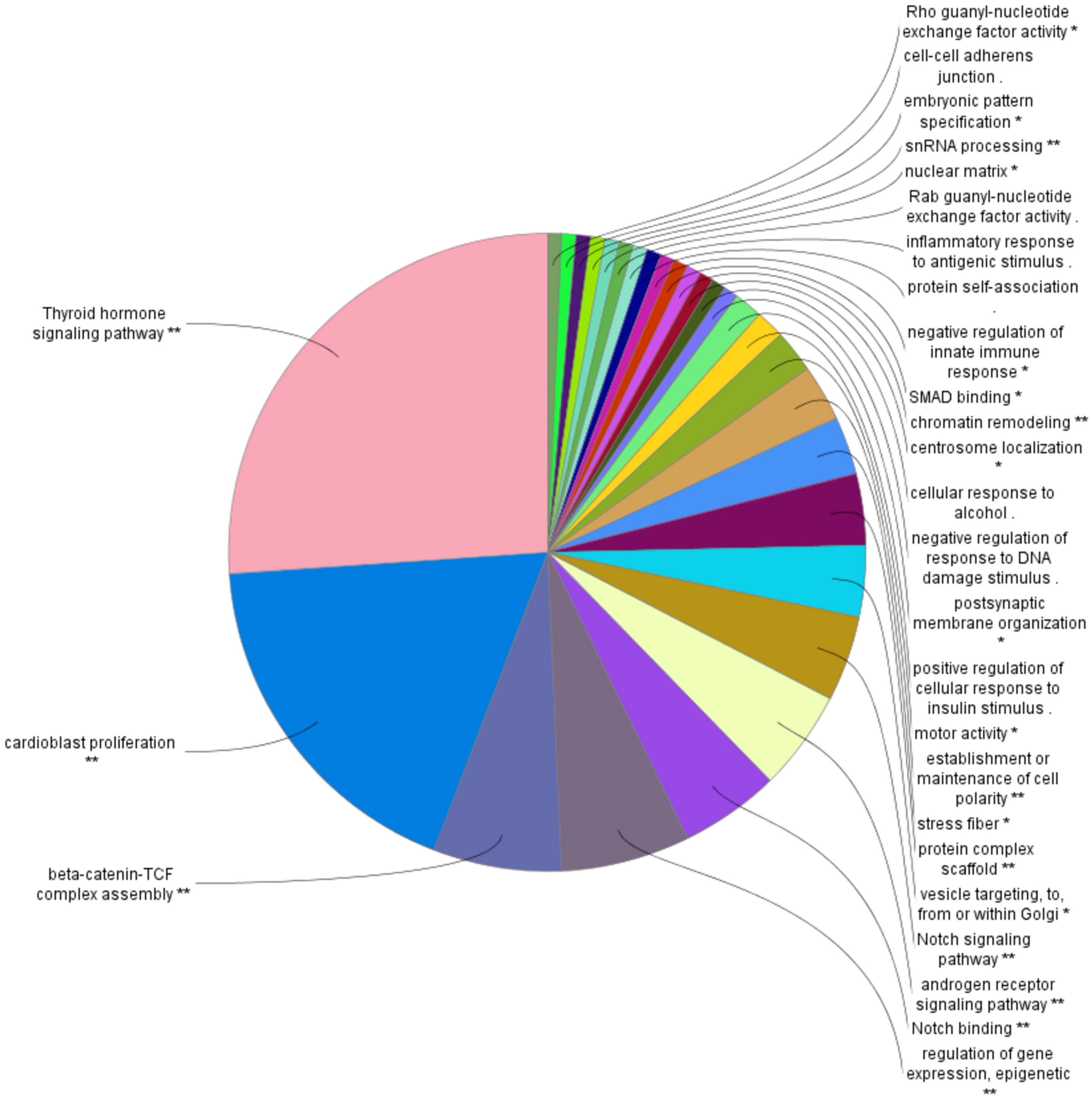

A
ASD vs. ID
P=0.01443
Density
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
0 20 40 60 80 100
Number
B
ASD vs. EE
P=0.00002
Density
0.00 0.02 0.04 0.06 0.08 0.10 0.12
0 20 40 60 80 100 120
Number
C
ASD vs. DD&NDD
P=0.00001
Density
0.00 0.01 0.02 0.03 0.04 0.05 0.06
0 50 100 150 200
Number
D
ASD vs. SCZ
P=0.00004
Density
0.00 0.02 0.04 0.06 0.08
0 50 100 150
Number
E
DD&NDD vs. EE
P=0.00002
Density
0.00 0.05 0.10 0.15 0.20
0 20 40 60 80 100
Number
F
DD&NDD vs. ID
P=0.00001
Density
0.00 0.05 0.10 0.15 0.20
0 20 40 60 80 100
Number
G
DD&NDD vs. SCZ
P=0.00028
Density
0.00 0.02 0.04 0.06 0.08 0.10 0.12
0 20 40 60 80 100
Number
H
EE vs. ID
P=0.00048
Density
0.0 0.2 0.4 0.6 0.8
0 5 10 15
Number
I
EE vs. SCZ
P=0.17864
Density
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35
0 5 10 15 20
Number
G
ID vs. SCZ
P=0.00063
Density
0.0 0.1 0.2 0.3 0.4
0 5 10 15 20
Number

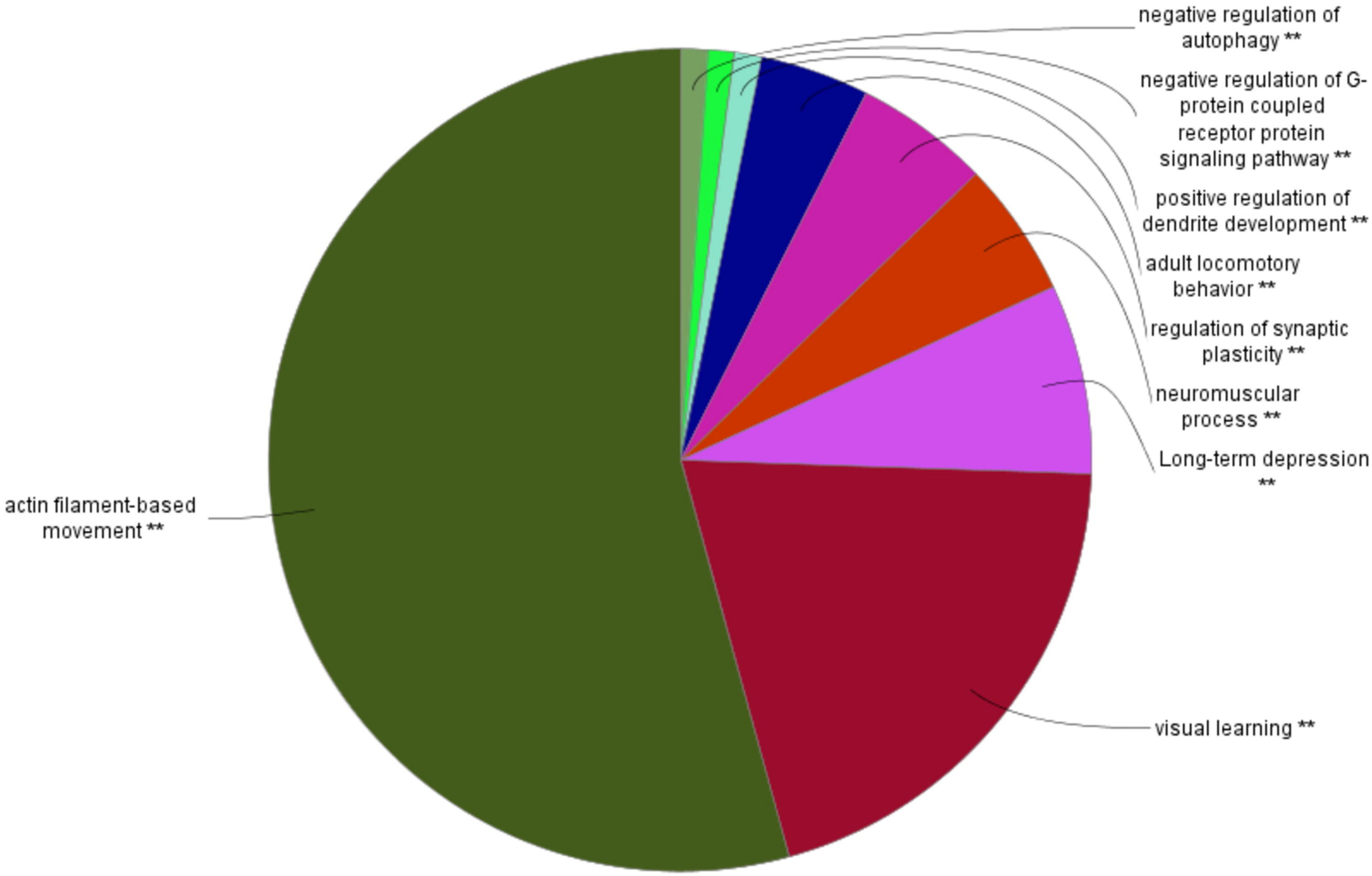

postsynaptic density **
Thyroid hormone signaling pathway **
caveola **
intracellular steroid hormone receptor signaling pathway **
regulation of axonogenesis **
dopamine receptor binding **
myofibril **
chromatin remodeling **
embryonic hindlimb morphogenesis **
cytoplasmic region **
motor activity **
modulation of synaptic transmission **
regulation of gene expression, epigenetic **


actin filament-based movement
microfilament motor activity
actomyosin
nuclear migration
motor activity
microtubule motor activity
microtubule associated complex
histone methylation
negative regulation of gene expression, epigenetic
regulation of chromatin organization
chromatin silencing
histone lysine methylation
macromolecule methylation
regulation of gene expression, epigenetic
gene silencing
genetic imprinting
regulation of synaptic transmission, GABAergic
regulation of synaptic plasticity
cognition
modulation of synaptic transmission
learning
visual learning
learning or memory
memory
cytoplasmic region
cell cortex
establishment of spindle localization
establishment or maintenance of cell polarity
import across plasma membrane
caveola
dopamine receptor binding
regulation of oligodendrocyte differentiation
embryonic hindlimb morphogenesis
chromatin-mediated maintenance of transcription
chromatin remodeling
Z disc
myofibril
Thyroid hormone signaling pathway
regulation of axonogenesis
Melanoma
intracellular steroid hormone receptor signaling pathway
postsynaptic density


A
MDS plot
Scaling Dimension 2
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
0.0
0.2
0.4
Scaling Dimension 1
B
Relationship between module eigengenes
−0.02
0.02
0.06
0.10
MEblue
0.15
MEgrey
0.09
0.10
MEturquoise
−0.15
−0.05
0.00
0.05
−0.10
−0.05
0.00
0.05
C
Clustering tree based of the module eigengenes
Height
0.52
0.50
0.48
0.46
0.44
MEblue
MEgrey
MEturquoise
as.dist(dissimME)
hclust (*, "average")
D
Eigengene adjacency heatmap
1
0.8
0.6
0.4
0.2
0

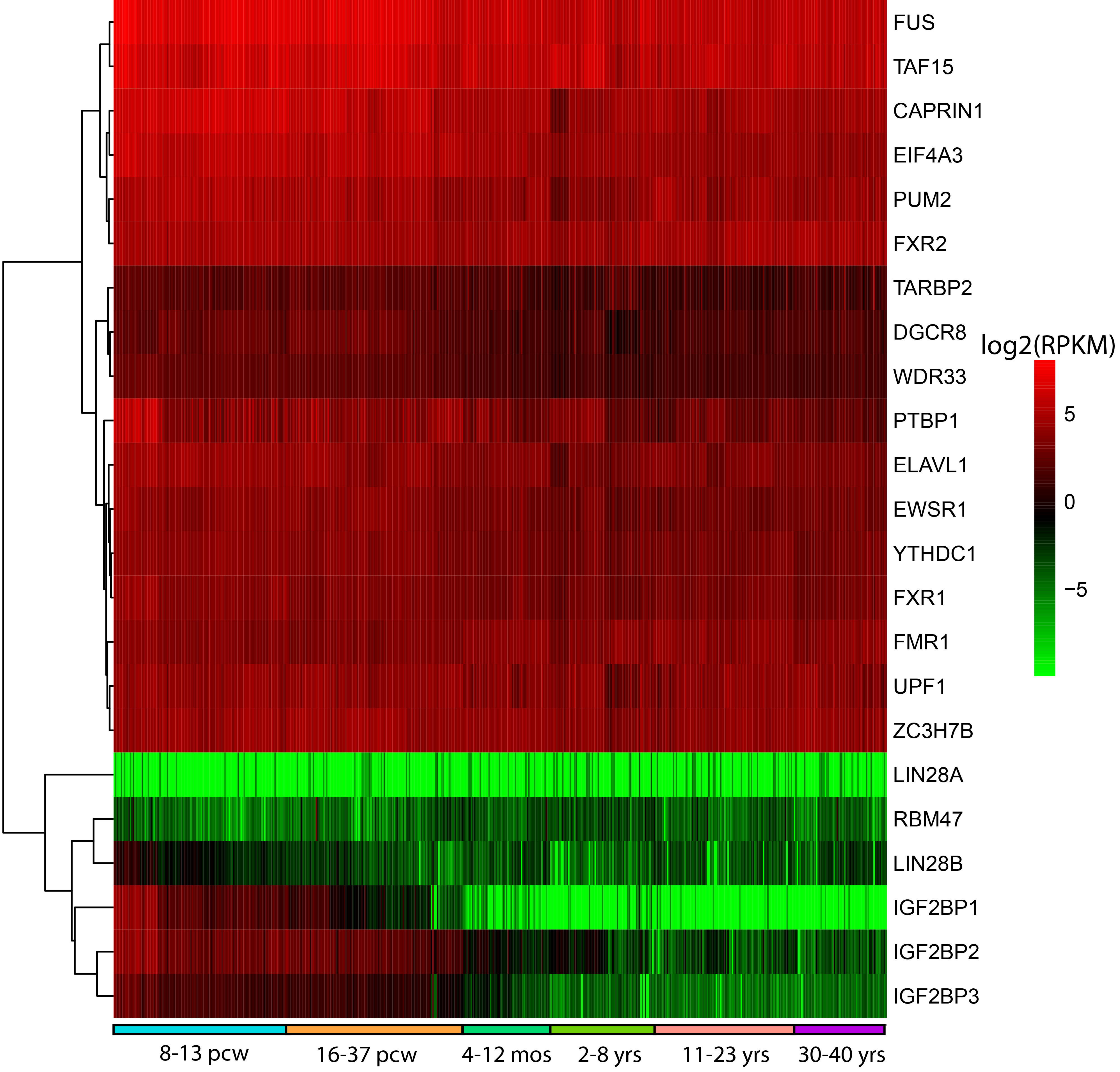

POGZ
KAT6B
MUC5B
KPNA1
BCORL1
USP38
CHAMP1
ZNF644
ZNF445
NEDD9
PIK3R3
DYRK1A
DNAH17
TNRC6A
ZNF248
MED13
CHD2
MED13L
EIF4G2
CNOT3
YLPM1
PHIP
UBR5
TLK2
IRS1
GNPTG
NUMA1
UBR4
SETD5
SPATA5
EP300
TRIO
CNOT1
SSH2
TCF4
DISP1
SON
SRRM2
SYNE2
DNMT3A
IMPDH2
KDM5B
KCTD20
RBM33
TRRAP
WDR19
PTPN21
FOXK2
CUL3
DIS3L2
JARID2
HNRNPU
KIRREL
CAD
SMARCA4
PLXNA1
ADNP
BRD4
ACACA
WIZ
PTK7
MAST2
TNRC18
DLG5
CTCF
ZC3H3
MECP2
PXDN
ARID1B
SETDB1
DOT1L
SMAP2
CLSTN3
ATP1A1
STXBP1


Thyroid hormone signaling pathway **
cardioblast proliferation **
beta-catenin-TCF complex assembly **
Rho guanyl-nucleotide exchange factor activity *
cell-cell adherens junction .
embryonic pattern specification *
snRNA processing **
nuclear matrix *
Rab guanyl-nucleotide exchange factor activity .
inflammatory response to antigenic stimulus .
protein self-association .
negative regulation of innate immune response *
SMAD binding *
chromatin remodeling **
centrosome localization *
cellular response to alcohol .
negative regulation of response to DNA damage stimulus .
postsynaptic membrane organization *
positive regulation of cellular response to insulin stimulus .
motor activity *
establishment or maintenance of cell polarity **
stress fiber *
protein complex scaffold **
vesicle targeting, to, from or within Golgi *
Notch signaling pathway **
androgen receptor signaling pathway **
Notch binding **
regulation of gene expression, epigenetic **


negative regulation of autophagy **
negative regulation of G-protein coupled receptor protein signaling pathway **
positive regulation of dendrite development **
adult locomotory behavior **
regulation of synaptic plasticity **
neuromuscular process **
Long-term depression **
actin filament-based movement **
visual learning **

A
Levl 2
Levl 3
Levl 4
Levl 5
mortality/aging
abnormal survival
partial lethality
preweaning lethality
lethality during fetal growth through weaning
prenatal lethality
postnatal lethality
partial postnatal lethality
perinatal lethality
prenatal lethality
embryonic lethality
P > 0.05
0.01 < P < 0.05
0.001 < P < 0.01
P < 0.001
skeleton phenotype
abnormal skeleton morphology
abnormal axial skeleton morphology
behavior/neurological phenotype
abnormal behavior
abnormal motor capabilities/coordination/movement
abnormal motor coordination/ balance
craniofacial phenotype
abnormal craniofacial morphology
embryogenesis phenotype
abnormal embryogenesis/ development
growth/size/body region phenotype
abnormal postnatal growth/weight/body size
abnormal postnatal growth
abnormal head morphology
muscle phenotype
abnormal muscle morphology
abnormal muscle development
abnormal muscle physiology
nervous system phenotype
abnormal nervous system physiology
abnormal synaptic transmission
abnormal nervous system morphology
abnormal nervous system development
abnormal brain development
abnormal neuron morphology
abnormal brain morphology
abnormal brain development

B
Levl 2
Levl 3
Levl 4
Levl 5
mortality/aging
abnormal survival
preweaning lethality
lethality during fetal growth through weaning
prenatal lethality
postnatal lethality
prenatal lethality
embryonic lethality
lethality throughout fetal growth and development
P > 0.05
0.01 < P < 0.05
0.001 < P < 0.01
P < 0.001
cellular phenotype
abnormal cell physiology
abnormal cell proliferation
craniofacial phenotype
abnormal craniofacial morphology
embryogenesis phenotype
abnormal embryogenesis/ development
abnormal embryo size
endocrine/exocrine gland phenotype
abnormal gland morphology
abnormal pancreas morphology
growth/size/body region phenotype
abnormal prenatal growth/weight/body size
abnormal prenatal body size
abnormal embryo size
prenatal growth retardation
abnormal embryonic growth/weight/body size
abnormal embryo size
abnormal head morphology
muscle phenotype
abnormal muscle physiology
nervous system phenotype
abnormal nervous system morphology
abnormal nervous system development
abnormal brain development
abnormal brain morphology
abnormal brain development


FUS
TAF15
CAPRIN1
EIF4A3
PUM2
FXR2
TARBP2
DGCR8
WDR33
PTBP1
ELAVL1
EWSR1
YTHDC1
FXR1
FMR1
UPF1
ZC3H7B
LIN28A
RBM47
LIN28B
IGF2BP1
IGF2BP2
IGF2BP3
log2(RPKM)
5
0
−5
8-13 pcw
16-37 pcw
4-12 mos
2-8 yrs
11-23 yrs
30-40 yrs