HHS Public Access

Author manuscript

Mol Psychiatry. Author manuscript; available in PMC 2016 April 20.

Published in final edited form as:

Mol Psychiatry. 2016 February ; 21(2): 290–297. doi:10.1038/mp.2015.40.

Genes with *de novo* mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database

Jinchen Li[1,2,3,*], **Tao Cai**[4,*], **Yi Jiang**[2], **Huiqian Chen**[2], **Xin He**[5], **Chao Chen**[3,6], **Xianfeng Li**[3], **Qianzhi Shao**[2], **Xia Ran**[2], **Zhongshan Li**[2], **Kun Xia**[3], **Chunyu Liu**[3,6,‡], **Zhong Sheng Sun**[1,2,‡], and **Jinyu Wu**[1,2,‡]

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

[2]Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China

[3]State Key Laboratory of Medical Genetics, Central South University, Changsha, China

[4]Experimental Medicine Section, NIDCR/NIH, Bethesda, Maryland, USA

[5]Department of Human Genetics, University of Chicago, Chicago, USA

[6]Department of Psychiatry, University of Illinois at Chicago, Chicago, USA

Abstract

Currently, many studies on neuropsychiatric disorders have utilized massive trio-based whole-exome sequencing (WES) and whole-genome sequencing (WGS) to identify numerous *de novo* mutations (DNMs). Here, we retrieved 17,104 DNMs from 3,555 trios across four neuropsychiatric disorders: autism spectrum disorder (ASD), epileptic encephalopathy (EE), intellectual disability (ID), schizophrenia (SCZ), in addition to unaffected siblings (Control), from 36 studies by WES/WGS. After eliminating non-exonic variants, we focused on 3,334 exonic DNMs for evaluation their association with these diseases. Our results revealed a higher prevalence of DNMs in the probands of all four disorders than the one in the controls ($P < 1.3 \times 10^{-7}$). The elevated DNM frequency is dominated by loss-of-function/deleterious single nucleotide variants and frameshift indels (i.e., extreme mutations, $P < 4.5 \times 10^{-5}$). With extensive annotation of these "extreme" mutations, we prioritized 764 candidate genes in these four disorders. A combined analysis of Gene Ontology, microRNA targets, and transcription factor targets revealed shared biological process and non-coding regulatory elements of candidate genes in the pathology of neuropsychiatric disorders. In addition, weighted gene co-expression network analysis (WGCNA) of human laminar-specific neocortical expression data showed that candidate genes are convergent on eight shared modules with specific layer-enrichment and biological process features. Furthermore, we identified that 53 candidate genes are associated with more than one disorder ($P < 0.000001$), suggesting a possibly shared genetic etiology underlying these disorders. Particularly, DNMs of the *SCN2A* gene are frequently occurred across all four disorders. Finally,

‡Correspondence should be addressed to: Jinyu Wu (wujy@mail.biols.ac.cn), Zhong Sheng Sun (sunzs@mail.biols.ac.cn), or Chunyu Liu (liucy@uic.edu).

*These authors contributed equally to this work.

Conflict of Interest: The authors declare no conflict of interest.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

we constructed a freely available NPdenovo database, which provides a comprehensive catalog of the DNMs identified in neuropsychiatric disorders.

Introduction

Over the last decade, next-generation sequencing (NGS) has become one of the most effective tools for identifying the genetic causes of Mendelian, complex, and undiagnosed diseases[1-3]. Recent whole-exome sequencing (WES) and whole-genome sequencing (WGS) studies of neuropsychiatric disorders have indicated that *de novo* mutations (DNMs) play prominent roles in these disorders[4-20] despite their high heritabilities and genetic heterogeneities[21, 22]. DNMs including single nucleotide variants (SNVs), small insertions and deletions (indels), copy-number variants (CNVs), and structural variants (SVs) are extremely rare and regarded as more deleterious, having a stronger disruptive effect on biological functions due to less stringent evolutionary selection[23, 24]. Therefore, DNMs offer considerable insights into the genetic bases and clinical interpretations of sporadic cases in which inheritance seems to offer no explanation for disease etiology[7, 10, 25]. Trio-based WES/WGS is revolutionizing the identification of DNMs, and has been performed on more than 3,000 controls and patients with neuropsychiatric disorders, mostly including autism spectrum disorder (ASD), epileptic encephalopathy (EE), intellectual disability (ID), and schizophrenia (SCZ). These studies identified several dozen candidate genes harboring recurrent loss-of-function (LoF) DNMs that are crucial to pathogenesis of these disorders, such as *CHD8*, *SCN2A*, *NTNG1* and *KATNAL2* in ASD[11-14, 26, 27], *GABRB3*, *ALG13* and *CACNA1A* in EE[9], *DYNC1H1*, *STXBP1*, *SYNGAP1* and *SCN2A* in ID[16, 17], and *LAMA2*, *DPYD*, *TRRAP* and VPS39 in SCZ[18, 19]. However, the genetic etiologies of these disorders remain difficult to decipher due to limited sample sizes, high genetic heterogeneity, and complex pathogenesis[10, 22].

In addition, DNMs are so rare; it has been difficult to be statistically evaluated in terms of the relevance of most detected DNMs to these diseases. To facilitate DNM interpretation, we curated and cataloged all DNMs reported to date in ASD, EE, ID, SCZ, and unaffected siblings or controls by WES/WGS. We subjected all DNMs to consistent quality control standards to characterize the frequency of different classes of DNMs and prioritized genes associated with each disorder. Functional enrichment and co-expression network analysis revealed that some genetic etiologies are shared among these four neuropsychiatric disorders. In addition, the developed NPdenovo database here is a useful tool for future studies in elucidating the mechanisms and underlying the genetic etiologies of these diseases.

Materials and Methods

Data collection and annotation

In total, 3,555 trios from four disorders (ASD, EE, ID and SCZ) together with the unaffected siblings/controls were collected from currently available trios-based WES/WGS studies, in which 17,104 DNMs were identified (Figure 1 and Supplementary Table 1). Comprehensive annotation was performed for each DNM by ANNOVAR[28] with RefSeq (hg19, from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

UCSC), including: 1) Gene information (gene region, effect, mRNA GenBank accession number, amino acid change, cytoband, et al.); 2) Functional prediction for missense mutations by twelve bioinformatics tools; 3) Allele frequency in different populations of public database (different version of dbSNP, 1000 Genomes, ESP6500 and CG69); 4) Disease-related database (ClinVar, HGMD, COSMIC, MGI, OMIM); and 5) Genome features for non-coding variations (segmental duplication, VISTA enhancer, transcription factor, DNase I hypersensitivity, chromatin state segmentation and non-coding RNA from ENCODE).

Identification of extreme mutations

To identify pathogenic mutations, firstly, we removed all DNMs with minor allele frequency (MAF) > 0.1% in dbSNP138, 1000-Genome (released in April, 2012), and ESP6500. Synonymous and non-frameshift mutations were eliminated due to their low possibility to contribute to disorders. The LoF mutations, such as nonsense/splicing SNVs, frameshift indels, were directly considered to be damaging. For missense mutations, which account for the majority of DNMs, though many tools or methods were developed to predict degree of damages based on evolutionary conservation or functional disruption, all of them have inevitable limitations and biases. A proposed solution for this is to use consensus prediction or majority vote of many methods[29]. Consequently, twelve generic tools and methods were applied for functional prediction of damaging missense mutations, expected with more robust statistical power than any single tool, including SIFT[30], Polyphen2_hvar[31], Polyphen2_hdiv[31], MutationTaster[32], MutationAssessor[33], LRT[34], FATHMM[35], GERP++[36], PhyloP[37], SiPhy[38, 39], RadialSVM and **MetaLR**. The predicted damaging scores of twelve tools were sourced from dbNSFP 2.0 database[29] and integrated in ANNOVAR. Then, for each missense mutation, total damaging score is the summed number of tools predicted to be "deleterious" or "conserved". Since missense mutations in cases were found to be more likely to have higher damaging scores ($\geq$ 8) than controls, only the DNMs with a damaging score higher than eight are considered to be deleterious mutations (Supplementary Figure 1). The genes harboring rare LoF/deleterious SNVs and rare frame-shift indels, which we refer to as extreme mutations, were used for candidate genes prioritization (Figure 1).

Prioritize candidate genes

Our recently developed TADA program (Transmission And De novo Association)[40], which can predict 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 corresponding disorders with default parameters. To gain a clear view on risk genes in each disorder, according to the *P*-value generated by the TADA program, we classified genes into five different types in each disorder: "strong" ($P_{TADA} \leq 0.0001$), "suggestive" ($0.0001 < P_{TADA} \leq 0.001$), "positive" ($0.001 < P_{TADA} \leq 0.01$), "possible" ($0.01 < P_{TADA} \leq 0.05$) and "negative" ($P_{TADA} > 0.05$). Genes with $P_{TADA} \leq 0.05$ (i.e. candidate genes) were used for the analysis of functional enrichment and neocortical expression profiles (Figure 1).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Neocortical expression profiles

Human laser micro-dissection data (LMD) of four prenatal brains (15-21 post conceptual weeks, pcw) from BrainSpan Atlas (http://brainspan.org)[41] were used to investigate the expression patterns of candidate genes[42]. In this data, neocortex was divided into ~25 areas in each sample, delineating nine layers per area from subpial granular zone (SZ) to ventricular zone (VZ), corresponding to 526 neocortical samples (Supplementary Table 2). In this study, signed hybrid weighted gene co-expression network analysis (WGCNA) was performed across all neocortical samples[42] using the standard method with power of 4 (as it was the smallest threshold that resulted in a scale-free R^2 fit of 0.9; Figure 1).

Functional enrichment analysis

Functional enrichment of genes (such as Gene Ontology, transcription factor targets, MicroRNA targets, and KEGG pathways, et al.) were all performed by WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) with default parameters[43].

Results

To characterize the DNMs in neuropsychiatric disorders, we retrieved 17,104 DNMs identified from trio-based WES/WGS from ASD, EE, ID, SCZ and normal controls of 36 studies (Supplementary Table 1). After eliminating non-exonic variants, 3,334 exonic DNMs located in the coding DNA sequence (CDS) regions were left, comprising 3,112 *de novo* SNVs and 222 *de novo* indels.

High prevalence of extreme DNMs in neuropsychiatric disorders

Although several studies have documented that cases have a significant high rate of LoF DNMs related to controls[11-14], they have not presented statistically compelling evidence due to the absence of a large sample set. With the combined data from 3,555 trios, we observed that disease groups have higher prevalence of DNMs (both *de novo* SNVs and *de novo* indels) than controls (two-sample Poisson rate test : ASD, OR = 1.36, corrected $P = 5.36 \times 10^{-10}$; EE, OR = 1.52, $P = 2.84 \times 10^{-9}$; ID, OR = 1.60, $P = 1.52 \times 10^{-9}$; SCZ, OR = 1.31, $P = 1.3 \times 10^{-7}$; Table 1 and Supplementary Figure 2A). Interestingly, we found that the significant increase of DNMs is dominated by LoF/deleterious SNVs and frameshift indels, rather than synonymous SNVs, tolerant missense and non-frameshift indels (Table 1 and Supplementary Figure 2B-E). In addition, the frequency of extreme mutations in cases and controls showed significant difference (two-sample Poisson rate test : ASD, OR = 1.80, adjusted $P = 5.96 \times 10^{-11}$; EE, OR = 2.54, $P = 1.89 \times 10^{-15}$; ID, OR = 3.03, $P = 1.30 \times 10^{-19}$; SCZ, OR = 1.50, $P = 4.4 \times 10^{-5}$; Table 1 and Supplementary Figure 2F), suggesting that these extreme DNMs are likely contributed to the pathogenesis of these four neuropsychiatric disorders.

Prioritization of candidate genes

Using the TADA program[44], we identified a total of 764 potential candidate genes with P_{TADA} 0.05 in four disorders: 330 genes for ASD, 109 for EE, 106 for ID, and 277 for SCZ (Table 2 and Supplementary Table 3). We identified six genes with strong associations in ASD, nine genes in EE, 18 genes in ID, and seven genes in SCZ (P_{TADA} 0.0001), most

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of which harbor recurrent extreme DNMs. Many of the 764 candidate genes have been previously reported to be severely implicated in neuropsychiatric disorders: *CHD8*, *SCN2A*, *RLEN*, *NRXN1*, and *NRXN2* in ASD; *SCN1A*, *SCN2A*, *SCN8A*, *STXBP1*, *GABRB3*, and *CDKL5* in EE; *SCN2A*, *DYNC1H1*, *CTCF*, *TCF4*, and *DEAF1* in ID; *LAMA2*, *MIF*, *TRH*, and *HSPA8* in SCZ. In addition to those known candidate genes, we also identified several novel candidates, such as *SUV420H1*, *KATNAL2*, *TBR1*, *NR3C2*, *TUBA1A*, *KIRREL3*, and *UBE3C* in ASD; *SLC35A2*, *THAP1*, *RAB5C*, *VPS37A*, *WDR45*, *IQSEC2*, and *CHD2* in EE; *RB1*, *DEAF1*, *CNGA3*, *RBL2*, *PACS1*, and *FHDC1* in ID; *TAF13*, *ESAM*, *RB1CC1*, and *MKI67* in SCZ.

Pathway shared across disorders

To characterize the function of the 764 identified candidate genes harboring extreme mutations in the four disorders, we conducted an enrichment analysis of Gene Ontology (GO) terms in each disorder (Supplementary Table 4). Although neuropsychiatric disorders have complex etiologies and are genetically heterogeneous, the overlaps suggested that some biological processes share in to all four neuropsychiatric disorders (such as nervous system development, and multicellular organismal signaling), indicating shared molecular pathologies may underlying this subtype of complex diseases. In addition, our results showed that candidate genes in each disorder are highly enriched in respective biological processes. For example, the terms transmission of nerve impulse, synaptic transmission and ion transport are over-represented in the set of EE genes.

Candidate genes that converge on these functional pathways from the neuropsychiatric disorders suggest that they are likely regulated by common special regulatory elements in their non-coding regions. Thus, we performed enrichment analysis of transcription factors (TF) and microRNA targets, which regulates the 5-upstream promoter region and downstream regulatory region (3′-UTR), respectively (Supplementary Table 4). Interestingly, we found that several TF binding sites and microRNA target sites of candidate genes are shared among four neuropsychiatric disorders. For instance, the transcription factor binding sites, “hsa_GGGAGGRR_V$MAZ_Q6”, and “hsa_CAGGTG_V$E12_Q6” are shared by all four disorders; the MicroRNA target sites, “hsa_TTGCACT”, and “hsa_TGAATGT” are shared by ASD and SCZ. In addition, each of the disorders was enriched with some special regulatory elements. For example, ID genes showed enrichment in the “hsa_AAGCCAT” site ($P = 2.71 \times 10^{-12}$, OR = 3.9), which is the target of miR-29A, miR-29B and miR-29C. It should be noted that brain-specific knockdown of miR-29 was found to result in neuronal cell death[45]. Our analysis thus demonstrated a plausible link between given transcription factors or microRNAs with these disorders, although the specific factors/microRNAs remain to be identified and verified.

Shared neocortical expression profiles

Human brain development involves complex regulation of cellular proliferation, signaling, and transcription pathways, and requires orchestration of many relevant genes[46]. We performed weighted gene co-expression network analysis (WGCNA) with our candidate genes in all 526 laminar neocortical samples (Supplementary Materials and Methods). The identified co-expression network includes eight modules of varying sizes from 36 to 170

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

genes with highly coordinated spatio-temporal expression patterns of each module in the human neocortex (M1-M8, Figure 2A). All the modules consist of candidate genes of all four disorders, each is presented with laminar-specific enrichment (Figure 2B and Supplementary Table 5) and distinct biological processes (Figure 2C and Supplementary Table 6). For example, the largest module (M1) included 170 genes enriched in the intermediate zone and was convergent on GO terms “multicellular organismal signaling” ($P = 1.36 \times 10^{-12}$, OR = 4.2) and “transmission of nerve impulse” ($P = 4.07 \times 10^{-12}$, OR = 4.16). Moreover, candidate genes associated with ASD, EE, ID, and SCZ contribute to all the modules, highlighting the shared genetic etiology underlying these disorders.

Shared genes in neuropsychiatric disorders

Among the 764 candidate genes, we identified 12 genes harboring recurrent DNMs in 1,038 ASD trios, eight from 1,024 SCZ trios, eight from 291 EE trios, and 15 from 220 ID trios (Figure 3A). Most of them have been previously linked with neuropsychiatric disorders. However, no single gene was found to harbor recurrent extreme DNMs in 982 controls. More significantly, we identified 53 genes that are associated with more than one disorder (permutation test, $P < 0.000001$ based on random resampling, Figure 3B, Supplementary Figure 3A, Supplementary Materials and Methods). On the contrary, the genes harboring extreme mutations in controls did not significantly correlate with ASD ($P = 0.16$, Supplementary Figure 3B), EE ($P = 0.11$, Supplementary Figure 3C), ID ($P = 0.14$, Supplementary Figure 3D), and SCZ ($P = 0.31$, Supplementary Figure 3E). And the extreme DNM-containing genes in the controls was also not significantly overlapped with the 53 shared genes in the four disorders ($P = 0.12$, Supplementary Figure 3F).

Furthermore, these 53 shared genes are significantly enriched in several neural function-associated GO terms, such as “regulation of transmission of nerve impulse”, “regulation of excitatory postsynaptic membrane potential”, and “regulation of neurological system process” (Supplementary Table 7 and Supplementary Table 8). Briefly, we found that 34% of shared genes (18 of 53) are included in the largest co-expression module M1 (hypergeometric test, $P = 0.026$), suggesting that these shared genes may account for the common etiology of all four neuropsychiatric disorders (Figure 2A, Supplementary Table 5).

More interestingly, different extreme DNMs of *SCN2A* were frequently identified in all four disorders, including five extreme DNMs in ASD ($P_{TADA} = 1.51 \times 10^{-8}$), five in ID ($P_{TADA} = 8.26 \times 10^{-7}$), four in EE ($P_{TADA} = 4.72 \times 10^{-8}$), and one in SCZ ($P_{TADA} = 0.0047$) (Figure 3B and Supplementary Table 9). Of the 15 extreme mutations, nine are frame-shift indels, stop-gain or splicing site SNVs (i.e., LoF DNMs), compared to zero rare LoF mutations of *SCN2A* in the ESP6500 database and controls (MAF < 0.01%, Supplementary Table 9). It is known that *SCN2A* encodes the alpha subunit of the voltage-gated sodium channel, which is responsible for generation and propagation of action potentials in excitable cells, such as nerve and neuroendocrine cells. *SCN2A* and its homologs such as *SCN1A* and *SCN8A*, have been found to be associated with ASD, ID, ataxia, and elevated sensitivity to pain[47]. In support of those previous discoveries, in the combined neocortex co-expression network, *SCN2A* was included in the largest module M1. This means that *SCN2A* is widely expressed in cortex, and is enriched in the intermediate zone (outer cortical plate, CPo; inner

cortical plate, CPi; subplate zone, SP; Supplementary Figure 4). These observations suggested that *SCN2A* may involve common molecular pathways and contribute to overlapped phenotypes in these four disorders.

Moreover, we found that *MYH9*, *LRP1* and *POGZ* were shared by ASD, ID, and SCZ; *GRIN2B* and *STXBP1* were shared by ASD, EE and ID (Figure 3B). Several other candidate genes are shared by two disorders, such as *CHD8*, *RELN*, and *NRXN1* shared by ASD and SCZ; *SETBP1*, *SETD5*, *BRD3*, *CTNNBP2*, *LRP2*, and *SLC6A1* shared by ASD and ID, *WDR45*, *IQSEC2*, *CHD2*, *KCNQ3*, and *SCN8A* shared by EE and ID (Figure 3B). Despite the apparently distinct pathogenesis of these four neuropsychiatric disorders, functional association of these overlapping genes suggests that these disorders share some genetic architecture and molecular pathway features (Supplementary Table 8). For example, *GRIN2B* (glutamate receptor, ionotropic, *N*-methyl D-aspartate 2B) is involved in synaptic transmission and its mutations have been reported in patients with West syndrome and intellectual disability, or behavioral phenotypes[48]. *SETBP1* (SET Binding Protein 1) is associated with ASD (c.2716delC, p.P906fs, $P_{TADA} = 0.004$) and ID (c.1774A>T, p.K592X, $P_{TADA} = 0.0009$). A recent study of more than 30,000 cases with developmental delay showed that *SETBP1* is frequently mutated (both *de novo* SNVs/indels and *de novo* CNVs) in patients with intellectual disability and loss of expressive language[49].

The NPdenovo database

To make our findings easily accessible to the research community, we have developed the NPdenovo database (http://122.228.158.106/NPdenovo/) for storage and retrieval of DNMs, candidate genes, and their brain expression patterns, and for exploring the genetic etiology of neuropsychiatric disorders (Supplementary Figure 5, Supplementary Materials and Methods).

Discussion

To gain insight into the biological implication of DNMs in diseases, datasets from multiple independent studies need to be integratively analyzed in a comprehensive source[7, 50, 51]. In doing so, we found that extreme mutations in the four neuropsychiatric disorders are significantly more frequent than in controls. Specially, EE and ID patients are more likely to harbor recurrent DNMs than ASD and SCZ, suggesting that ID and EE have lower heterogeneities than ASD and SCZ[7]. Numerous studies have revealed that patients with recurrent mutations are prone to present with similar clinically recognizable phenotypes[8, 52-54]. Thus, the identified recurrent DNMs may provide a “genotype-first” approach for complex disease-subtype diagnosis and therapy[55]. For example, patients with disruptive *CHD8* (chromodomain helicase DNA binding protein 8) mutations were characterized as a subtype of autism with macrocephaly, distinct faces, and gastrointestinal complaints, responsible for 0.4% (15/3,730) of the patients with developmental delay or ASD[56]. *ADNP* (activity-dependent neuroprotector) is another ASD-associated gene, which is frequently mutated in 0.17% (10/5,776) of ASD cases with shared clinical phenotypes of intellectual disability and facial dysmorphisms[57].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Next, by employing the TADA program[40] to prioritize associated genes, we identified 764 genes with $P_{TADA} < 0.05$, of which many were previously identified candidate genes as well as a large number of novel candidate genes. Although most of these candidate genes have not been verified or validated by in vitro experiments or animal models, our analysis provided meaningful reference and a downsized list of candidate genes for further studies of these DNMs in neuropsychiatric disorders. For example, *SUV420H1* (suppressor of variegation 4-20 homolog 1) was found to harbor two damaging missense mutations (p.W264S and p.A513V) and one splicing site SNV (c.977+1G>A) in unrelated ASD cases, and was thus prioritized by this study as a strong candidate gene for ASD (P_{TADA} = 0.00009). Furthermore, two damaging *de novo* missense mutations (p.Q264P and p.I228S) in *DEAF1* (DEAF1 transcription factor) were described in two independent ID studies. *DEAF1* is expressed in the neurons and is associated with anxiety and depression phenotypes and behavioral problems[58, 59]. In fact, numerous candidate genes harboring only one extreme DNM were identified with significant *P*-values. Additional experimental data may be required to determine their pathogenic role due to the heterogeneities in these disorders. For example, a dopamine transporter gene, *SLC6A3* (solute carrier family 6 member 3) showed a significant *P* value (P_{TADA} = 0.01). A recent study demonstrated that the *de novo* extreme mutation in this gene (c.1067C>T, p.T356M) can confer risk for ASD based on animal model[60]. Therefore, candidate genes from our study provide new avenue to build up etiological network of these complex neuropsychiatric disorder.

Recently, several studies documented that LoF DNMs in ASD are often involved in chromatin remodeling, wnt signaling, transcriptional regulation and synaptic function[25-27]. Besides, activity-dependent neuronal signaling networks[61] and disruption of neuroplasticity[62] also play a key role in the etiology of ASD. Moreover, DNMs in schizophrenia have been implicated in its etiological development in the fetal prefrontal cortical network[20] and synaptic networks[19]. GO analysis in this study revealed that candidate genes in each disorder are enriched for some unique biological processes, but some are obviously shared, suggesting the overlapped molecular pathways in neuropsychiatric disorders. Our study provides a global view on the molecular etiologies of DNMs in neuropsychiatric disorders.

More importantly, some special regulatory elements are enriched in candidate genes. These elements include not only upstream control regions regulated by particular TFs, but also downstream regulatory regions mediated by particular microRNAs. Most previous studies have focused on the biological function of genes in coding regions. Given by the facts that mutations in non-coding regulatory regions are also involved in etiology of human diseases[63, 64], regulation in non-coding regions is particularly important for brain development[44, 65] and neuropsychiatric disorders[66, 67], we strongly propose that DNMs located in transcription factor and microRNA target sites may participate in the pathology of these four disorders.

To investigate the underlying relationship between these disorders, we performed WGCNA on the 764 candidate genes and identified eight modules with the similar expression patterns in prenatal human neocortex, each module representing specific gene ontology biological processes. Previous studies have demonstrated that ASD[51, 68, 69] and SCZ[20] genes display

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

specific co-expression networks in the human brain or neocortex. Our analysis showed that at least eight co-expression networks of candidate genes are associated with neuropsychiatric disorders. Some of these display superficial layer-enrichment or intermediate layer-enrichment consistent with previous studies[68, 69]. Others display extreme high/medial/low expression profiles in whole neocortex or deep layer-enrichment, which has not been detected in previous studies. However, no single module clearly represents an individual disorder, supporting the existence of interconnected molecular pathways in these four disorders[7, 70, 71]. This assumption is also consistent with the results of a previous protein–protein interaction (PPI) analysis[72].

Previous studies have identified numerous genes associated with multiple neuropsychiatric diseases, such as *SCN2A*, *GRIN2B*, *STXBP1*, *GABRB3*, *RELN*, *GABRA1* and *MECP2*. A recent trio-based WES study found that DNMs associated with chromatin modeling in schizophrenia (*CHD8*, *MECP2*, and *HUWE1*) overlap with ASD and ID[73]. In this study, we identified 53 genes that overlapped in all four disorders, support a genetic overlap between these diseases[7, 70, 71]. Only the extreme DNMs of *SCN2A* were identified in all four disorders. Based on our analysis, *SCN2A* is responsible for 0.5% (5/1038) of patients with ASD, 1.4% (4/291) of EE, 2.3% (5/220) of ID, and 0.1% (1/1024) of SCZ, making it one of the most frequently mutated genes in these neuropsychiatric disorders.

In summary, we have provided new insights into the shared genetic basis of DNMs in neuropsychiatric disorders. We also provide new evidence that some candidate genes, molecular pathways, regulatory elements, and expression profiles are shared among ASD, EE, ID, and SCZ (Supplementary Table 10). All these data can be easily discovered from our online NPdenovo database. In conclusion, our findings may improve understanding of their genetic etiology and facilitate the diagnosis and genetic counseling of these disorders in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project was funded by the National Natural Science Foundation of China (31171236/C060503), the National Basic Research Program of China (No. 2012CB517902 and 2012CB517904), the National “12th Five-Year” scientific and technological support projects (No. 2012BAI03B02), and the Special Funds of National Health and Family Planning Commission of China (No. 201302002).

References

1. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics. 2011; 12(11):745–755.
2. Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, et al. Sequencing studies in human genetics: design and interpretation. Nature Reviews Genetics. 2013; 14(7):460–470.
3. MacArthur D, Manolio T, Dimmock D, Rehm H, Shendure J, Abecasis G, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014; 508(7497):469–476. [PubMed: 24759409]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4. Ku C, Polychronakos C, Tan E, Naidoo N, Pawitan Y, Roukos D, et al. A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. Molecular psychiatry. 2013; 18(2):141–153. [PubMed: 22641181]
5. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 2012; 151(7):1431–1442. [PubMed: 23260136]
6. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al. Using whole-exome sequencing to identify inherited causes of autism. Neuron. 2013; 77(2):259–273. [PubMed: 23352163]
7. Hoischen A, Krumm N, Eichler EE. Prioritization of neurodevelopmental disease genes by discovery of new mutations. Nature neuroscience. 2014; 17(6):764–772. [PubMed: 24866042]
8. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci. 2014; 37(2):95–105. [PubMed: 24387789]
9. Phenome E, Consortium EK. De novo mutations in epileptic encephalopathies. Nature. 2013
10. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nature Reviews Genetics. 2012; 13(8):565–575.
11. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012; 74(2):285–299. [PubMed: 22542183]
12. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012; 485(7397):242–245. [PubMed: 22495311]
13. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012; 485(7397): 246–250. [PubMed: 22495309]
14. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485(7397):237–241. [PubMed: 22495306]
15. Jiang, Y-h; Yuen, RK.; Jin, X.; Wang, M.; Chen, N.; Wu, X., et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. The American Journal of Human Genetics. 2013; 93(2):249–263. [PubMed: 23849776]
16. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. New England Journal of Medicine. 2012; 367(20):1921–1929. [PubMed: 23033978]
17. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. The Lancet. 2012; 380(9854):1674–1682.
18. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012; 44(12):1365–1369. [PubMed: 23042115]
19. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014
20. Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell. 2013; 154(3):518–529. [PubMed: 23911319]
21. Gratten J, Visscher PM, Mowry BJ, Wray NR. Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. Nature genetics. 2013; 45(3):234–238. [PubMed: 23438595]
22. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics. 2012; 13(8):537–551.
23. Crow JF. The origins, patterns and implications of human spontaneous mutation. Nature Reviews Genetics. 2000; 1(1):40–47.
24. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nature Reviews Genetics. 2007; 8(8):610–618.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

25. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends in neurosciences. 2014; 37(2):95–105. [PubMed: 24387789]
26. Poultney CS, Samocha K, Kou Y, Liu L, Walker S, Singh T, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515(7526):209–215. [PubMed: 25363760]
27. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515(7526):216–221. [PubMed: 25363768]
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38(16):e164. [PubMed: 20601685]
29. Liu X, Jian X, Boerwinkle E. dbNSFP v2. 0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. Human mutation. 2013; 34(9):E2393–E2402. [PubMed: 23843252]
30. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009; 4(7):1073–1081. [PubMed: 19561590]
31. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7(4):248–249. [PubMed: 20354512]
32. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods. 2010; 7(8):575–576. [PubMed: 20676075]
33. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic acids research. 2011; 39(17):e118–e118. [PubMed: 21727090]
34. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome research. 2009; 19(9):1553–1561. [PubMed: 19602639]
35. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human Mutation. 2013; 34(1):57–65. [PubMed: 23033316]
36. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++ PLoS computational biology. 2010; 6(12):e1001025. [PubMed: 21152010]
37. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome research. 2010; 20(1):110–121. [PubMed: 19858363]
38. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009; 25(12):i54–i62. [PubMed: 19478016]
39. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011; 478(7370):476–482. [PubMed: 21993624]
40. He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS genetics. 2013; 9(8):e1003671. [PubMed: 23966865]
41. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic acids research. 2013; 41(D1):D996–D1008. [PubMed: 23193282]
42. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014
43. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic acids research. 2013; 41(W1):W77–W83. [PubMed: 23703215]
44. Wenger AM, Clarke SL, Notwell JH, Chung T, Tuteja G, Guturu H, et al. The enhancer landscape during early neocortical development reveals patterns of dense regulation and co-option. PLoS genetics. 2013; 9(8):e1003728. [PubMed: 24009522]
45. Roshan R, Shridhar S, Sarangdhar MA, Banik A, Chawla M, Garg M, et al. Brain-specific knockdown of miR-29 results in neuronal cell death and ataxia in mice. rna. 2014; 20(8):1287–1297. [PubMed: 24958907]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

46. Rakic P. Evolution of the neocortex: a perspective from developmental biology. Nature Reviews Neuroscience. 2009; 10(10):724–735. [PubMed: 19763105]
47. Liao Y, Anttonen A-K, Liukkonen E, Gaily E, Maljevic S, Schubert S, et al. SCN2A mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain. Neurology. 2010; 75(16):1454–1458. [PubMed: 20956790]
48. Lemke JR, Hendrickx R, Geider K, Laube B, Schwake M, Harvey RJ, et al. GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Annals of neurology. 2014; 75(1): 147–154. [PubMed: 24272827]
49. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014
50. Liu L, Sabo A, Neale BM, Nagaswamy U, Stevens C, Lim E, et al. Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. PLoS genetics. 2013; 9(4):e1003443. [PubMed: 23593035]
51. Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. Mol Psychiatry. 2012; 10
52. Schuurs-Hoeijmakers JH, Oh EC, Vissers LE, Swinkels ME, Gilissen C, Willemsen MA, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. American journal of human genetics. 2012; 91(6): 1122–1127. [PubMed: 23159249]
53. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013; 45(7):825–830. [PubMed: 23708187]
54. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012; 338(6114):1619–1622. [PubMed: 23160955]
55. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. Cell. 2014; 156(5):872–877. [PubMed: 24581488]
56. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. Cell. 2014
57. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014; 46(4):380–384. [PubMed: 24531329]
58. Albert PR, Vahid-Ansari F, Luckhart C. Serotonin-prefrontal cortical circuitry in anxiety and depression phenotypes: pivotal role of pre-and post-synaptic 5-HT1A receptor expression. Frontiers in Behavioral Neuroscience. 2014; 8:199. [PubMed: 24936175]
59. Vulto-van Silfhout AT, Rajamanickam S, Jensik PJ, Vergult S, de Rocker N, Newhall KJ, et al. Mutations Affecting the SAND Domain of DEAF1 Cause Intellectual Disability with Severe Speech Impairment and Behavioral Problems. The American Journal of Human Genetics. 2014; 94(5):649–661. [PubMed: 24726472]
60. Hamilton PJ, Campbell NG, Sharma S, Erreger K, Hansen FH, Saunders C, et al. De novo mutation in the dopamine transporter gene associates dopamine dysfunction with autism spectrum disorder. Molecular psychiatry. 2013; 18(12):1315–1323. [PubMed: 23979605]
61. Ebert DH, Greenberg ME. Activity-dependent neuronal signalling and autism spectrum disorder. Nature. 2013; 493(7432):327–337. [PubMed: 23325215]
62. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. Nature Reviews Genetics. 2014; 15(2):133–141.
63. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337(6099):1190–1195. [PubMed: 22955828]
64. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science. 2013; 342(6154):1235587. [PubMed: 24092746]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

65. Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS. A microRNA array reveals extensive regulation of microRNAs during brain development. Rna. 2003; 9(10):1274–1281. [PubMed: 13130141]

66. Maffioletti E, Tardito D, Gennarelli M, Bocchio-Chiavetto L. Micro spies from the brain to the periphery: new clues from studies on microRNAs in neuropsychiatric disorders. Frontiers in cellular neuroscience. 2014; 8:75. [PubMed: 24653674]

67. Xu B, Hsu PK, Karayiorgou M, Gogos JA. MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. Neurobiol Dis. 2012; 46(2):291–301. [PubMed: 22406400]

68. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013; 155(5): 1008–1021. [PubMed: 24267887]

69. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013; 155(5):997–1007. [PubMed: 24267886]

70. Hormozdiari F, Penn O, Borenstein E, Eichler E. The discovery of integrated gene networks for autism and related disorders. Genome research. 2014 gr. 178855.178114.

71. Zhu X, Need AC, Petrovski S, Goldstein DB. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nat Neurosci. 2014; 17(6):773–781. [PubMed: 24866043]

72. Cristino A, Williams S, Hawi Z, An J, Bellgrove M, Schwartz C, et al. Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. Molecular psychiatry. 2013; 19(3):294–301. [PubMed: 23439483]

73. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol Psychiatry. 2014; 19(6):652–658. [PubMed: 24776741]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

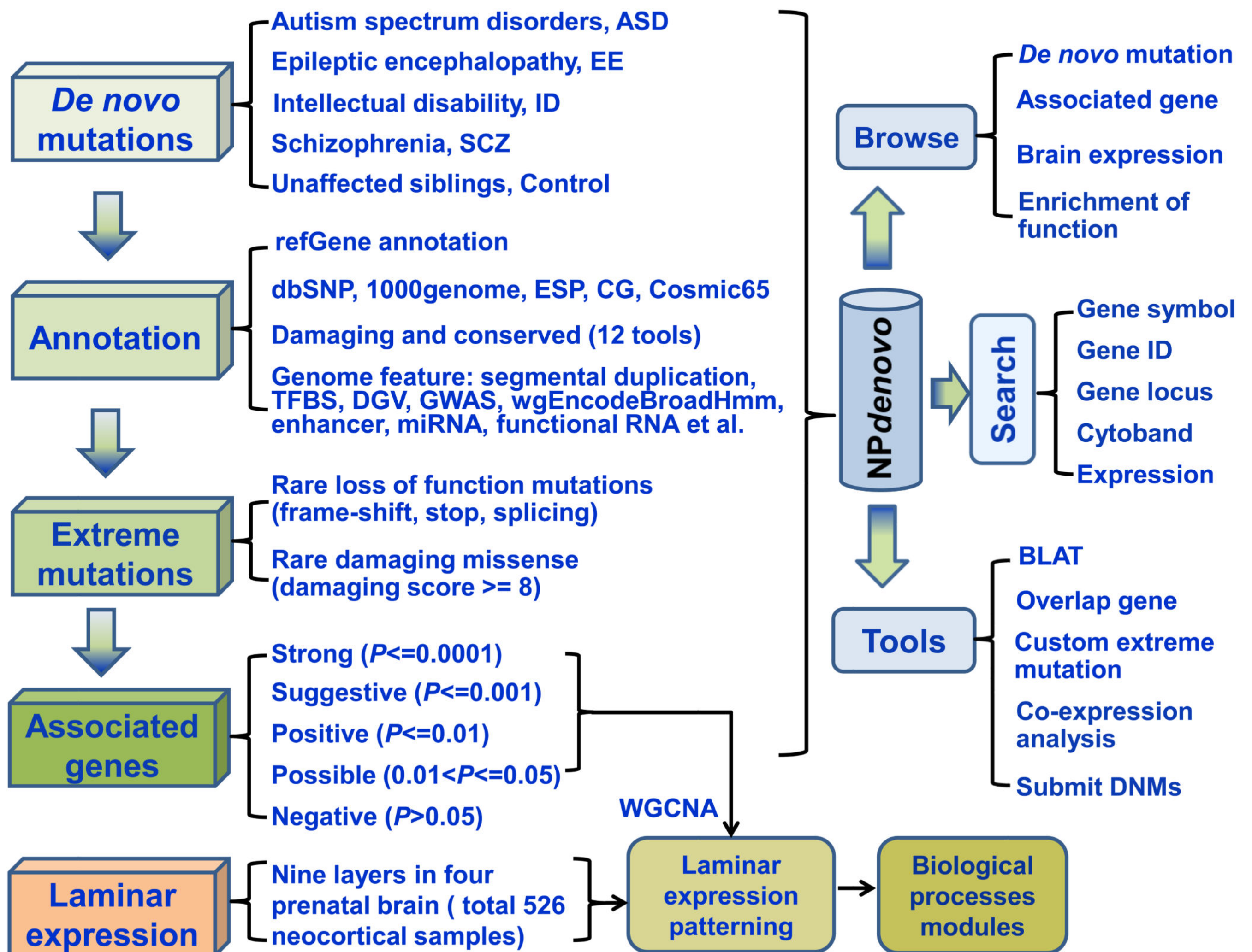

Figure 1. Flowchart of the NPdenovo database

Totally six main parts are included in our analysis of DNMs in neuropsychiatric disorders: 1) Collection of DNMs; 2) Comprehensive annotation of DNMs; 3) Identification of extreme mutations (rare and damaging/LoF mutations); 4) Prioritization of candidate genes with statistic support; 5) Laminar expression patterning of candidate genes by WGCNA; and 6) Development of the NPdenovo database.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

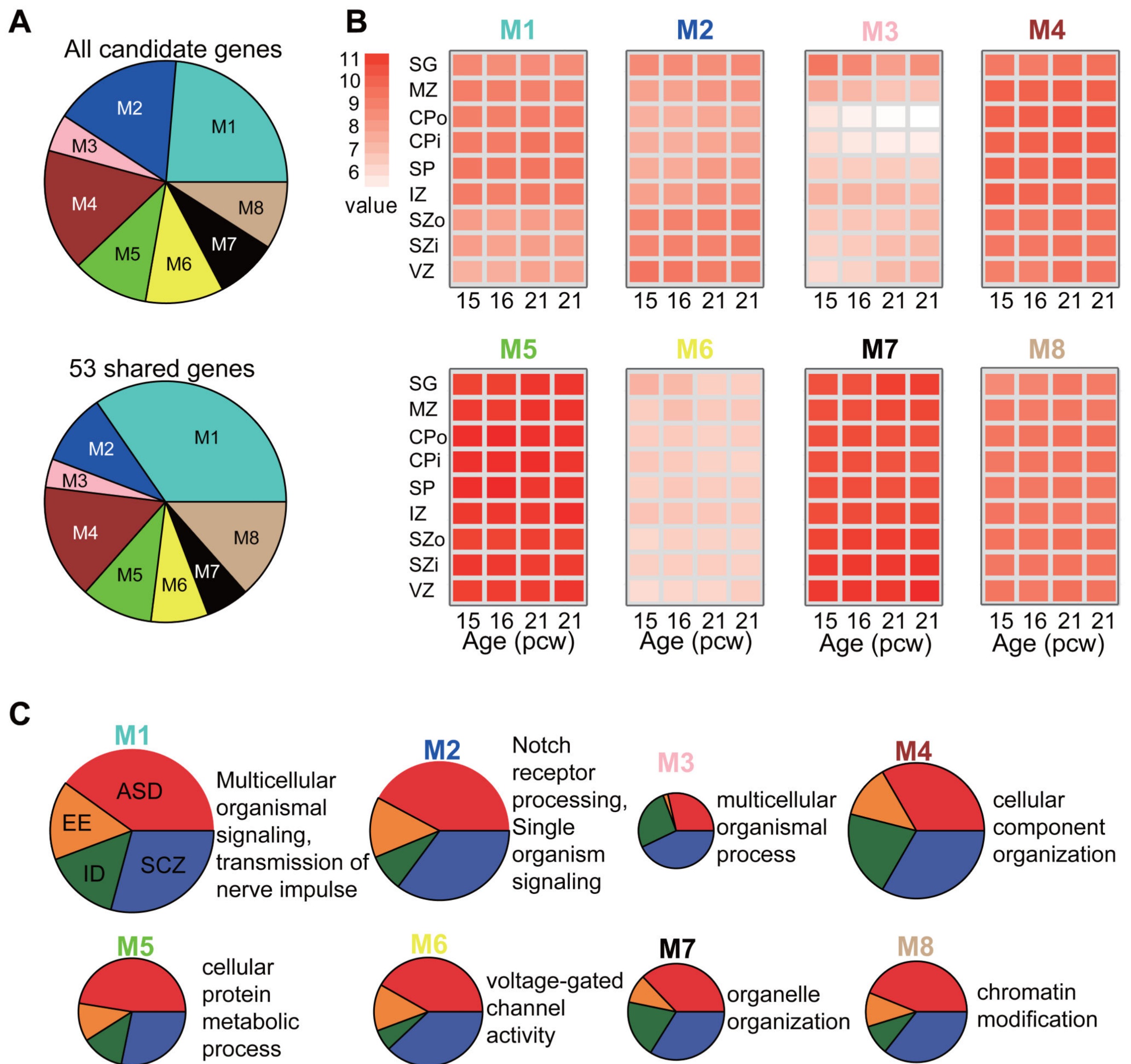

Figure 2. Laminar expression patterning of candidate genes

(A) Distribution of all candidate genes and 53 shared genes in eight modules (M1-M8), which are clustered by WGCNA on the basis of laminar neocortical expression data. Each module is assigned with a color arbitrarily by WGCNA. (B) Module eigengene expression of eight modules in the cortical network. For each module, each box corresponds to the average expression level of genes across the nine layers of neocortex (rows) in four samples (columns). The nine layers correspond to SZ, MZ, CPo, CPi, SZ, IZ, SZo, SZi and VZ (Supplementary Table 2). The four samples correspond to the four high-quality mid-gestational brains, two from 15 and 16 pcw (post-conceptual weeks) and two from 21 pcw. White, low expression; red, high expression. (C) The distribution of candidate genes and

their biological processes in four disorders (ASD, EE, ID and SCZ). Candidate genes in each module are shown as a pie chart in four disorders. The size of each pie chart is proportional to the number of genes in corresponding modules. The enrichment analysis of GO was performed by WebGestalt.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

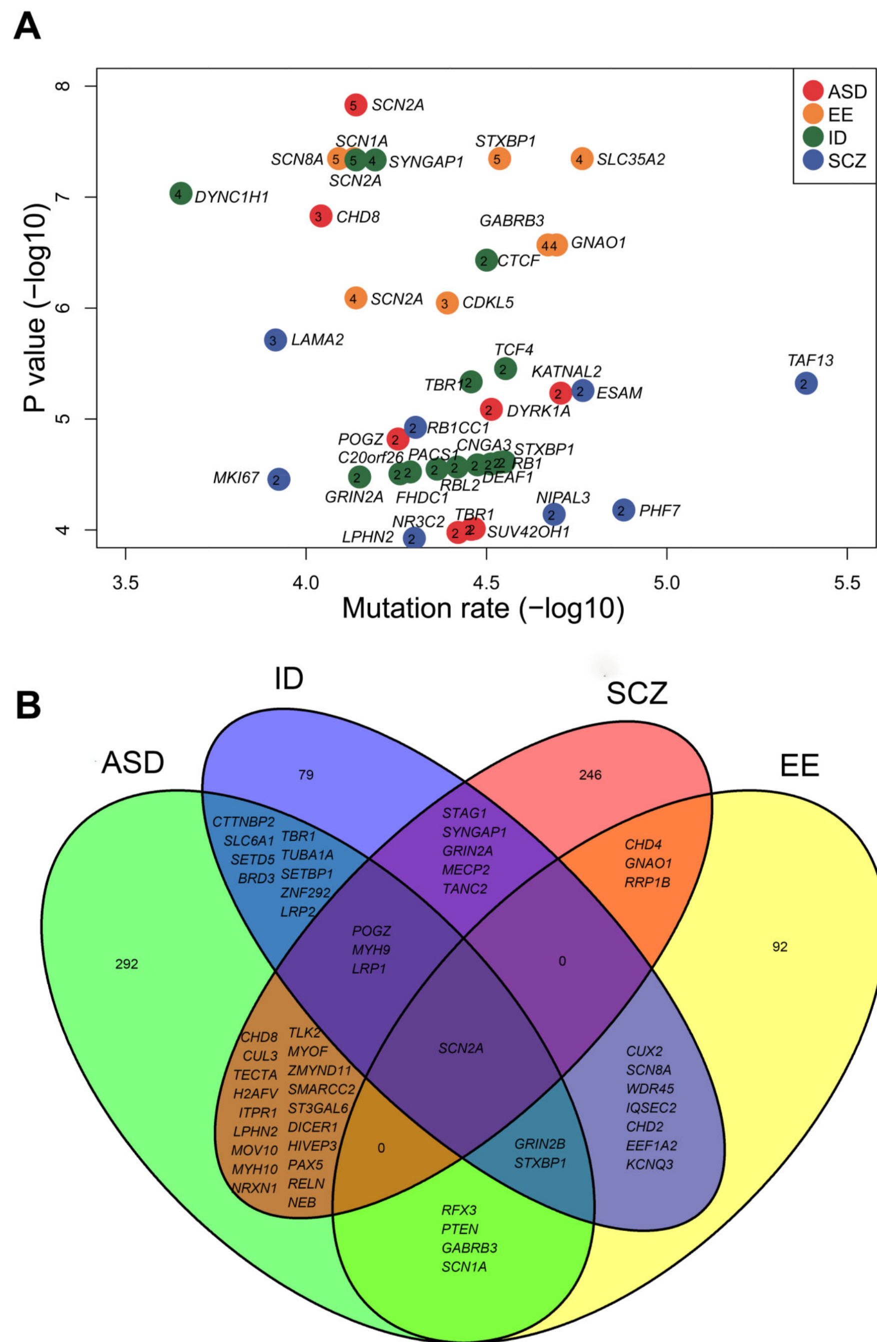

Figure 3. Candidate genes in neuropsychiatric disorders

(A) Scatter diagram of candidate genes with recurrent DNMs. The numbers in each plot indicate the total quantity of extreme DNMs in different disorders. Y-axis represents the value of $-\log^{10}$ (*P*-value), indicating the predicted degrees of association. X-axis denotes $-\log^{10}$ (mutation rate in TADA program). (B) Intersections of candidate genes between each individual disorder. The overlap area in Venn diagram shows the common candidate genes between/among different disorders.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Odds ratio of functional classes of DNMs in coding region

Group	Trios	VS Control	Total DNM	*De novo* SNVs				*De novo* InDels			Extreme mutations
				Total SNVs	Damaging SNVs	Tolerant missense	Synonymous	Total	Frameshift	Non- frameshift	
ASD	1038	DNMs	1040	967	329	394	244	73	63	10	364
		OR	1.36	1.31	2.03	1.18	1.21	2.66	3.31	1.18	1.80
		95% CI	1.24-1.50	1.19-1.45	1.67-2.48	1.02-1.38	1.00-1.47	1.68-4.33	1.94-5.94	0.42-3.45	1.51-2.16
		P-value	1.34E-10	3.27E-08	6.69E-14	0.027	0.055	7.12E-06	1.76E-06	0.82	1.49E-11
		$P_{corrected}$	**5.36E-10**	**1.31E-07**	**2.67E-13**	0.11	0.22	**2.85E-05**	**7.02E-06**	1	**5.96E-11**
EE	291	DNMs	327	303	124	123	56	24	21	3	144
		OR	1.52	1.47	2.73	1.32	0.99	3.11	3.94	1.26	2.54
		95% CI	1.34-1.74	1.28-1.68	2.14-3.49	1.06-1.63	0.72-1.34	1.71-5.64	2.00-7.84	0.22-5.27	2.03-3.18
		P-value	7.11E-10	5.30E-08	1.13E-15	0.01	1	0.000121	2.65E-05	0.722003	4.74E-16
		$P_{corrected}$	**2.84E-09**	**2.12E-07**	**4.54E-15**	0.042	1	**0.00048**	**0.0001**	1	**1.89E-15**
ID	220	DNMs	259	229	108	78	43	30	26	4	130
		OR	1.60	1.47	3.15	1.10	1.00	5.15	6.45	2.23	3.03
		95% CI	1.38-1.85	1.26-1.71	2.44-4.06	0.85-1.42	0.70-1.41	2.94-9.07	3.40-12.48	0.49-8.33	2.41-3.82
		P-value	3.81E-10	1.16E-06	4.97E-18	0.433678	1	3.19E-09	2.12E-09	0.25	3.24E-20
		$P_{corrected}$	**1.52E-09**	**4.64E-06**	**1.99E-17**	1	1	**1.28E-08**	**8.49E-09**	1	**1.30E-19**
SCZ	1024	DNMs	986	917	275	416	226	69	54	15	299
		OR	1.31	1.26	1.72	1.27	1.13	2.54	2.88	1.79	1.50
		95% CI	1.19-1.44	1.14-1.40	1.41-2.11	1.09-1.47	0.93-1.38	1.60-4.16	1.66-5.21	0.72-4.90	1.25-1.81
		P-value	3.29E-08	3.14E-06	4.00E-08	0.0016	0.20	2.09E-05	4.53E-05	0.21	1.10E-05
		$P_{corrected}$	**1.31E-07**	**1.26E-05**	**1.60E-07**	0.0066	0.81232	**8.36E-05**	**0.00018**	0.85	**4.40E-05**
Control	982	DNMs	722	696	153	315	191	26	18	8	191

In total, we collected 3,334 exonic DNMs from 3,555 trios, including 1040 DNMs from1038 ASD trios, 327 DNMs from 291 EE trios, 259 DNMs from 220 ID trios, 986 DNMs from 1,024 SCZ trios and 722 DNMs from 982 trios of control. In each group, exonic DNMs were classified into different functional classes. Compared to control group, *P*-values were calculated on the basis of two-sample Poisson rate test (Supplementary Materials and Methods). Bonferroni correction was used to counteract the problem of multiple comparisons. We referred to LoF/deleterious SNVs as damaging SNVs. The rare LoF/deleterious SNVs and rare frame-shift indels, which were regarded as as extreme mutations. *P* values below 0.005 are highlighted in bold.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Number of associated genes in each neuropsychiatric disorder

Disorder	Strong (P 0.0001)	Suggested (P 0.001)	Positive (P 0.01)	Possible (P 0.05)	Total associated
ASD	6	23	135	166	330
EE	9	22	77	1	109
ID	18	29	59	0	106
SCZ	7	19	115	136	277

We classified all 764 potential candidate genes (P_{TADA} 0.05) into four different types in each disorder on the basis of *P*-value in TADA program. The detail information can be found in Supplementary Table 3.