Identification of *de novo* mutations in prenatal neurodevelopment-associated genes in schizophrenia in two Han Chinese patient-sibling family-based cohorts

Running head: Identification of *de novo* mutations in schizophrenia in Chinese families

Shan Jiang[1,*], Daizhan Zhou[2,3,*], Yin-Ying Wang[1], Peilin jia[1], Chunling Wan[2,3], Xingwang Li[2,3], Guang He[2,3], Dongmei Cao[2], Xiaoqian Jiang[4], Kenneth S. Kendler[5], Ming Tsuang[6], Travis Mize[7,8], Jain-Shing Wu[9], Yimei Lu[9], Lin He[2,3,10,#], Jingchun Chen[9,#], Zhongming Zhao[1,11,12,#] and Xiangning Chen[13,#]

[1]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

[2]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Collaborative Innovation Center for Brain Science, Shanghai Jiao Tong University, Shanghai, China.

[3]Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

[4]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

[5]Virginia Institute of Psychiatric and Behavioral Genetics, Medical College of Virginia and Virginia Commonwealth University, Richmond, VA, 23298, USA.

[6]Department of Psychiatry, University of California at San Diego, San Diego, CA, 92093, USA.

[7]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO 80309, USA.

[8]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO 80309, USA.

[9]Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA.

[10]Institute of Neuropsychiatric Science and Systems Biological Medicine, Shanghai Jiao Tong University, Shanghai, China.

[11]MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA.

[12]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

[13]410 AI, LLC, Germantown, MD 20876, USA.

[*]These authors contributed equally.

[#]Correspondence: Lin He (helin@bio-x.cn) or Jingchun Chen (Jingchun.chen@unlv.edu) or Zhongming Zhao (Zhongming.Zhao@uth.tmc.edu) or Xiangning Chen (va.samchen@gmail.com).

Abstract

Schizophrenia (SCZ) is a severe psychiatric disorder with a strong genetic component. High heritability of SCZ suggests a major role for transmitted genetic variants. Furthermore, SCZ is also associated with a marked reduction in fecundity, leading to the hypothesis that alleles with large effects on risk might often occur de novo. In this study, we conducted whole-genome sequencing for 23 families from two cohorts, Taiwan and Shanghai, with matched unaffected siblings and parents. Two nonsense *de novo* mutations (DNMs) in *GJC1* and *HIST1H2AD* were identified in SCZ patients. By performing DNM-based burden test, we found that *DPYSL2, NBPF1, SDK1, ZNF595, ZNF718, GCNT2, SNX9, AACS, KCNQ1* and *MSI2* genes carried more DNMs in SCZ patients than their unaffected siblings. Developmental expression analyses revealed that the loss-of-function DNM genes and DNM-burdened genes showed significantly higher expression in prefrontal cortex in prenatal stage. The biased high expression of *GJC1* and *SDK1* in prenatal stage were preserved in other brain regions and conserved across species. Data from genome-wide association studies suggested that common variants in the *GJC1*, *HIST1H2AD* and *SDK1* genes may contribute to the risks of SCZ and SCZ-related traits. Analyses on *GJC1* revealed that genes co-expressed with *GJC1* were involved in SCZ, SCZ-associated pathways and drug targets. *GJC1* may be a potential target for SCZ in prenatal and early neurodevelopment.

Key words: schizophrenia, *de novo* mutation, whole genome sequencing, neurodevelopment, loss-of-function

Introduction

Schizophrenia (SCZ) is a severe psychiatric disorder that profoundly affects cognitive, behavior and emotional processes, yet its etiology and pathophysiology are still largely unknown. The high heritability of SCZ suggests that genetic risk factors contribute to a significant proportion of the etiology [1, 2]. However, the marked reduction in fecundity in SCZ patients suggests the removal of risk variants with the largest effects from the population by natural selection. Thus these variants often occur *de novo*. Indeed, the strongest genetic risk factors for SCZ identified so far are *de novo* large copy number variants (CNV) [3].

The availability of next-generation sequencing permits the detection of *de novo* mutation (DNM) events at the genome level. Through large-scale sequencing in parent-offspring trios, DNM have been increasingly discovered from an array of severe neurodevelopmental disorders, including Autism Spectrum Disorders (ASD) [4], Attention-Deficit Hyperactivity Disorder (ADHD) [5] and epileptic encephalopathy [6]. For SCZ, Xu *et al.* showed a large excess of *de novo* nonsynonymous changes and DNMs presented with greater potential to affect protein structures and functions in SCZ patients [7]; Girard *et al.* reported increased exonic DNM rate in SCZ patients [8]; and Fromer *et al.* found that DNMs in SCZ implicated synaptic networks and DNM-affected genes in SCZ overlapped with those mutated in other neurodevelopmental disorders [9]. Whole exome sequencing (WES) was applied predominantly in DNM identification thus far, however, few DNM studies have been based on whole genome sequencing (WGS) as WES remains a cost-effective strategy.

Unfortunately, only 1.22% of DNMs are within exonic regions [10] and meaningful mutations may occur outside of exons, such as in regulatory elements (i.e. transcriptional promoters, enhancers and suppressors) thereby altering expression level of governed genes. Similarly, mutations

within exon-intron junction regions may influence splice sites and thus lead to inappropriate expression of particular isoforms [11]. Emerging roles of non-coding RNAs such as microRNA, circRNA and lncRNA further call for attention to explore DNMs in non-coding regions.

We therefore conducted WGS on two cohorts of Chinese families with matched patient-sibling to capture all classes of DNMs and to more fully describe the genetic architecture of SCZ. This study was conducted to identify potential DNMs in SCZ in the Asian population, adding to the growing body of information regarding ethnicity-specific DNMs.

Methods

Subjects

Subjects were drawn from two distinct cohorts, Taiwan and Shanghai, of Han Chinese origin. Recruitment of subjects from the Taiwan cohort was described in previous publications [12, 13]. Briefly, families with at least three siblings, two of whom were diagnosed with SCZ, were recruited in the Taiwan Schizophrenia Linkage Study (TSLS) from 1998 to 2002. All recruited subjects were interviewed using the Diagnostic Interview for Genetic Studies (DIGS) [14], accompanied with the Family Diagnostic Interview for Genetic Studies (FIGS) (https://www.nimhgenetics.org/resources/clinical-instruments/figs/list-of-figs). Final diagnostic assessment was based on the criteria of the fourth edition of the Diagnostic and Statistical Manual (DSM-IV), joined with the record of DIGS, FIGS, interviewer notes, and hospital anamnesis. For the Shanghai cohort, families from the Bio-x SCZ Biobank with at least three siblings, two of whom were diagnosed with SCZ, were selected. All families from the Bio-x SCZ Biobank were recruited from the city of Shanghai and the provinces of Hebei, Liaoning and Guangxi from 2001 to 2003. All individuals with SCZ were interviewed by two independent

psychiatrists and diagnosed according to DSM-IV criteria. A total of 23 families (10 from Taiwan and 13 from Shanghai), with SCZ patients, matched unaffected siblings and parents, were used in this study (Figure 1). For detailed demographic characteristics of all individuals in the 23 families, please refer to Supplementary Table 1. All subjects gave written informed consent with the approval of the local research ethics committees.

Whole genome sequencing

For the subjects of Taiwan cohort, whole blood samples were collected with anticoagulant (EDTA) tubes and sent to the National Institute of Mental Health (NIMH) Repository and Genomics Resource (RGR). Lymphocytes from the whole blood samples were transformed into immortalized lymphoblastoid cell lines and stored. The DNA samples extracted from the cell lines were used for WGS. WGS was carried out on the Illumina HiSeq 2000 platform using paired-end chemistry with 75 base-pair read length through NovoGene, Inc. (Beijing, China). For detailed description please refer to previous publication [15]. WGS data for the subjects of Taiwan cohort can be accessed in BioProject of NCBI: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA551447. For the subjects from Shanghai cohort, whole blood samples were also collected with anticoagulant tubes. DNA was extracted from blood lymphocytes by standard procedures using FlexiGene DNA kits (Fuji). DNA libraries were prepared using protocols recommended by Illumina (Illumina, San Diego, CA). WGS was performed on Illumina HiSeq-X Ten platform with 150 base-pair read length through Cloud Health Genomics Ltd. (Shanghai, China).

Quality control and variant calling

FastQC (v0.11.8) was used to perform quality checks on all samples across Taiwan and Shanghai cohorts (Supplementary Figure 1).

The GATK best practices of variant calling were applied to process all raw reads from both Taiwan and Shanghai cohorts [16]. Raw sequencing reads in FASTQ format were aligned to the GRCh37 build of the human reference genome with BWA-mem [17]. Then the aligned reads in BAM format were sorted, indexed and marked with duplicate reads with Picard Tools. Reads containing indels were realigned with GATK's IndelRealigner tool. Next, GATK was used to perform Base Quality Score Recalibration (BQSR). Quality control after alignment was performed using Picard (v2.20.4) CollectAlignmentSummaryMetrics (Supplementary Table 2). Variant calling was performed across all samples with GATK (v4.1.1.0) HaplotypeCaller. For detailed variant calling pipeline, please refer to Supplementary Figure 2.

Kinship analysis, DNM calling and annotation

We used PLINK to perform kinship analysis [18]. Briefly, for a given family, if the father was shown to be within the third degree relative to the mother, the family would be excluded; if the child was not shown to be the first degree relative to the parents, the child would be excluded.

To ensure the DNM calls with high confidence, three tools, GATK PhyseByTransmission (PBT), TrioDeNovo (v0.0.6) and DeNovoGear (v.develop), were used to evaluate the calls [19-21]. Only DNMs called by all three tools consistently were considered as candidate DNMs.

PBT was run with the following command:

```
java –jar GenomeAnalysisTK.jar -T PhaseByTransmission -R human_g1k_v37_decoy.fasta -V parent_child_trio.vcf -prior 1.0e-8 -mvf PBT_result.vcf
```

TrioDeNovo was run with the following command:

```
triodenovo --ped parent_child_pedigree.ped --in_vcf parent_child_trio.vcf --mu 1.0e-8 --out_vcf TrioDeNovo_result.vcf
```

DeNovoGear was run with the following command:

```
dng dnm --ped parent_child_pedigree.ped –vcf parent_child_trio.vcf -s 1.0e-8 --write DeNovoGear_result.vcf
```

After obtaining consistently called DNMs, the following five criteria were applied to retain high quality calls: (1) quality score is greater than or equal to 30; (2) genotypes of the parents are homozygotes; (3) genotype of the child is heterozygote; (4) phred-scaled maximum likelihood of heterozygote for the parents is less than 50; and (5) phred-scaled maximum likelihood of heterozygote of the child is 0. DNMs with MAF greater than 0.01 in East Asian populations were further excluded by querying gnomAD (https://gnomad.broadinstitute.org/) and EXac (http://exac.broadinstitute.org/). Please refer to Supplementary Figure 2 for detailed pipeline of DNM calling. Remaining DNMs were annotated using ANNOVAR [22].

Polymerase chain reaction (PCR)-based Sanger sequencing validation

For the families where candidate DNMs were found, DNA from all members of the family was subjected to PCR-based Sanger sequencing by capillary electrophoresis according to standard molecular biology practices (ABI 3130 genetic analyzer, ThermoFisher Scientific). Primer3Plus was used to design the PCR primers [23]. For the *GJC1* mutation, the forward primer sequence was 5’-TTAGGTTTGGGTTGGCTCTG -3’ and the reverse primer sequence was 5’-CACGGTGAAGCAGACAAGAA -3’. For the *HIST1H2AD* insertion, the forward primer

sequence was 5'- CTCGTTTTACTTGCCCTTGG -3' and the reverse primer sequence was 5'-ACAACAAGAAGACCCGCATC -3'.

Reactions were performed on an Eppendorf MasterCycler (Eppendorf North America, New York, USA) under the following cycling conditions: denaturation at 95 °C for 3 min, 35 cycles of 95 °C for 15 sec, 55 °C for 20 sec, 72 °C for 30 sec, and a final extension at 72 °C for 5 min. Sanger sequencing data was then analyzed using Chromas software (https://technelysium.com.au/wp/).

Identification of genes implicated by coding DNMs in SCZ

As the spontaneous background mutation rates vary greatly between genes, those carrying relatively more protein-coding DNMs in cases are not necessarily implicated by DNMs [24]. To identify those genes implicated by protein-coding DNMs or of which the DNMs occurred higher than the background mutation rates, the R package denovolyzeR was used to analyze protein-coding DNMs based on a mutation model developed previously [24]. Briefly, denovolyzeR estimates underlying mutation rate based on trinucleotide context and incorporates exome depth and divergence adjustments based on macaque-human comparisons over a ± 1-Mb window and accommodates known mutational biases, such as CpG hotspots. By applying the underlying mutation rate estimates, denovolyzeR generates prior probabilities for observing a specific number and class of mutations (synonymous, missense, nonsense, splice-site and frameshift) for a given gene.

DNM burden test

To determine whether some genes carry more DNMs in SCZ patients than expected by chance, we performed the DNM burden test for each gene potentially implicated by DNM. Human brain-

specific gene enhancer information was included from PsychENCODE (http://resource.psychencode.org/) and human-specific gene promoter information from Eukaryotic Promoter Database (EPD, https://epd.epfl.ch//index.php) [25, 26]. For a given gene, DNMs occurred in its gene body, brain-specific gene enhancer and gene promoter regions were considered as burdens. For a given gene, we compared the number of DNMs mapped to these regions in SCZ patients with the number of DNMs mapped to these regions in unaffected siblings, and then assessed the significance of the comparison using 10,000 within-sibship case-control label-swapping permutations. *P* value was calculated as the proportion of permutations with relative risk (RR) as or more extreme than in the observed data.

Developmental expression of DNM-implicated genes

Multiple lines of evidence showed prenatal maternal infection, malnutrition and stress are risk factors for SCZ [27-29]. The neurodevelopmental model of SCZ posits that a perturbation in early brain development leads to an altered brain developmental trajectory that is sensitive to molecular changes associated with development and environmental experience, consequently converging on the emergence of SCZ in early adulthood [30]. It was hypothesized that DNMs drove dysfunction of genes in early brain development and that this dysfunction confers risks for subsequent SCZ. To determine whether DNM-implicated genes (loss-of-function DNM genes and DNM-burdened genes) were involved in early brain development, human brain developmental expression data from BrainCloud and BrainSpan was evaluated [31, 32]. The expression data of BrainCloud consisted of data derived from the prefrontal cortices of 269 individuals [31], whereas the expression data of BrainSpan consisted of 42 brain specimens across 13 developmental states in 8-16 brain structures [32]. BrainCloud expression data was examined by comparing the expression of a DNM gene in prenatal stage to the expression in postnatal stage

and BrainSpan expression data was evaluated by leveraging the developmental effect scores curated in a previous publication [33]. A developmental effect score measures the effect of age on expression per gene per brain structure, with a higher developmental effect score of a gene in a given brain structure indicating a stronger involvement of brain development in that structure. By 10,000 gene label-swapping permutation, *P* value of a DNM gene for a given brain structure was calculated as the proportion of permutations with developmental effect scores as or more extreme than the observed value. DNM genes were then further evaluated in brain developmental expression data from other species (macaque and mouse) [34, 35].

Genetic susceptibilities of DNM-implicated genes in SCZ and SCZ-related traits

To determine whether DNM-implicated genes are among the loci found by the genome-wide association study (GWAS) of SCZ or SCZ-related traits, we used GWAS summary statistics from SCZ [1], ASD [36], ADHD [37], bipolar disorder (BD) [38], major depressive disorder (MDD) [39], intelligence [40], educational attainment (EA) [41], cognitive performance (CP) [41], and smoking and drinking [42]. The criterion for annotating a single nucleotide polymorphism (SNP) to a given gene was that the GWAS SNP is located within 10 kb of the gene boundaries.

Co-expression and enrichment analysis

To explore the potential functions or pathological pathways affected by DNM-implicated genes, expression data from BrainCloud was leveraged by retrieving genes highly co-expressed with the candidate DNM genes. To identify enrichments in gene ontologic features, biological pathways, diseases and drug targets, WebGestalt was used (http://www.webgestalt.org/) [43]. In WebGestalt, genes co-expressed with the candidate DNM gene were input as the target gene set and all genes

in the BrainCloud expression data were input as the reference gene set. Enrichments with B-H FDR-corrected *P* values less than 0.05 were considered significantly enriched.

Cell-type specific expression analysis

To examine cell-type specific expression of DNM-implicated genes, brain tissue single nucleus RNA-seq (snRNA-seq) data of middle temporal gyrus (MTG), of which the cells have been sub-typed from Allen Brain Atlas (https://celltypes.brain-map.org/rnaseq) and single cell expression data from PsychENCODE were utilized. Single cell expression data from PsychENCODE were merged from multiple brain regions, including frontal cortex, visual cortex and cerebellar hemisphere [44]. Raw read count data was normalized by log-transformation using R package Seurat [45]. For a given DNM gene, the dominant cell type(s) with high expression were determined by pair-wise Wilcoxon test.

Results

The goal of our study was to identify DNM disturbed genes, of which the dysfunctions can contribute to the pathogenesis of SCZ. We identified coding DNM-implicated genes and genes carrying more DNM burdens in SCZ patients than their unaffected siblings. These DNM-implicated genes were detrimental and subsequent analyses were performed to explore their pathological effects (Figure 1).

Identification of DNMs and genes implicated by DNMs

To ensure parental unrelatedness and that the children are indeed biological offspring of their parents, kinship analysis was performed for each individual family. For the Taiwan cohort, one

family (Family ID: 35-04560) was excluded as the father is within the third degree relative of the mother, one family (Family ID: 35-93405) was excluded as the father was not found to be related to any of the children and one child in one family was excluded as he was found to be unrelated to his parents (Individual ID: 35-02497-01) (Supplementary Table 1). Three children from Shanghai cohort were excluded as they were unrelated to their parents (Individual IDs: CHG000225, CHG000236 and CHG000246) (Supplementary Table 1).

In this study, 70.71 ± 6.83 *de novo* point mutations and 6.31 ± 3.64 *de novo* indel mutations with high confidence per individual were identified. The observed *de novo* point mutation rate of 1.145×10^{-8} was consistent with the neutral expectation of 1.140×10^{-8} ($P = 0.95$, two-sided exact binomial test) [21]. The observed *de novo* indel mutation rate of 1.022×10^{-9} was consistent with the neutral expectation of 1.420×10^{-9} ($P = 0.50$, two-sided exact binomial test) [21]. All DNMs were checked visually by Integrative Genomics Viewer. Overall, the DNMs occurred in non-coding regions predominantly in intergenic regions (Supplementary Figure 3). No obvious difference was observed between unaffected siblings and SCZ patients with regards to the distributions of DNM locations relative to genes (Supplementary Figure 3). The nonsynonymous-to-synonymous ratio in SCZ patients was not found to be different from the ratio in unaffected siblings [46], which might be attributed to the limited sample size. Two nonsense loss-of-function DNMs implicating *GJC1* and *HIST1H2AD* respectively were identified in SCZ patients (Table 1 and Supplementary Table 4). The two loss-of-function DNMs were confirmed by Sanger sequencing (Supplementary Figure 6 and 7) and visual verification is shown in Supplementary Figures 4 and 5. To examine whether the occurrence rates of nonsense loss-of-function DNMs in *GJC1* and *HIST1H2AD* in SCZ patients were higher than expected by chance, we compared the nonsense mutation probabilities of *GJC1* and *HIST1H2AD* to those of

all genes calculated in Samocha *et al.* [47]. Nonsense DNMs were not prone to occur in DNA regions of *GJC1* and *HIST1H2AD* (Supplementary Figure 8). No loss-of-function DNMs were identified in unaffected siblings on the whole genome scale (Supplementary Table 3).

As DNMs occurred in coding regions may change protein structures physically, we focused on the DNMs in coding regions to find damaging DNMs (Figure 1 and Supplementary Table 3). A detailed list of these DNMs are shown in Supplementary Table 4. To identify genes implicated by these protein-coding DNMs, denovolyzeR was used. Due to the relatively small sample size of this study, Bonferroni correction was used to reduce the false discovery rate. Only *GJC1* and *HIST1H2AD* carrying nonsense DNMs were significantly identified to be implicated by DNMs (Supplementary Table 5). We applied another background mutation rate-based model, referred to as the chimpanzee-human divergence model, which also showed *GJC1* and *HIST1H2AD* were significantly implicated by the nonsense DNMs (P value = 1.56×10^{-5} for *GJC1* and P value = 1.82×10^{-5} for *HIST1H2AD*) [48]. Of note, the *GJC1* nonsense DNM was predicted to be extremely close to the most severe 0.1% of mutations by combined annotation dependent depletion (CADD) (Supplementary Table 4; mutations with CADD phred-like scores greater than or equal to 30 is the most severe 0.1% of mutations) [49].

Identification of genes burdened with DNMs in SCZ patients

99% of the DNMs occurred in non-coding regions [10]. DNMs occurring in regulatory regions can potentially disturb the bindings of transcriptional factors and thus influence gene expression. If a gene or regulatory regions of the gene carry more DNMs in SCZ patients than their unaffected siblings, this gene may have a role in disease predisposition. Based on this assumption, for a given gene, we compared the number of DNMs that occurred in brain-specific enhancer, promoter and gene body in SCZ patients to the number of DNMs in unaffected siblings by

within-sibship case-control label-swapping permutation. Before calculation, to reduce the falsely identified genes by chance, we excluded genes with low total DNM counts in both SCZ patients and unaffected siblings (≤ 3). Due to the relatively small sample size, Bonferroni correction was used to reduce the false discovery rate. Ten genes, *DPYSL2*, *NBPF1*, *SDK1*, *ZNF595*, *ZNF718*, *GCNT2*, *SNX9*, *AACS*, *KCNQ1* and *MSI2*, were identified to carry more DNMs in SCZ patients than their unaffected siblings. Multiple DNMs implicated enhancers for *DPYSL2*, *NBPF1*, *SNX9* and *MSI2* (Supplementary Table 6), suggesting the dysregulation of these genes, especially in the prenatal stage, may predispose an individual to SCZ. One single enhancer can regulate the expression of multiple genes. Therefore, single DNM implicating an individual enhancer can disturb the expression of multiple genes. We leveraged transcription factor (TF)-enhancer-target gene linkage data from PsychENCODE to establish the gene regulatory network disturbed by the DNMs implicating the enhancers identified by DNM burden test (Figure 2). For detailed list of the DNMs found in the ten genes reported here, please refer to Supplementary Table 6. Interestingly, no DNM occurred in promoters of the ten genes.

DNMs occurred in genes involving in early brain development

In the following analyses, coding DNM-implicated genes (*GJC1* and *HIST1H2AD*, Supplementary Table 5) and DNM-burdened genes (Table 2) were served as detrimental DNM genes to explore their potential functions that are potentially related to SCZ.

Early neurodevelopmental events have been implicated in pathogenesis of SCZ [30]. Genes implicated in SCZ function in processes important to fetal brain development [46, 50]. To determine whether the detrimental DNM genes, including *GJC1* and *HIST1H2AD* implicated by nonsense mutations (Supplementary Table 5) and genes carrying more DNMs in SCZ patients (Table 2), involved in brain development, we leveraged the developmental expression data of prefrontal

cortices (PFC), one of the most highly implicated brain regions in SCZ, from BrainCloud. Nine of the 13 detrimental DNM genes, including the two loss-of-function DNM-implicated genes *GJC1* and *HIST1H2AD*, and seven DNM-burdened genes *DPYSL2*, *NBPF1*, *SDK1*, *ZNF595*, *ZNF718*, *KCNQ1* and *SNX9*, showed biased higher expression in prenatal stage compared to that of the expression in postnatal stage (Figure 3). The proportion of 9/12 detrimental DNM genes with higher expression in prefrontal cortices was more than expected by chance (Supplementary Figure 9, $P = 0.05$, hypergeometric test). We next sought to examine whether the involvements of detrimental DNM genes in PFC development were preserved in other brain regions. The genes with available developmental data in BrainSpan were analyzed. It was found that *GJC1*, *SDK1* and *GCNT2* were still involved in the developments of other brain regions (Supplementary Figure 10). To examine whether the involvements of detrimental DNM genes in human brain development were conserved across species, we first queried the brain developmental expression data of macaque, a primate species evolutionarily close to *Homo sapiens*. Only DNM genes with expression data of macaque were analyzed. *GJC1*, *HIST1H2AD*, *DPYSL2*, *SDK1* and *MSI2* showed higher expression in prenatal stage than the expression in postnatal stage (Supplementary Figure 11). We then queried the brain developmental expression data of mouse. Since the developmental data of mouse were limited in postnatal stage, we applied linear regression to examine the change of DNM gene expression level with time. The expression levels of *GJC1*, *DPYSL2*, *SDK1* and *AACS* significantly decreased with time (Supplementary Figure 12), which implied their involvements in prenatal neurodevelopment of mouse. Specifically for *GJC1*, Leung et al. showed that *GJC1* displayed high expression in embryonic stage followed by a massive postnatal decrease in the rat midbrain-floor, where dopaminergic neurons were mostly populated [51].

Common variants in detrimental DNM genes may influence risks of SCZ-associated traits

To determine whether the detrimental DNM genes are involved more broadly in SCZ, we asked whether common variants present in the DNM genes confer risk for SCZ and SCZ-associated traits. We extracted all SNPs that occur within 10 kb of the DNM genes. For a given trait, the SNP with the minimum *P* value mapped to the gene was used to represent the gene risk on the trait. Genes with minimum SNP *P* values less than the suggestive threshold of 1×10^{-5} were considered as risk genes for the trait. Interestingly, except for *HIST1H2AD* close to HLA locus [52], the DNM genes influence risk for SCZ-associated traits, including intelligence, educational attainment, cognitive performance, smoking and drinking primarily (Supplementary Table 7). Of note, *GJC1*, *HIST1H2AD* and *SDK1* even confer risks for multiple SCZ-associated traits with SNPs passed the genome-wide significance threshold of 5×10^{-8}.

GJC1 co-expressed with multiple potassium channel genes and is a potentially target for SCZ

Next, we focused on the loss-of-function DNM gene *GJC1*, a member of the connexin gene family, which showed strong evidence of SCZ susceptibility (Supplementary Table 8). DNMs in other brain-associated connexin genes were also found in psychiatric patients (Supplementary Table 9). *GJC1* showed higher expression in prenatal stage than the expression in postnatal stage in PFC and multiple other brain regions, and its involvement in early brain development was conserved across species (Figures 3; Supplementary Figures 10, 11 and 12). Moreover, common variants present in *GJC1* confer risks for intelligence, educational attainment, cognitive performance and alcohol abuse, which were associated with SCZ (Supplementary Table 7) [53-56].

To explore the pathogenic effects of *GJC1*, we first sought to identify the genes that might be impacted by the dysfunction of *GJC1*. We performed co-expression analysis with Pearson

correlation coefficient greater than 0.8 or less than -0.8 in PFC to identify these genes (Figure 4A and supplementary Table 10). Then we leveraged the genes co-expressed with *GJC1* to perform enrichment analysis to identify functions or pathways involved by *GJC1*. Interestingly, the genes co-expressed with GJC1 were enriched in SCZ, SCZ drug zuclopenthixol and SCZ-associated functions or pathways, including potassium ion transport (Figure 4B and Supplementary Table 11). Of note, multiple potassium ion channel genes were negatively co-expressed with *GJC1* (Figure 4A), suggesting their dysfunctions may be subsequent to the dysfunction of *GJC1*.

To determine whether the expression of *GJC1* was specific to a certain cell type, we performed cell type specific analysis of *GJC1* expression. *GJC1* was predominantly expressed in glutamatergic excitatory neurons (Figure 4C and 4D, refer to Supplementary Tables 12 and 13 for pair-wise *P* values among different cell types derived from Wilcoxon rank sum test). This should be interesting as glutamatergic dysfunction has long been implicated in SCZ [57].

Discussion

In this study, we used two matched SCZ-sibling family cohorts of Han Chinese origin to investigate DNMs and DNM-implicated genes in SCZ. By integrating this information with publicly available data, including brain developmental expression profiles and summary GWAS statistics, we identified DNM-implicated genes that were involved in fetal neurodevelopment. These genes may confer risks for SCZ or SCZ-associated traits, and are potential drug therapy targets for SCZ.

Connexin 45 (Cx45) encoded by *GJC1* is a component of the gap junction channel. It is one of the two connexin genes expressed particularly in neurons [58]. Our results showed *GJC1* was

expressed predominantly in glutamatergic excitatory neurons in the human brain (Figure 4C and 4D). Particularly for the nonsense mutation case in our study, its occurrence might trigger the activation of surveillance pathway nonsense-mediated mRNA decay (NMD), which reduces aberrant proteins to be formed [59]. The reduction of *GJC1* may perturb the co-expressed genes and thus influence the functions and pathways involved (Figure 4A and 4B). Krüger *et al*. reported Cx45-deficient embryos exhibited striking abnormalities in vascular development and died between embryonic day 9.5 and 10.5 [60]. Kumai *et al*. reported Cx45-deficient embryos displayed an endocardial cushion defect in early cardiogenesis and died of heart failure at around embryonic day 9 [61]. Nishii *et al*. showed mice lacking Cx45 conditionally in cardiac myocytes displayed embryonic lethality and the requirement of Cx45 for developing cardiac myocytes [62]. This evidence indicates the critical role of Cx45 in embryonic development. Our results showed that *GJC1* displayed biased high expression across multiple brain regions in prenatal stage, suggesting it is required for neurodevelopment. The high expression of *GJC1* in prenatal/early stage in other species indicated its involvement in development is phylogenetically conserved. The dysfunction or loss-of-function can cause lethal effects, thereby such mutation would be eliminated by natural selection in evolution. Interestingly, all potassium ion channel genes were negatively co-expressed with *GJC1*. Therefore, the reduction of *GJC1* indicates an increase or overactivity of potassium channels. Miyake *et al*. reported overexpression of Ether-a`-go-go potassium channel gene *KCNH3* in the forebrain can impair the performances of working memory, reference memory and attention [63]. Ghelardini *et al*. reported administrations of potassium channel openers bring out amnesic effect which can be reversed by potassium channel blockers [64]. This evidence suggested the cognitive or memory deficits in the particular case were caused by increased expression of potassium channel genes subsequent to the reduction of *GJC1*

activity (Figure 5). The various connexins of gap junctions may be capable of differentiating between the operation qualities of the cognate synapses defined by the neurotransmitter types [65]. Mitterauer *et al*. raised a hypothesis that if the function of glial gap junction proteins is lost, the brain is incapable of distinguishing between the same and different qualities of information processing and thus cause severe cognitive impairments in SCZ [65]. Our findings supports these hypotheses. Overall, the loss-of-function of Cx45 may be a driver for pathogenesis of some SCZ cases.

HIST1H2AD, the other nonsense DNM gene close to HLA locus, encodes the member D in histone cluster 1 H2A family. Studies have demonstrated that several SCZ candidate genes are especially susceptible to changes in transcriptional activity as a result of histone modification [66, 67].Therefore, a deficit in the histone itself can implicate multiple SCZ candidate genes. Moreover, epigenetic regulation effects of histone deacetylase inhibitors were potentially suggested to treat SCZ [68], which implies the fundamental role of histone in the pathology of SCZ.

Gene level DNM burden test identified a list of genes carrying more DNMs in SCZ patients than unaffected siblings. Here we show the experimental or clinical evidences demonstrating their potential association with SCZ. One of the DNM-burdened genes, *DPYSL2*, is a member of the collapsin response mediator protein (CRMP) family. CRMP forms homo- and hetero-tetramers and facilitates neuron guidance, growth and polarity. It also plays a role in synaptic signaling through interactions with calcium channels. Lee *et al*. revealed that *DPYSL2* was downregulated in the PFC and hippocampus of prenatally stressed (PNS) adult rats that underperformed in behavioral tests [69]. Bruce *et al*. conducted potassium channel-targeted SNP association analyses with SCZ and SCZ-associated phenotypes. rs8234 in *KCNQ1*, a DNM-burdened gene in our study, was associated to processing speed [70]. Geschwind *et al*. found that patients with voltage-

gated potassium channel complex antibody (VGKCC-Abs) had particular impairment in memory and executive functions when they evaluated cognitive function and imaging data in patients with VGKCC-Abs associated encephalopathy [71]. Interestingly, the impaired brain functions coincide with the two most implicated brain regions in SCZ, dorsolateral prefrontal cortex and the hippocampus [72]. Luan et al. reported SNPs in *MSI2*, another DNM-burdened gene, are strongly associated with SCZ in the Chinese population [73]. In the GWAS of PGC, the SNPs in *MSI2* were also associated with SCZ (Supplementary Table 7). This evidence demonstrated the efficacy of our DNM burden test by incorporating enhancer and promoter regions into consideration. It also indicated DNMs contributing to the genesis of disease are not limited in coding regions, underling the value of WGS.

Except for *HIST1H2AD*, common variants in these identified DNM genes were not strongly associated to psychiatric diseases. However, a few of them, *GJC1*, *HIST1H2AD* and *SDK1*, were strongly associated to SCZ-associated traits, including intelligence, educational attainment, smoking and drinking. It suggested the significance of these DNM genes in neurodevelopment and that the lethal mutations occurred in these genes were eliminated by marked reduction of fecundity in psychiatric patients.

The associations of these identified DNM genes, especially *GJC1*, with SCZ or related traits were strongly supported by subsequent functional analysis. However, the results should be interpreted with caution due to the limited samples in the pilot study. More samples with family design sequenced at whole genome scale should be incorporated and functional experiments should be carried out to validate our findings in the future.

In summary, we identified a list of DNM-implicated genes which are involved in prenatal neurodevelopment. Common variants in these DNM-implicated genes had been previously

reported to be associated with SCZ and related traits in previous GWASs, which is consistent with our analyses. DNMs implicating the enhancers may contribute to pathogenesis of SCZ by dysregulating the expression of genes. *GJC1* is one of these DNM-implicated genes that is primarily expressed in glutamatergic neurons and may be involved in the modulation of ion channel functions, which has also been implicated in SCZ in previous studies [74, 75]. Overall, our study provided new evidence that DNMs have a significant role in SCZ. Further study of these DNM-implicated genes with functional analyses could lead to better understanding of the pathology of SCZ.

Acknowledgements

This work was supported in part by grants from National Institutes of Health (R01MH101054 to X.C. and R01LM012806 to Z.Z.), the National Natural Science Foundation of China (grant 81421061), the National Key Research and Development Program (2016YFC0906400), Shanghai Key Laboratory of Psychotic Disorders (13dz2260500), Cancer Prevention and Research Institute of Texas (RR180012 to X.J.), and UT Stars award to X.J. The DNA samples of the subjects from Taiwan cohort were obtained through NIMH Genetics Repository. The DNA samples of the subjects from Shanghai cohort were from the Bio-x SCZ Biobank in Shanghai, China. Computational resources from the school of biomedical informatics at The University of Texas Health Science Center at Houston were used in data analysis. The data from the TSLS were collected with funding from grant R01MH59624 from K. S. K. and M. T.. The whole genome sequencing of samples from Taiwan cohort were supported by grant from National Institutes of Mental Health (1RO1-MH085560) to M.T.. We acknowledge Dr. Hai-Gwo Hwu and Dr. Wei J. Chen for their recruitment of families, collection of clinical data and

preprocess of blood samples in Taiwan cohort. We acknowledge the help from Dr. Lukas Simon on single cell analysis and the collection of GWAS summary statistics from Dr. Yulin Dai. We acknowledge the altruism of the participants and their families and support staff at each of the participating sites for their contributions to this study.

Author contributions

S.J. designed the study, performed the analyses, interpreted the results and wrote the manuscript. D.Z. collected the demographic data and performed whole genome sequencing, data processing, quality control and cleaning from Shanghai cohort. Y.W., P.J., C.W., X.L., G.H., D.C., X.J., T.M. and J.S.W. contributed to data processing, quality control and cleaning. K.S.K. and M.T. contributed to data collection and whole genome sequencing. J.S.W., Y.L. and J.C. conducted Sanger sequencing experiment. T.M. revised the manuscript. L.H., J.C., Z.Z., P.J. and X.C. conceived the project, designed the study, collected the data, interpreted the results and wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Figure legends

Figure 1. **Schematic of genetic data processing, DNM identification and functional analysis in 23 families with schizophrenia patients and matched unaffected siblings.** SNV, single nucleotide variant; INDEL, insertion and deletion.

Figure 2. **Gene regulatory network disturbed by the DNMs implicating the enhancers identified by DNM burden test.** Light green arrow nodes are the transcription factors. The yellow round nodes are the enhancers perturbed by the DNMs. The Orange round nodes are the genes. The edge between transcription factor and enhancer represent the transcription factor can bind to the enhancer without the perturbation of DNM. The edge between enhancer and gene represent the enhancer can enhance the expression of the gene without the perturbation of DNM. Red labeled genes were those identified by DNM burden test.

Figure 3. **Detrimental DNM genes implicated in early brain development in PFC.** A, *GJC1*; B, *HIST1H2AD*; C, *DPYSL2*; D, *NBPF1*; E, *SDK1*; F, *ZNF595*; G, *ZNF718*; H, *KCNQ1*; I, *SNX9*. Black dots represent samples from prenatal stage. Blue dots represent samples from postnatal stage. *P* values were derived from Wilcoxon rank sum test by comparing gene expression from prenatal stage to postnatal stage.

Figure 4. **Genes co-expressed with *GJC1* were enriched in schizophrenia and schizophrenia-associated pathways and drug.** A, *GJC1*-hubed co-expression network. The nodes of genes positively co-expressed with *GJC1* ($r_{pearson} > 0.8$) were labeled in red and the nodes of genes negatively co-expressed with *GJC1* ($r_{pearson} < -0.8$) were labeled in blue. The names of potassium channel genes co-expressed with *GJC1* were marked in red. B, significantly enriched terms by genes co-expressing with *GJC1*. C. Cell-type specific expression of *GJC1* in

middle temporal gyrus of human from Allen Brain Atlas. D. Cell-type specific expression of *GJC1* in mixed brain regions, including frontal cortex, visual cortex and cerebellum hemisphere, from PsychENCODE.

Figure 5. **Schematic illustration of possible mechanism of dysfunction of *GJC1* leads to SCZ.** Loss of gap junction formed by connexin 45 blocked the pass of ions and small molecules between brain cells, which caused the state of homeostatic imbalance in cell. Increased potassium channel subsequent to the loss of gap junction leads to cognitive impairment and memory loss in SCZ patients.

References

1. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510)**:** 421-427.

2. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* 2016; **530**(7589)**:** 177-183.

3. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017; **49**(1)**:** 27-35.

4. 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.

5. Kim DS, Burt AA, Ranchalis JE, Wilmot B, Smith JD, Patterson KE *et al.* Sequencing of sporadic Attention-Deficit Hyperactivity Disorder (ADHD) identifies novel and potentially pathogenic de novo variants and excludes overlap with genes associated with autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet* 2017; **174**(4)**:** 381-389.

6. Epi KC. De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. *Am J Hum Genet* 2016; **99**(2)**:** 287-298.

7. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011; **43**(9)**:** 864-868.

8. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011; **43**(9)**:** 860-863.

9. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**(7487)**:** 179-184.

10. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I *et al.* Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet* 2015; **47**(7)**:** 822-826.

11. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; **8**(10)**:** 749-761.

12. Hwu HG, Faraone SV, Liu CM, Chen WJ, Liu SK, Shieh MH *et al.* Taiwan schizophrenia linkage study: the field study. *Am J Med Genet B Neuropsychiatr Genet* 2005; **134B**(1)**:** 30-36.

13. Faraone SV, Hwu HG, Liu CM, Chen WJ, Tsuang MM, Liu SK *et al.* Genome scan of Han Chinese schizophrenia families from Taiwan: confirmation of linkage to 10q22.3. *Am J Psychiatry* 2006; **163**(10)**:** 1760-1766.

14. Chen WJ, Hsiao CK, Hsiao LL, Hwu HG. Performance of the Continuous Performance Test among community samples. *Schizophr Bull* 1998; **24**(1)**:** 163-174.

15. Chen J, Wu JS, Mize T, Moreno M, Hamid M, Servin F *et al.* A Frameshift Variant in the CHST9 Gene Identified by Family-Based Whole Genome Sequencing Is Associated with Schizophrenia in Chinese Population. *Sci Rep* 2019; **9**(1)**:** 12717.

16. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **43:** 11 10 11-33.

17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**(14)**:** 1754-1760.

18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3)**:** 559-575.

19. Francioli LC, Cretu-Stancu M, Garimella KV, Fromer M, Kloosterman WP, Genome of the Netherlands c *et al.* A framework for the detection of de novo mutations in family-based sequencing data. *Eur J Hum Genet* 2017; **25**(2)**:** 227-233.

20. Wei Q, Zhan X, Zhong X, Liu Y, Han Y, Chen W *et al.* A Bayesian framework for de novo mutation calling in parents-offspring trios. *Bioinformatics* 2015; **31**(9)**:** 1375-1381.

21. Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, Cartwright RA *et al.* DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat Methods* 2013; **10**(10)**:** 985-987.

22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**(16)**:** e164.

23. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 2007; **35**(Web Server issue)**:** W71-74.

24. Ware JS, Samocha KE, Homsy J, Daly MJ. Interpreting de novo Variation in Human Disease Using denovolyzeR. *Curr Protoc Hum Genet* 2015; **87:** 7 25 21-15.

25. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* 2018; **362**(6420).

26. Dreos R, Ambrosini G, Perier RC, Bucher P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res* 2015; **43**(Database issue)**:** D92-96.

27. Khandaker GM, Zimbron J, Lewis G, Jones PB. Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol Med* 2013; **43**(2)**:** 239-257.

28. Susser E, St Clair D, He L. Latent effects of prenatal malnutrition on adult health: the example of schizophrenia. *Ann N Y Acad Sci* 2008; **1136:** 185-192.

29. Khashan AS, Abel KM, McNamee R, Pedersen MG, Webb RT, Baker PN *et al.* Higher risk of offspring schizophrenia following antenatal maternal exposure to severe adverse life events. *Arch Gen Psychiatry* 2008; **65**(2)**:** 146-152.

30. Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci* 2017; **18**(12)**:** 727-740.

31. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 2011; **478**(7370)**:** 519-523.

32. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A *et al.* Transcriptional landscape of the prenatal human brain. *Nature* 2014; **508**(7495)**:** 199-206.

33. Grote S, Prufer K, Kelso J, Dannemann M. ABAEnrichment: an R package to test for gene set expression enrichment in the adult and developing human brain. *Bioinformatics* 2016; **32**(20)**:** 3201-3203.

34. Fertuzinhos S, Li M, Kawasawa YI, Ivic V, Franjic D, Singh D *et al.* Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. *Cell Rep* 2014; **6**(5)**:** 938-950.

35. Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L *et al.* A comprehensive transcriptional map of primate brain development. *Nature* 2016; **535**(7612)**:** 367-375.

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

37. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019; **51**(1)**:** 63-75.

38. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019; **51**(5)**:** 793-803.

39. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**(5)**:** 668-681.

40. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 2018; **50**(7)**:** 912-919.

41. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018; **50**(8)**:** 1112-1121.

42. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; **51**(2)**:** 237-244.

43. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 2017; **45**(W1)**:** W130-W137.

44. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* 2018; **36**(1)**:** 70-80.

45. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; **36**(5)**:** 411-420.

46. 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. *Nat Genet* 2012; **44**(12)**:** 1365-1369.

47. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 2014; **46**(9)**:** 944-950.

48. 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.

49. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**(3)**:** 310-315.

50. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci* 2012; **15**(12)**:** 1723-1728.

51. Leung DS, Unsicker K, Reuss B. Expression and developmental regulation of gap junction connexins cx26, cx32, cx43 and cx45 in the rat midbrain-floor. *Int J Dev Neurosci* 2002; **20**(1)**:** 63-75.

52. Mokhtari R, Lachman HM. The Major Histocompatibility Complex (MHC) in Schizophrenia: A Review. *J Clin Cell Immunol* 2016; **7**(6).

53. Aukes MF, Alizadeh BZ, Sitskoorn MM, Kemner C, Ophoff RA, Kahn RS. Genetic overlap among intelligence and other candidate endophenotypes for schizophrenia. *Biol Psychiatry* 2009; **65**(6)**:** 527-534.

54. Le Hellard S, Wang Y, Witoelar A, Zuber V, Bettella F, Hugdahl K *et al.* Identification of Gene Loci That Overlap Between Schizophrenia and Educational Attainment. *Schizophr Bull* 2017; **43**(3)**:** 654-664.

55. Liddle PF. Schizophrenic syndromes, cognitive performance and neurological dysfunction. *Psychol Med* 1987; **17**(1)**:** 49-57.

56. Nielsen SM, Toftdahl NG, Nordentoft M, Hjorthoj C. Association between alcohol, cannabis, and other illicit substance abuse and risk of developing schizophrenia: a nationwide population based register study. *Psychol Med* 2017; **47**(9)**:** 1668-1677.

57. Olney JW, Farber NB. Glutamate receptor dysfunction and schizophrenia. *Arch Gen Psychiatry* 1995; **52**(12)**:** 998-1007.

58. Yamasaki RJC, Neuroimmunology E. Connexins in health and disease. 2018; **9:** 30-36.

59. Hug N, Longman D, Caceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* 2016; **44**(4)**:** 1483-1495.

60. Kruger O, Plum A, Kim JS, Winterhager E, Maxeiner S, Hallas G *et al.* Defective vascular development in connexin 45-deficient mice. *Development* 2000; **127**(19)**:** 4179-4193.

61. Kumai M, Nishii K, Nakamura K, Takeda N, Suzuki M, Shibata Y. Loss of connexin45 causes a cushion defect in early cardiogenesis. *Development* 2000; **127**(16)**:** 3501-3512.

62. Nishii K, Kumai M, Egashira K, Miwa T, Hashizume K, Miyano Y *et al.* Mice lacking connexin45 conditionally in cardiac myocytes display embryonic lethality similar to that of germline knockout mice without endocardial cushion defect. *Cell Commun Adhes* 2003; **10**(4-6)**:** 365-369.

63. Miyake A, Takahashi S, Nakamura Y, Inamura K, Matsumoto S, Mochizuki S *et al.* Disruption of the ether-a-go-go K+ channel gene BEC1/KCNH3 enhances cognitive function. *J Neurosci* 2009; **29**(46)**:** 14637-14645.

64. Ghelardini C, Galeotti N, Bartolini A. Influence of potassium channel modulators on cognitive processes in mice. *Br J Pharmacol* 1998; **123**(6)**:** 1079-1084.

65. Mitterauer B. Loss of function of glial gap junctions may cause severe cognitive impairments in schizophrenia. *Med Hypotheses* 2009; **73**(3)**:** 393-397.

66. Lewis DA, Hashimoto T, Volk DW. Cortical inhibitory neurons and schizophrenia. *Nat Rev Neurosci* 2005; **6**(4)**:** 312-324.

67. Guidotti A, Auta J, Davis JM, Dong E, Grayson DR, Veldic M *et al.* GABAergic dysfunction in schizophrenia: new treatment strategies on the horizon. *Psychopharmacology (Berl)* 2005; **180**(2)**:** 191-205.

68. Hasan A, Mitchell A, Schneider A, Halene T, Akbarian S. Epigenetic dysregulation in schizophrenia: molecular and clinical aspects of histone deacetylase inhibitors. *Eur Arch Psychiatry Clin Neurosci* 2013; **263**(4)**:** 273-284.

69. Lee H, Joo J, Nah SS, Kim JW, Kim HK, Kwon JT *et al.* Changes in Dpysl2 expression are associated with prenatally stressed rat offspring and susceptibility to schizophrenia in humans. *Int J Mol Med* 2015; **35**(6)**:** 1574-1586.

70. Bruce HA, Kochunov P, Paciga SA, Hyde CL, Chen X, Xie Z *et al.* Potassium channel gene associations with joint processing speed and white matter impairments in schizophrenia. *Genes Brain Behav* 2017; **16**(5)**:** 515-521.

71. Geschwind M, Gelfand J, Irani S, Neuhaus J, Forner S, Bettcher B. Neuropsychological Profiles Of Voltage-Gated Potassium Channel Complex And Other Autoimmune Encephalopathies; More Than Memory Impairment (S18. 005). AAN Enterprises2014.

72. Meyer-Lindenberg AS, Olsen RK, Kohn PD, Brown T, Egan MF, Weinberger DR *et al.* Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Arch Gen Psychiatry* 2005; **62**(4)**:** 379-386.

73. Luan Z, Lu T, Ruan Y, Yue W, Zhang D. The Human MSI2 Gene is Associated with Schizophrenia in the Chinese Han Population. *Neurosci Bull* 2016; **32**(3)**:** 239-245.

74. Pers TH, Timshel P, Ripke S, Lent S, Sullivan PF, O'Donovan MC *et al.* Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes. *Hum Mol Genet* 2016; **25**(6)**:** 1247-1254.

75. Peltola MA, Kuja-Panula J, Liuhanen J, Voikar V, Piepponen P, Hiekkalinna T *et al*. AMIGO-Kv2.1 Potassium Channel Complex Is Associated With Schizophrenia-Related Phenotypes. *Schizophr Bull* 2016; **42**(1)**:** 191-201.

Table 1. Loss-of-function DNMs identified in schizophrenia patients

Individual ID	Chr[a]	Position (hg19)	Gene	Reference allele	Mutant allele	Mutation type	Amino acid substitution
35-50505-02	17	42,882,819	*GJC1*	G	A	Nonsense	p.Q123X
35-06277-01	6	26,199,170	*HIST1H2AD*	A	ACTTTACCCAG	Nonsense	p.V101Afs×2

[a]Chr: chromosome.

Table 2. Genes significantly enriched with higher DNM burden in schizophrenia patients than unaffected siblings

Gene symbol	DNM count in unaffected siblings	DNM count in schizophrenia patients	*P*-value	Bonferroni *P*-value
DPYSL2	0	6	1×10^{-4}	3.4×10^{-3}
NBPF1	0	6	1×10^{-4}	3.4×10^{-3}
SDK1	1	5	1×10^{-4}	3.4×10^{-3}
ZNF595	1	5	1×10^{-4}	3.4×10^{-3}
ZNF718	1	5	1×10^{-4}	3.4×10^{-3}
GCNT2	0	5	1×10^{-4}	3.4×10^{-3}
SNX9	0	5	1×10^{-4}	3.4×10^{-3}
KCNQ1	1	4	1×10^{-4}	3.4×10^{-3}
AACS	1	3	1×10^{-4}	3.4×10^{-3}
MSI2	1	3	1×10^{-4}	3.4×10^{-3}

Figure 1

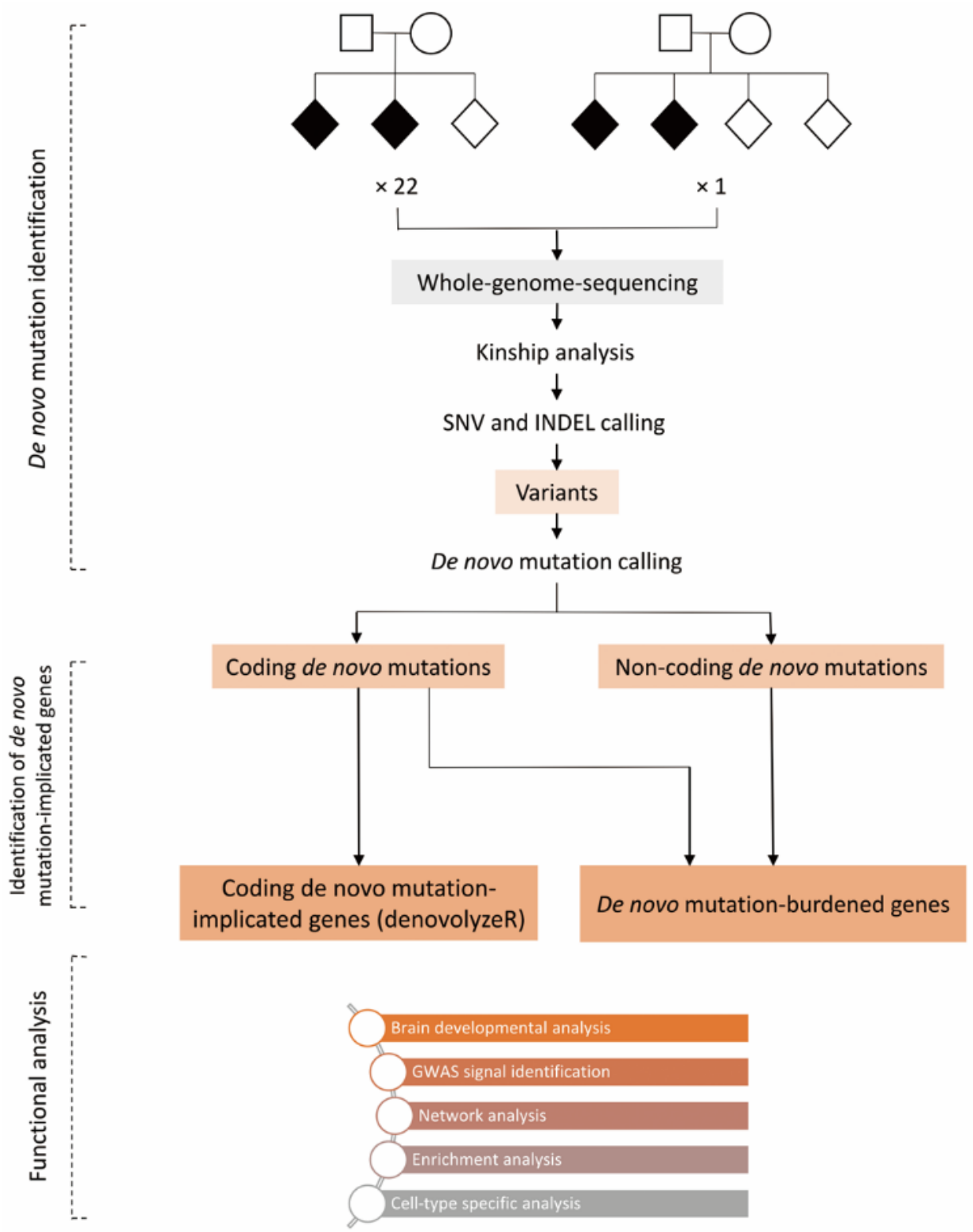

De novo mutation identification
× 22
× 1
Whole-genome-sequencing
Kinship analysis
SNV and INDEL calling
Variants
De novo mutation calling
Identification of de novo mutation-implicated genes
Coding de novo mutations
Non-coding de novo mutations
Coding de novo mutation-implicated genes (denovolyzeR)
De novo mutation-burdened genes
Functional analysis
Brain developmental analysis
GWAS signal identification
Network analysis
Enrichment analysis
Cell-type specific analysis

Figure 2

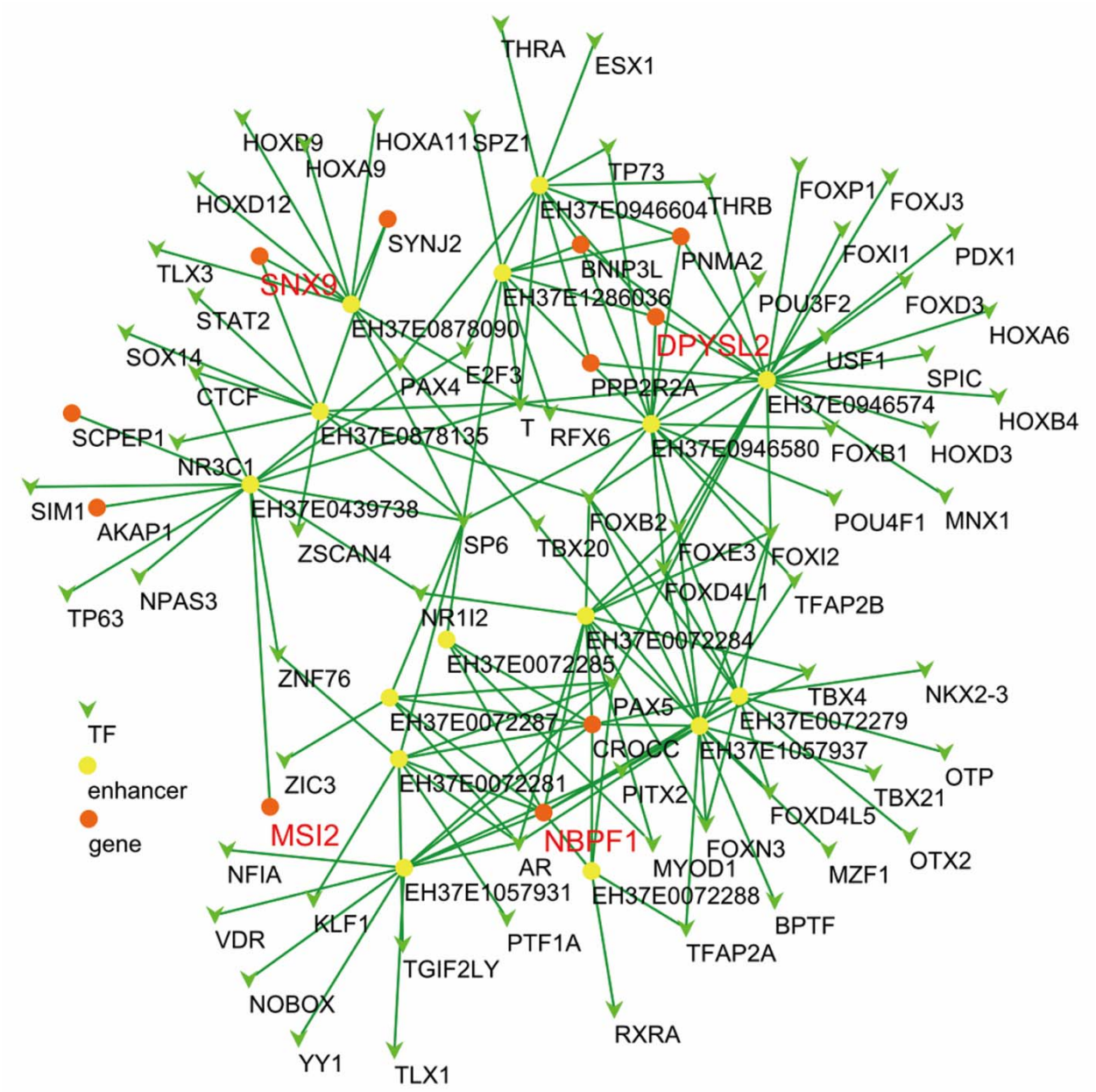

THRA
ESX1
HOXB9
HOXA11
SPZ1
HOXA9
TP73
HOXD12
EH37E0946604
THRB
FOXP1
FOXJ3
SYNJ2
BNIP3L
PNMA2
FOXI1
PDX1
TLX3
SNX9
EH37E1286036
POU3F2
FOXD3
STAT2
EH37E0878090
DPYSL2
HOXA6
SOX14
USF1
E2F3
SPIC
PAX4
PPP2R2A
CTCF
EH37E0946574
HOXB4
SCPEP1
EH37E0878135
T
RFX6
EH37E0946580
FOXB1
HOXD3
NR3C1
SIM1
EH37E0439738
FOXB2
POU4F1
MNX1
AKAP1
ZSCAN4
SP6
TBX20
FOXE3
FOXI2
FOXD4L1
TFAP2B
TP63
NPAS3
NR1I2
EH37E0072284
EH37E0072285
ZNF76
TBX4
NKX2-3
PAX5
EH37E0072287
EH37E0072279
TF
CROCC
EH37E1057937
OTP
enhancer
ZIC3
EH37E0072281
PITX2
TBX21
FOXD4L5
gene
MSI2
NBPF1
FOXN3
OTX2
NFIA
AR
MYOD1
MZF1
EH37E1057931
EH37E0072288
BPTF
VDR
KLF1
PTF1A
TFAP2A
TGIF2LY
NOBOX
RXRA
YY1
TLX1

Figure 3

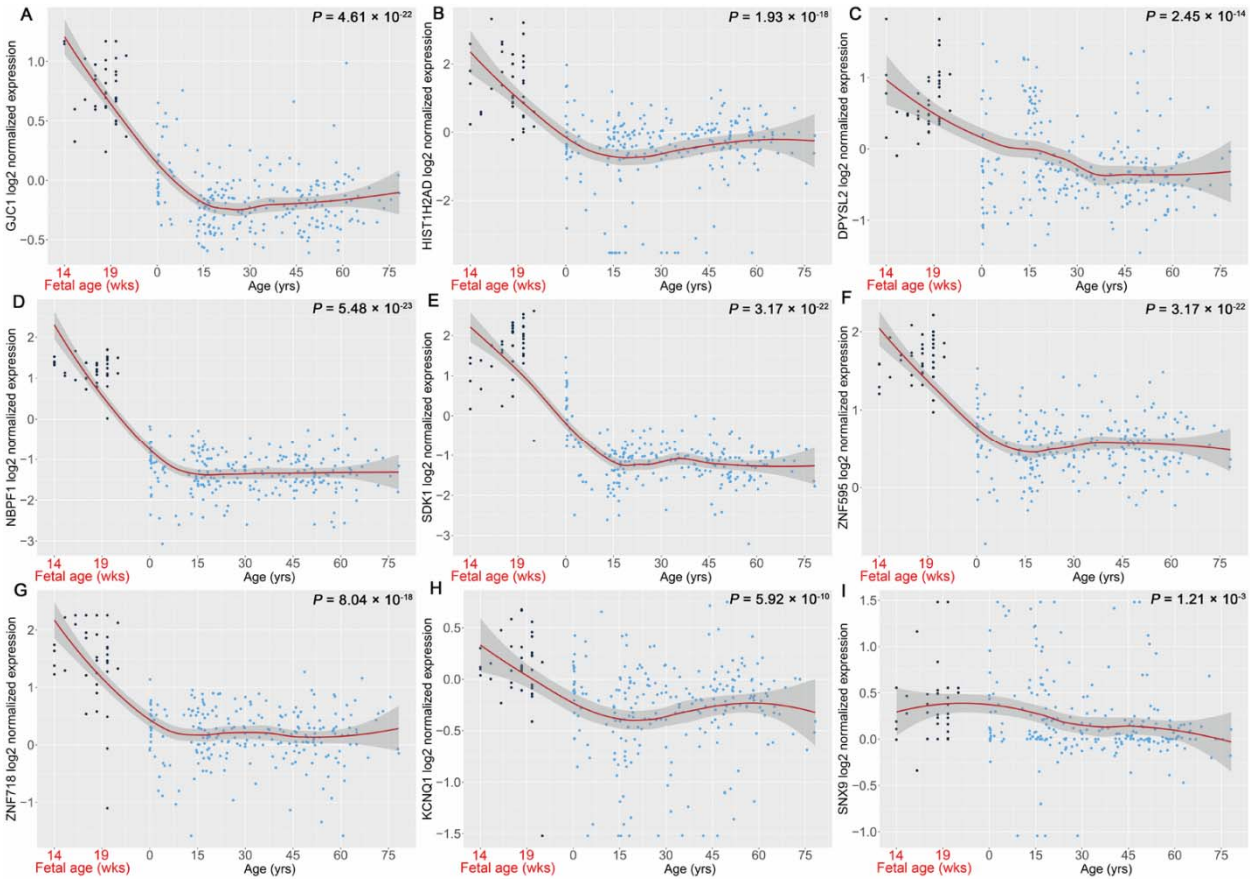
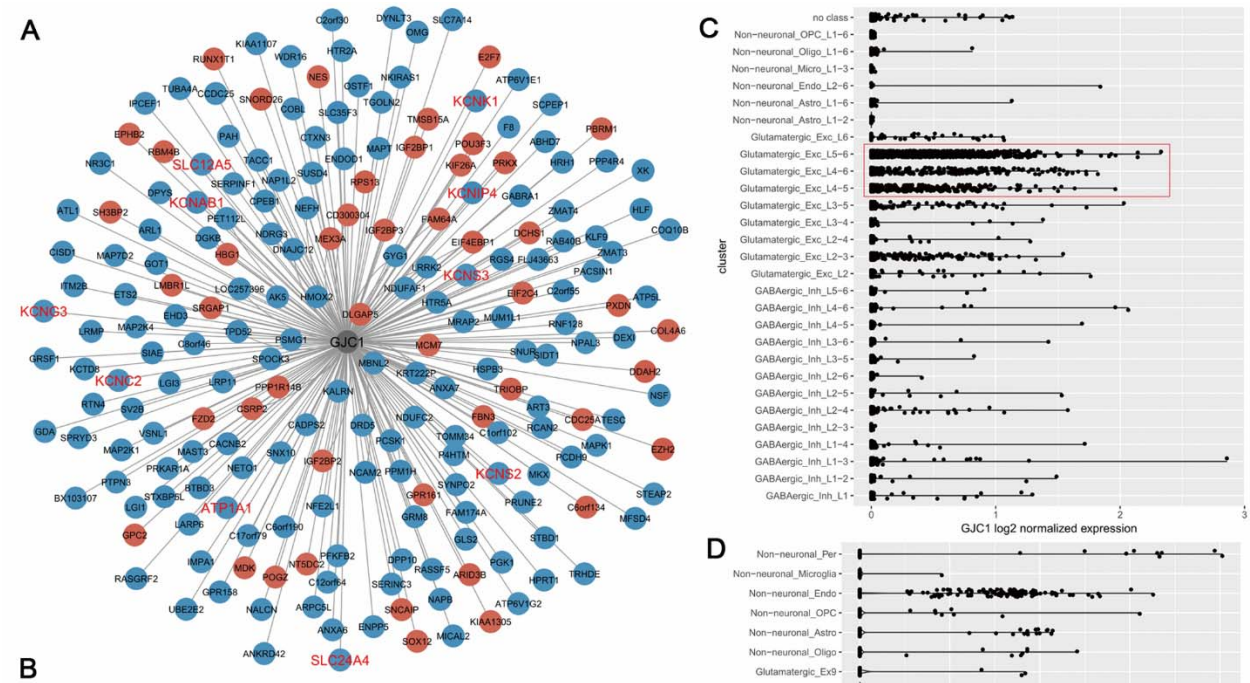

A
P = 4.61 × 10-22
GJC1 log2 normalized expression
B
P = 1.93 × 10-18
HIST1H2AD log2 normalized expression
C
P = 2.45 × 10-14
DPYSL2 log2 normalized expression
D
P = 5.48 × 10-23
NBPF1 log2 normalized expression
E
P = 3.17 × 10-22
SDK1 log2 normalized expression
F
P = 3.17 × 10-22
ZNF595 log2 normalized expression
G
P = 8.04 × 10-18
ZNF718 log2 normalized expression
H
P = 5.92 × 10-10
KCNQ1 log2 normalized expression
I
P = 1.21 × 10-3
SNX9 log2 normalized expression
14 19 0 15 30 45 60 75
Fetal age (wks)
Age (yrs)

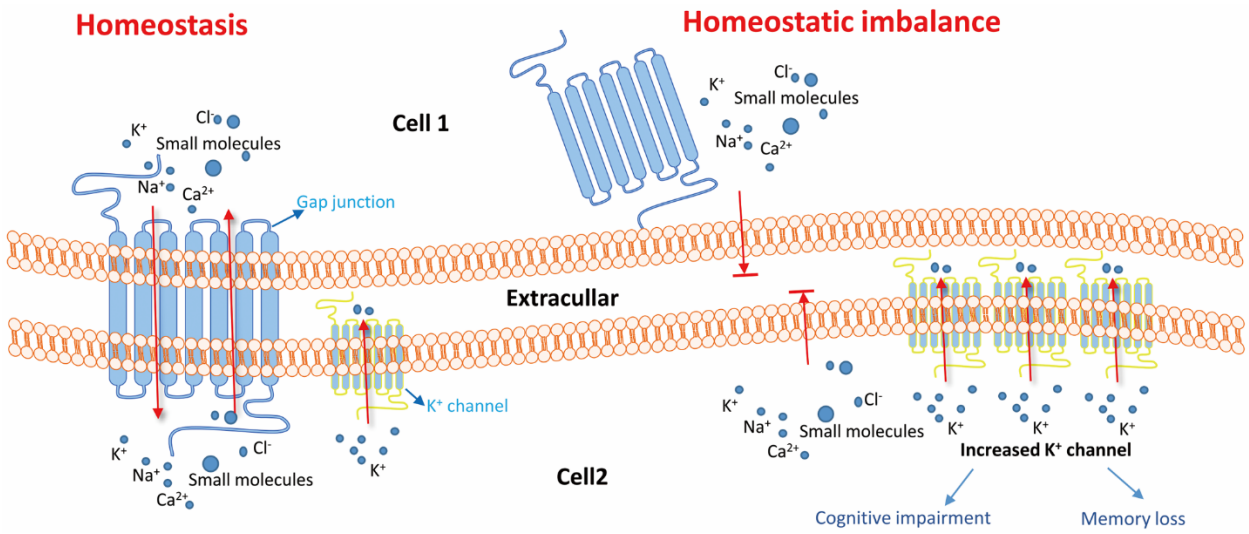
Figure 4



B

Enrichment category	Description	FDR
GO biological process	Potassium ion transport	2.49×10^{-4}
GO cell component	Axon	1.67×10^{-2}
	Dendrite	1.75×10^{-2}
	Post-synapse	1.75×10^{-2}
	Cell body	3.86×10^{-2}
GO molecular function	Translation regulator activity	1.47×10^{-2}
	Monovalent inorganic cation transmembrane transporter activity	2.95×10^{-2}
KEGG pathway	Dorso-ventral axis formation	4.05×10^{-2}
Disease	Schizophrenia	9.69×10^{-3}
	Adrenal Cortex Diseases	9.69×10^{-3}
	Mental Disorders	2.30×10^{-2}
Drug	Zuclopenthixol	3.46×10^{-2}

Figure 5


Homeostasis
Homeostatic imbalance
Cell 1
Cl-
K+
Small molecules
Na+
Ca2+
Gap junction
Extracullar
K+ channel
K+
Na+
Small molecules
Ca2+
Cl-
K+
Cell2
Cl-
Small molecules
K+
Na+
Ca2+
K+
Cl-
Na+
Small molecules
Ca2+
K+
K+
K+
Increased K+ channel
Cognitive impairment
Memory loss