medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

[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).

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 with matched unaffected siblings and parents. Two nonsense *de novo* mutations (DNMs) in *GJC1* and *HIST1H2AD* were identified in SCZ patients. Ten genes (*DPYSL2, NBPF1, SDK1, ZNF595, ZNF718, GCNT2, SNX9, AACS, KCNQ1* and *MSI2*) were found to carry more DNMs in SCZ patients than their unaffected siblings by burden test. Expression analyses indicated that these DNM implicated genes showed significantly higher expression in prefrontal cortex in prenatal stage. The DNM in the *GJC1* gene is highly likely a loss function mutation (pLI = 0.94), leading to the dysregulation of ion channel in the glutamatergic excitatory neurons. Analysis of rare variants in independent exome sequencing dataset indicates that *GJC1* has significantly more rare variants in SCZ patients than in unaffected controls. Data from genome-wide association studies suggested that common variants in the *GJC1* gene may be associated with SCZ and SCZ-related traits. Genes co-expressed with *GJC1* are involved in SCZ, SCZ-associated pathways and drug targets. These evidence suggest that *GJC1* may be a risk gene for SCZ and its function may be involved in prenatal and early neurodevelopment, a vulnerable period for developmental disorders such as SCZ.

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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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.

However, 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

within exon-intron junction regions may influence splice sites and thus lead to inappropriate expression of particular isoforms [11]. Takata *et al.* reported that cis-acting splicing quantitative trait loci from prefrontal cortices of human were in linkage disequilibrium with SCZ genome-wide association study (GWAS) loci and DNMs near splicing sites were enriched in SCZ patients as compared to controls [12, 13]. 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 [14, 15]. The rationale for combining the two cohorts was to increase the power to detect DNMs because each of the two cohorts has small sample size but they are ethnically homogeneous. 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) [16], 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 S1. 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 [17]. 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, Tokyo, Japan). DNA libraries were prepared using protocols recommended by Illumina (Illumina, San Diego,

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 S1).

The GATK best practices of variant calling were applied to process all raw reads from both Taiwan and Shanghai cohorts [18]. Raw sequencing reads in FASTQ format were aligned to the GRCh37 build of the human reference genome with BWA-mem [19]. 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 S2). Variant calling was performed across all samples with GATK (v4.1.1.0) HaplotypeCaller. For detailed variant calling pipeline, please refer to Supplementary Figure S2. To further systematically remove potentially false positively called variants, we applied the following quality filters for each variant: (1) the quality score normalized by allele depth is less than 2; (2) the root mean square of the mapping quality is less than 40; (3) the strand bias is more than 60; (4) three consecutive variants were clustered within 10 bases; and (5) the FILTER tag is not PASS.

Kinship analysis, DNM calling and annotation

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

We used PLINK to perform kinship analysis [20]. 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 [21-23]. 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 tagged with PASS in the FILTER annotation: (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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 S2 for detailed pipeline of DNM calling and Supplementary Figure S3 for the distributions of DNM quality scores. Remaining DNMs were annotated using ANNOVAR [24].

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 [25]. 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 [26]. To identify those genes implicated by protein-coding DNMs or of which the DNMs occurred higher

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

than the background mutation rates, the R package denovolyzeR was used to analyze protein-coding DNMs based on a mutation model developed previously [26]. 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.

To validate whether the genes implicated by coding DNMs were also implicated by rare variants, the population-based SCZ Swedish case-control cohort with whole exome sequences and variants called were used to perform the analysis [27]. Variants with MAF greater than 0.01 were excluded and only rare variants were retained subsequently. The association between the set of rare variants from Swedish case-control cohort located in the exons of a given gene implicated by coding DNM and phenotype was tested using SKAT [28].

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) [29, 30]. For a given gene, DNMs occurring in its gene body, which includes both exons and introns, 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 have shown that prenatal maternal infection, malnutrition and stress are risk factors for SCZ [31-33]. 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 [34]. 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 [35, 36]. The expression data of BrainCloud consisted of data derived from the prefrontal cortices of 269 individuals [35], whereas the expression data of BrainSpan consisted of 42 brain specimens across 13 developmental states in 8-16 brain structures [36]. 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 [37]. 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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) [38, 39].

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 [40], ADHD [41], bipolar disorder (BD) [42], major depressive disorder (MDD) [43], intelligence [44], educational attainment (EA) [45], cognitive performance (CP) [45], and smoking and drinking [46]. 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/) [47]. 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 [48]. Raw read count data was normalized by log-transformation using R package Seurat [49]. 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 potentially 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 S1). Three children from Shanghai cohort were excluded as they were unrelated to their respective parents (Individual IDs: CHG000225, CHG000236 and CHG000246) (Supplementary Table S1). A total of 21 families were retained for analyses. For detailed information of the families recruited and the individuals

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

in each family in the Taiwan and Shanghai cohorts and the families or individuals excluded due to unrelatedness, please refer to Supplementary Table S1.

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) [23]. The observed *de novo* indel mutation rate of 1.022×10^{-9} was consistent with the neutral expectation of 1.420×10^{-9} ($P = 0.500$, two-sided exact binomial test) [23]. All DNMs were checked visually by Integrative Genomics Viewer (IGV). Overall, the DNMs occurred in non-coding regions predominantly in intergenic regions (Supplementary Figure S4). No obvious difference was observed between unaffected siblings and SCZ patients with regards to the distributions of DNM locations relative to genes (Supplementary Figure S4). No obvious differences were observed between unaffected siblings and SCZ patients with regards to the percentages of DNMs occurring in exons (unaffected siblings: 29.85% versus SCZ patients: 30.29%) and introns (unaffected siblings: 1.53% versus SCZ patients: 2.03%). The nonsynonymous-to-synonymous ratio in SCZ patients was not found to be different from the ratio in unaffected siblings [50], 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 S4). The two loss-of-function DNMs were visually verified by IGV (Supplementary Figure S5 and S6) and then were confirmed by Sanger sequencing (Supplementary Figure S7 and S8). 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.* [51]. Nonsense DNMs were not

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

prone to occur in DNA regions of *GJC1* and *HIST1H2AD* (Supplementary Figure S9). No loss-of-function DNMs were identified in unaffected siblings on the whole genome scale (Supplementary Table S3).

Except for the two nonsense loss-of-function DNMs, we also investigated other DNMs in coding regions to find damaging DNMs because DNMs occurring in coding regions may change protein structures physically (Figure 1 and Supplementary Table S3). A detailed list of these DNMs are shown in Supplementary Table S4. 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 S5). 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*) [52]. 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 S4; mutations with CADD phred-like scores greater than or equal to 30 is the most severe 0.1% of mutations) [53]. Moreover, *GJC1* falls into the haploinsufficient category with a high probability of loss-of-function intolerance (pLI) of 0.94, therefore is extremely intolerant of loss-of-function variation [54].

To investigate whether *GJC1* and *HIST1H2AD* were also implicated by rare variants as DNMs in SCZ, we analyzed SCZ Swedish case-control cohort comprised of 4969 SCZ patients and 6245 controls. Rare variants in the exons of *GJC1* were significantly associated to SCZ (Table 2).

Identification of genes burdened with DNMs in SCZ patients

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Previous studies have shown that 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 S6), suggesting the dysregulation of these genes 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 S6. Interestingly, no DNM occurred in promoters of the ten genes.

DNMs occurred in genes involving in early brain development

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

In the following analyses, coding DNM-implicated genes (*GJC1* and *HIST1H2AD*, Supplementary Table S5) and DNM-burdened genes (Table 3) were as assumed to be detrimental DNM genes to explore their potential functions that might be related to SCZ.

Early neurodevelopmental events have been implicated in the pathogenesis of SCZ [34]. Genes implicated in SCZ function in processes important to fetal brain development [50, 55]. To determine whether the detrimental DNM genes, including *GJC1* and *HIST1H2AD* implicated by nonsense mutations (Supplementary Table S5) and genes carrying more DNMs in SCZ patients (Table 3), were 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 large proportion of 9/12 potentially detrimental DNM genes with higher expression in prefrontal cortices was more than expected by chance (Supplementary Figure S10, $P = 0.05$, 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 S11). 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

expression in postnatal stage (Supplementary Figure S12). 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 13), 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 [56].

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. We found that 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* 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 S8). DNMs

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

in other brain-associated connexin genes were also found in psychiatric patients (Supplementary Table S9). *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 S11, S12 and S13). Moreover, common variants present in *GJC1* confer risks for intelligence, educational attainment, cognitive performance and alcohol abuse, which were associated with SCZ (Supplementary Table S7) [57-60].

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 S10). 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 S11). 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 S12 and S13 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 [61].

Discussion

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 predominantly in neurons [62]. Our analyses showed that *GJC1* was expressed highly in prenatal stage as compared to that in postnatal stage and was expressed predominantly in glutamatergic excitatory neurons in the human brain (Figure 4C and 4D). For the nonsense mutation identified in our study, its occurrence might trigger the activation of surveillance pathway nonsense-mediated mRNA decay (NMD), which reduces aberrant proteins to be formed [63]. The reduction of *GJC1* may perturb the co-expressed genes and thus influence the potassium ion transport and axon/dendrite formations, especially in early neurodevelopment (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 [64]. Kumai *et al*. reported that Cx45-deficient embryos displayed an endocardial cushion defect in early cardiogenesis and died of heart failure at around embryonic day 9 [65]. Nishii *et al*. showed that mice lacking Cx45 conditionally in cardiac myocytes displayed embryonic lethality and the requirement of Cx45 for developing cardiac myocytes [66]. These studies indicate 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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 [67]. Ghelardini *et al*. reported that administrations of potassium channel openers bring out amnesic effect which can be reversed by potassium channel blockers [68]. This evidence suggested that the cognitive or memory deficits in this case were caused by the increased expression of potassium channel genes due 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 [69]. 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 [69]. Our findings support these hypotheses. Overall, the loss-of-function of Cx45 may be a driver of pathogenesis for some SCZ cases.

HIST1H2AD, the other nonsense DNM gene close to HLA locus, encodes the member D in the 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 [70, 71]. 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 [72], which implies the fundamental role of histone in the pathology of SCZ.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

458 Gene level DNM burden test identified a list of genes carrying more DNMs in SCZ patients than
459 their unaffected siblings. In the literature, there were experimental or clinical evidences
460 demonstrating their potential association with SCZ. Luan et al. reported SNPs in *MSI2* are
461 strongly associated with SCZ in the Chinese population [73]. In the GWAS of PGC, the SNPs in
462 *MSI2* were also associated with SCZ (Supplementary Table 7). Another DNM-burdened genes,
463 *DPYSL2*, is a member of the collapsin response mediator protein (CRMP) family. CRMP forms
464 homo- and hetero-tetramers and facilitates neuron guidance, growth and polarity. It also plays a
465 role in synaptic signaling through interactions with calcium channels. Lee *et al*. revealed that
466 *DPYSL2* was downregulated in the PFC and hippocampus of prenatally stressed (PNS) adult rats
467 that underperformed in behavioral tests [74]. Bruce *et al*. conducted potassium channel-targeted
468 SNP association analyses with SCZ and SCZ-associated phenotypes. rs8234 in *KCNQ1*, a DNM-
469 burdened gene in our study, was associated to processing speed [75]. Geschwind *et al*. found that
470 patients with voltage-gated potassium channel complex antibody (VGKCC-Abs) had particular
471 impairment in memory and executive functions when they evaluated cognitive function and
472 imaging data in patients with VGKCC-Abs associated encephalopathy [76]. Interestingly, the
473 impaired brain functions coincide with the two most implicated brain regions in SCZ,
474 dorsolateral prefrontal cortex and the hippocampus [77]. This evidence demonstrated the efficacy
475 of our DNM burden test by incorporating enhancer and promoter regions into consideration. It
476 also indicated DNMs contributing to the genesis of disease are not limited in coding regions,
477 underling the value of WGS.

478 Except for *HIST1H2AD*, common variants in these identified DNM genes were not strongly
479 associated to psychiatric diseases. However, a few of them, *GJC1*, *HIST1H2AD* and *SDK1*, were
480 strongly associated to SCZ-associated traits, including intelligence, educational attainment,

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

481 smoking and drinking. It suggested the significance of these DNM genes in neurodevelopment
482 and that the lethal mutations occurred in these genes were eliminated by marked reduction of
483 fecundity in psychiatric patients.

484 The results should be interpreted with caution because of limited samples in the pilot study. A
485 few parents with other mental illness from the Taiwan cohort may complicate the interpretation
486 of the DNMs if the mental illness shares certain genetic liability with SCZ. Nevertheless, the
487 association of the nonsense DNM gene *GJC1* with SCZ is deemed reliable: First, the association
488 of *GJC1* with SCZ is supported by the analyses of rare variants of an independent exome
489 sequencing dataset where SCZ patients have more rare variants in the gene than controls[27].
490 Second, with its intolerance of loss-of-function index of pLI = 0.94, the loss-of-function
491 mutation is very likely to have detrimental consequences in the regulation of ion channels which
492 have been implicated in the pathogenesis of SCZ[78, 79]. To validate our findings, future studies
493 with more and independent Chinese family samples are necessary.

494 In summary, we identified a list of DNM-implicated genes which are involved in prenatal
495 neurodevelopment. Common variants in these DNM-implicated genes had been previously
496 reported to be associated with SCZ and related traits in previous GWASs, which is consistent
497 with our analyses. DNMs implicating the enhancers may contribute to pathogenesis of SCZ by
498 dysregulating the expression of genes. *GJC1* is one of these DNM-implicated genes that is
499 primarily expressed in glutamatergic neurons and may be involved in the modulation of ion
500 channel functions, which has also been implicated in SCZ in previous studies [78, 79]. Overall, our
501 study provided new evidence that DNMs have a significant role in SCZ. Further study of these
502 DNM-implicated genes with functional analyses could lead to better understanding of the
503 pathology of SCZ.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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

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

567

568

569

570

571

572

573

574

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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. Takata A, Matsumoto N, Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun* 2017; **8:** 14519.

13. Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. *Neuron* 2016; **89**(5)**:** 940-947.

14. 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.

15. 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.

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

17. 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.

18. 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.

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

20. 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.

21. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

22. 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.

23. 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.

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

25. 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.

26. 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.

27. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* 2016; **19**(4)**:** 571-577.

28. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet* 2013; **92**(6)**:** 841-853.

29. 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).

30. 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.

31. 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.

32. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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

697
698 34. Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of
699 schizophrenia. *Nat Rev Neurosci* 2017; **18**(12)**:** 727-740.

700
701 35. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT *et al.* Temporal dynamics
702 and genetic control of transcription in the human prefrontal cortex. *Nature* 2011;
703 **478**(7370)**:** 519-523.

704
705 36. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A *et al.* Transcriptional
706 landscape of the prenatal human brain. *Nature* 2014; **508**(7495)**:** 199-206.

707
708 37. Grote S, Prufer K, Kelso J, Dannemann M. ABAEnrichment: an R package to test for
709 gene set expression enrichment in the adult and developing human brain. *Bioinformatics*
710 2016; **32**(20)**:** 3201-3203.

711
712 38. Fertuzinhos S, Li M, Kawasawa YI, Ivic V, Franjic D, Singh D *et al.* Laminar and
713 temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex.
714 *Cell Rep* 2014; **6**(5)**:** 938-950.

715
716 39. Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L *et al.* A comprehensive
717 transcriptional map of primate brain development. *Nature* 2016; **535**(7612)**:** 367-375.

718
719 40. Autism Spectrum Disorders Working Group of The Psychiatric Genomics C. Meta-
720 analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a
721 novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* 2017; **8:**
722 21.

723
724 41. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of
725 the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat*
726 *Genet* 2019; **51**(1)**:** 63-75.

727
728 42. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V *et al.* Genome-
729 wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019;
730 **51**(5)**:** 793-803.

731
732 43. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.*
733 Genome-wide association analyses identify 44 risk variants and refine the genetic
734 architecture of major depression. *Nat Genet* 2018; **50**(5)**:** 668-681.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

54. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616)**:** 285-291.

55. 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.

56. 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.

57. 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.

58. 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.

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

60. 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.

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

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

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

64. 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.

65. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

66. 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.

67. 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.

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

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

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

71. 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.

72. 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.

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. 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.

75. 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.

76. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

77. 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.

78. 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.

79. 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.

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

871

872 **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

873 [a]Chr: chromosome.

Table 2. Rare variant association tests for the loss-of-function DNMs identified in schizophrenia patients

Gene	Rare variant frequency in SCZ	Rare variant frequency in control	SKAT *P* value
GJC1	0.0115	0.0091	5.60×10^{-3}
HIST1H2AD	0.0029	0.0043	6.28×10^{-1}

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

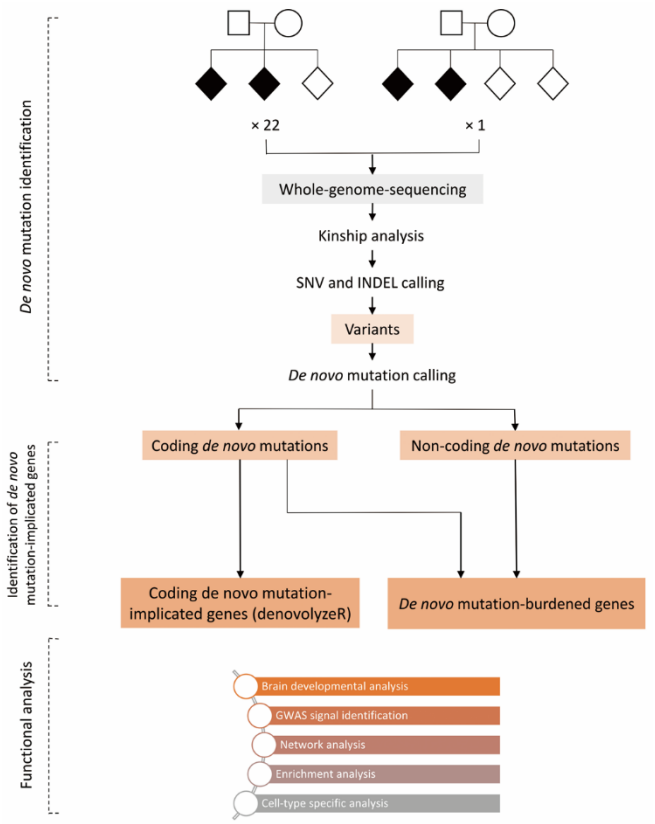
Table 3. Genes significantly enriched with higher DNM burden in schizophrenia patients than unaffected siblings

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}

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

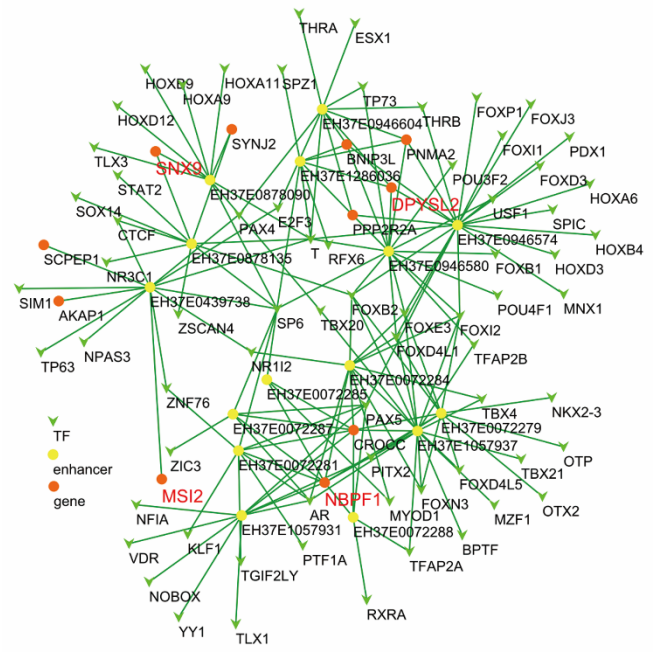
medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 1



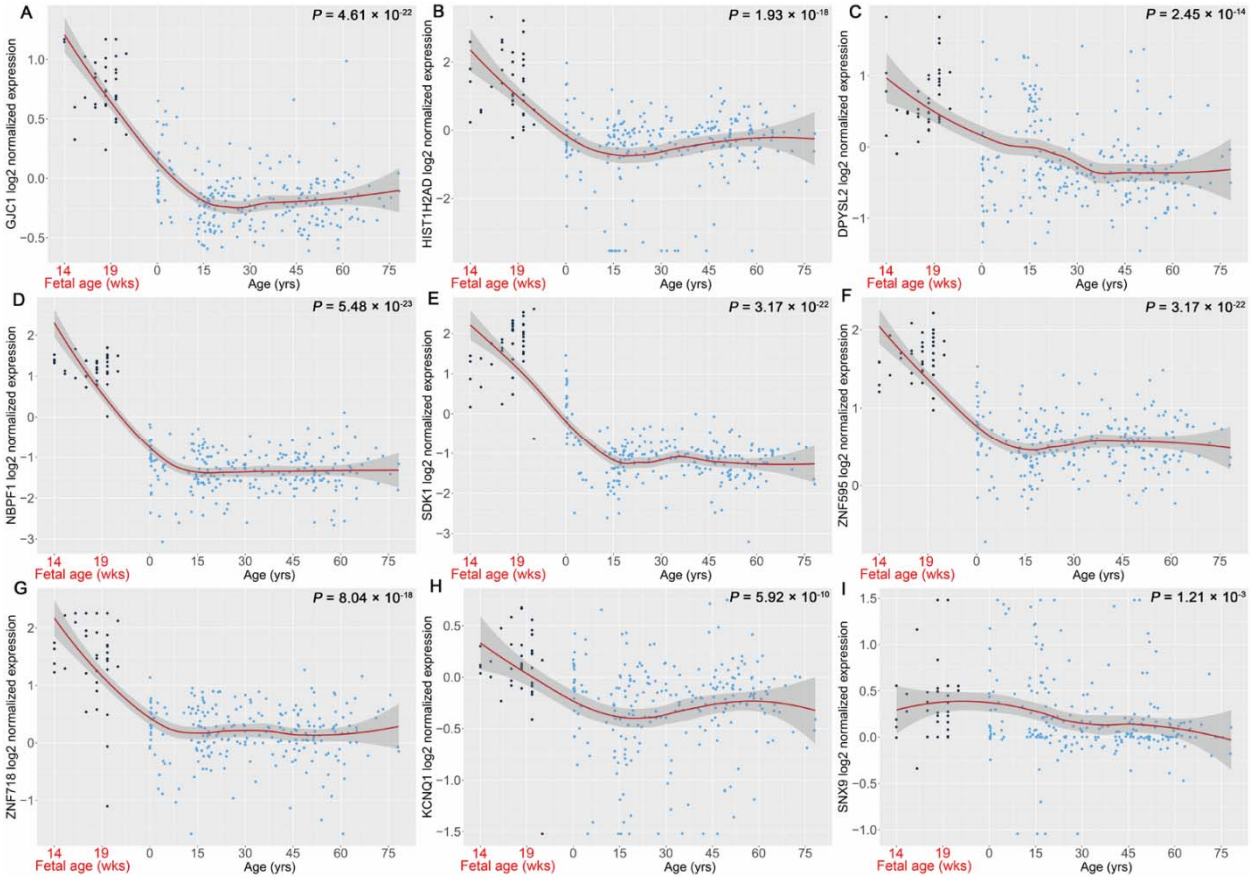
medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 2



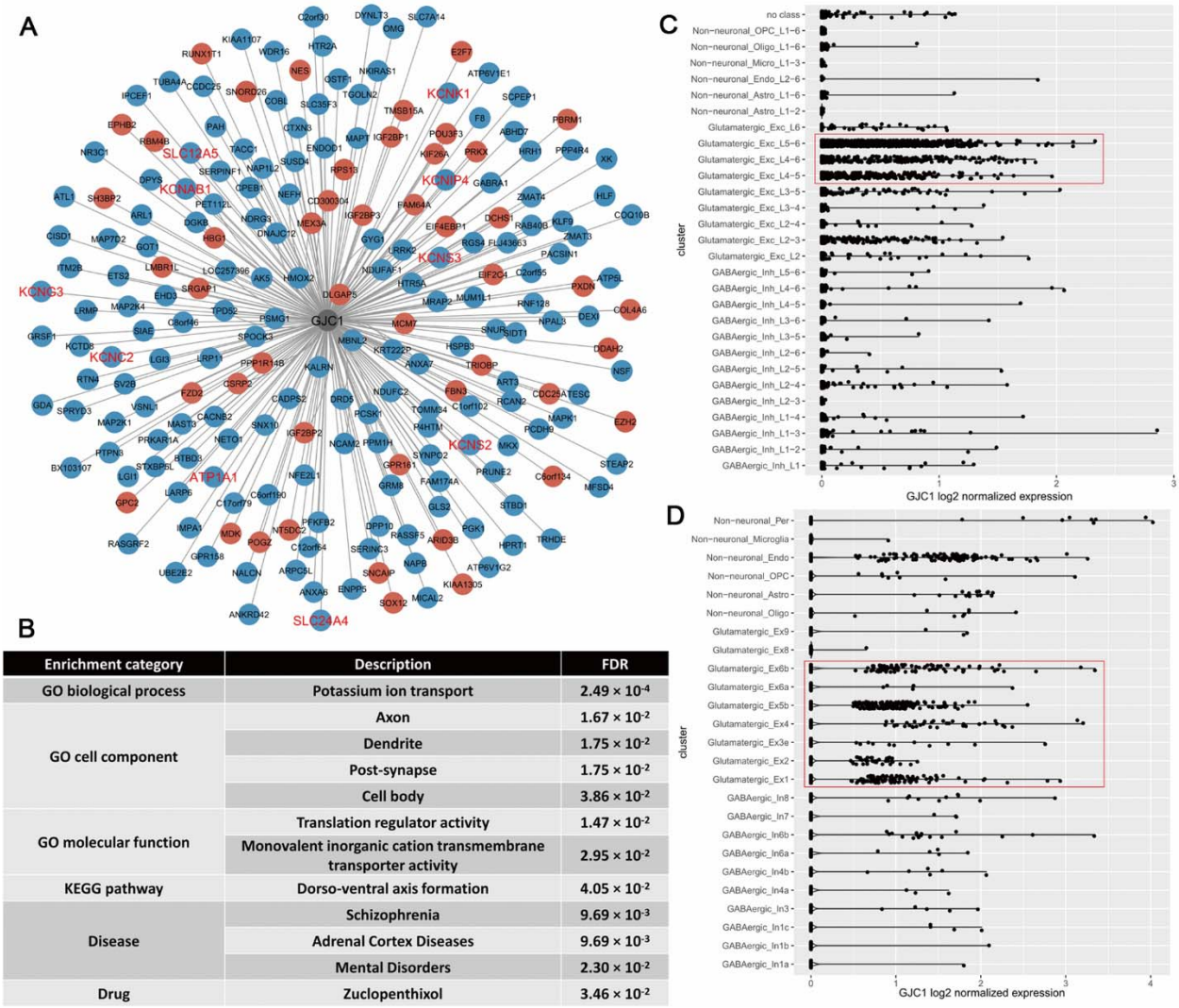
medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 3



medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 4



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}

medRxiv preprint doi: https://doi.org/10.1101/19011007; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 5