Nexus Between Genome-Wide Copy Number Variations and Autism Spectrum Disorder in Northeast Han Chinese Population

Shuang Qiu
Jilin University School of Public Health
Xianling Cong
China-Japan Union Hospital of Jilin University
Yan Li
Jilin University School of Public Health
Jikang Shi
Jilin University School of Public Health
Yingjia Qiu
China-Japan Union Hospital of Jilin University
Yanbo Guo
Jilin University School of Public Health
Yulu Gu
Jilin University School of Public Health
Yong Li
Jilin University School of Public Health
Xiaojuan Zhu
Northeast Normal University
Yunkai Liu
Jilin University First Hospital
Yichun Qiao
Jilin University School of Public Health
Yi Cheng
Jilin University First Hospital
Yawen Liu (ywliu@jlu.edu.cn)
Jilin University School of Public Health https://orcid.org/0000-0003-2859-4771

Research

Keywords: Autism spectrum disorder, Copy number variations, MicroRNAs, Duplication, Deletion, Array-based comparative genomic hybridization

Posted Date: January 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-154430/v1

Abstract

Background: Autism spectrum disorder (ASD) is a common neurodevelopmental condition, with an increasing prevalence worldwide. Copy number variation (CNV), as one of genetic factors, is involved in ASD etiology. However, there exist substantial differences in terms of location and frequency of some CNVs in the general Asian population. Whole-genome studies of CNVs in Northeast Han Chinese samples are still lacking, necessitating our ongoing work to investigate the characteristics of CNVs in a Northeast Han Chinese population with clinically diagnosed ASD.

Methods: We performed a genome-wide CNVs screening in Northeast Han Chinese individuals with ASD using array-based comparative genomic hybridization.

Results: We found 22 kinds of CNVs (six deletions and 16 duplications) were potential pathogenic. These CNVs were distributed in chromosome 1p36.33, 1p36.31, 1q42.13, 2p23.1-p22.3, 5p15.33, 5p15.33-p15.2, 7p22.3, 7p22.3-p22.2, 7q22.1-q22.2, 10q23.2-q23.31, 10q26.2-q26.3, 11p15.5, 11q25, 12p12.1-p11.23, 14q11.2, 15q13.3, 16p13.3, 16q21, 22q13.31-q13.33, and Xq12-q13.1. Additionally, we found 20 potential pathogenic genes of ASD in our population, including eight protein coding genes (six duplications [*DRD4*, *HRAS*, *OPHN1*, *SHANK3*, *SLC6A3*, and *TSC2*] and two deletions [*CHRNA7* and *PTEN*]) and 12 microRNAs genes (ten duplications [*MIR202*, *MIR210*, *MIR3178*, *MIR339*, *MIR4516*, *MIR4717*, *MIR483*, *MIR675*, *MIR6821*, and *MIR940*] and two deletions [*MIR107* and *MIR558*]).

Limitations: The sample size in our study may confer limited statistical power to discover significant findings. De novo or inherited of the CNVs were not be classified because of the lack of data from parents.

Conclusions: We identified CNVs and genes implicated in ASD risks, conferring perception to further reveal ASD etiology.

Background

Autism spectrum disorder (ASD) is a common neurodevelopmental condition, with an increasing prevalence worldwide [1, 2]. Persons with ASD manifest the wide range of symptoms and severity in perceivability and socialization with others, such as limited and repetitive patterns of behavior. Both genetic and environmental factors are involved in the pathogenesis of ASD. Environmental factors, including viral infections, medications during pregnancy, and air pollutants, may contribute to ASD [3]. Compared with environmental factors, genetic factors appear to be a prerequisite for ASD development: genetic changes (mutations) may increase ASD risks; and genes, such as *CHD8* [4], *CNTNAP2* [5], *DCC* [6], neurexin genes [7], *SHANK1* [8], *SHANK2* [9], *SHANK3* [10], and *WNT2* [11] may affect brain development or brain-cells communication. The heritability of ASD has been estimated to be 50%, reflecting that genetic factors contribute to main components in ASD etiology [12].

ASD begins in early childhood. Children with ASD usually show symptoms of autism within the first year, and regress during a period between one and two years of age. Although there is no specific medication for ASD patients [13], early treatment can confer a big difference in the lives of children with ASD. Gene-based test provides an impressive method to identify the potential infants with ASD [8].

Accumulating whole-genome, association, and linkage studies have strongly documented the roles of genes in ASD [14–16]. Copy-number variants (CNVs) are defined as deletions and duplications of DNA segments in the genome greater than one kilobase (Kb) [17, 18]. De novo CNV events have been found to be implicated in the etiology of depression, schizophrenia, bipolar disorder, attention deficit hyperactivity disorder, and ASD [19–22]. Array-based comparative genomic hybridization (aCGH) technology has proven to be a rapid method to detect the association between CNVs and ASD risks [23, 24]. Large simplex ASD cohort studies show that the rate of rare de novo CNVs is significantly higher in affected sibling (5.8–7.9 %) than that in unaffected siblings (1.7–1.9 %) [25, 26]. CNVs at 1q21.1, 2p16.3, 3q29, 7q11.23, 15q11.2–13.1, 16p11.2, 17p11.2, 17q12, and 22q11.2 are associated with ASD risks [24, 27]. Moreover, CNVs in *NRXN1*, *SETD5*, *HDAC9*, and *PARK2* are found to be associated with ASD risks [28–30]. However, there exist substantial differences in terms of location and frequency of some CNVs in the general Asian population [31]. In this paper, we investigated in CNVs in Northeast Han Chinese individuals with ASD.

Research Methods

Study subjects

We enrolled 16 individuals with ASD aged 2 to 7 years from the Chunguang Rehabilitation hospital in Jilin Province, after cases with fragile X syndrome, Rett syndrome, chromosomal abnormalities, and any neurological or psychiatric disorders were excluded. The individuals with ASD were diagnosed by Pediatric Neurology and Neurorehabilitation doctors using the Diagnostic and Statistical Manual of Mental Disorders (5th edition) [32]. All the individuals with ASD were northeast Han Chinese. This study was approved by the ethics committee of Jilin University. The parents or guardians of each individual with ASD signed the written informed consent forms.

DNA extraction and Detection of CNVs

Genomic DNA was extracted from peripheral blood samples using DNA extraction kits, according to the manufacturer's instructions (DP319 TIANamp Blood DNA Kit, TIANGEN BiotechCo. Ltd, Beijing, China) [33]. We used Nano Drop (Cat#ND-1000, ThermoFisher, Waltham, MA, US) and 1% agarose gel electrophoresis to check the quantity and quality of the isolated DNA. We used aCGH for genome-wide CNVs screening (Agilent SurePrint G3 Human CGH 60K). Male and female DNA samples were hybridized with male and female reference DNA samples (G1471, G1521, Promega), respectively.

aCGH data Analysis

We converted the raw data using FEATURE EXTRACTION software 10.7 and analyzed CNVs using Agilent CytoGenomics software 4.0.3.12 (Agilent technologies, Santa Clara, CA, US). The human genome assembly NCBI36/hg18 was used as a reference. The analysis settings for CNVs calling were Aberration Detection Method 2 algorithm, centralization threshold 6, bin size 10, and minimum number of adjacent probes 3. Thresholds were set via log2-ratio (log_2^R) (for detecting duplications, $log_2^R \geq 0.25$; for detecting deletions, $log_2^R \leq -0.25$).

Identification of potential pathogenic CNVs of ASD

We calculated the frequency of each overlapping or non-overlapping CNV in DNA samples from our subjects. CNVs with same overlapping sequence were defined as one kind of CNV, and a non-overlapping CNV was also sorted as one kind of CNV. The circular plot of CNVs distribution in chromosome was visualized using circlize package in R3.6.2 software [34]. We converted our bed file from exon coordinates for human build NCBI36 (hg18) into GRCh37 (hg19) using UCSC LiftOver tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). The classification of CNVs was based on Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home), Database of Genomic Structural Variation (dbVar, https://www.ncbi.nlm.nih.gov/dbvar), Clinical Genome Resources (ClinGen, https://clinicalgenome.org/), and Online Mendelian Inheritance in Man (OMIM, https://www.ncbi.nlm.nih.gov/omim). CNVs were classified as benign, likely benign, a variant of unknown significance (VOUS), likely pathogenic, and pathogenic using AnnotSV program (https://lbgi.fr/AnnotSV/) according to American College of Medical Genetics guideline [35].

CNVs were considered of strong putative interest when they reached the following criteria: (1) they were classified as likely pathogenic or pathogenic; (2) they were of large size (> 100kb); (3) they had been found in the knowledgebases for the genetic evidence of ASD (Simons Foundation Autism Research Initiative [SFARI, https://www.sfari.org/resource/sfari-gene/], or AutismKB [http://db.cbi.pku.edu.cn/autismkb_v2/index.php]); (4) they had been found in the Database of genomic variation and phenotype in Humans using Ensembl Resources (DECIPHER, https://decipher.sanger.ac.uk/about#overview); and (5) they contained previously reported ASD-relative genes. All potential pathogenic CNVs showing ≥ 90% overlap with at least one common variant of the same type in the DGV database were considered as common CNVs, and the others were rare CNVs [36, 37].

Identification of potential pathogenic genes of ASD

We selected potential pathogenic genes within potential pathogenic CNVs on the basis of the following criteria: (1) genes enriched in ASD-related pathways; and (2) within 363 genes classified as high-confidence or strong-candidate in SFARI, and 228 genes classified as high-confidence in AutismKB.

Identification of potential pathogenic microRNAs of ASD

MicroRNAs (miRNAs) are involved in the pathogenesis of ASD [30, 38]. Because genes implicated in CNVs that we found encode miRNAs, we further selected potential-pathogenic-CNVs-related miRNAs by retrieving PubMed according to experimental evidence documenting nervous system dysfunction.

Bioinformatic analysis

The Gene Ontology (GO) and KEGG pathway analyses of the genes from potential pathogenic CNVs were performed using clusterProfiler package in R3.6.2 software [39]. *P*-value < 0.05 was considered statistically significant. miRWalk 2.0 database, which contained 12 miRNA-target-prediction database, was used to predict target genes of CNVs-miRNAs [40]. We selected the target genes according to the criteria—target genes existed in at least seven of the 12 databases. Moreover, interactive relationship between CNVs-miRNAs and target genes was presented using Cytoscape 3.8.0 (http://www.cytoscape.org/).

Results

Identification CNVs

To detect CNVs, aCGH was performed in all DNA samples from the 16 subjects with ASD (13 males and 3 females). We identified 364 CNVs (153 deletions and 211 duplications) with an average genomic size of 211.982 kb (114.091 kb for deletions and 258.705 kb for duplications). The mean number of CNVs per subject was 22.750 (9.563 for deletions and 13.188 for duplications). The mean number of deletions in male (10.462) was greater than that in females (5.667) (Table 1).

Table 1
The Characteristics of genome-wide CNVs among our subjects

Characters	Number (proportion %) of CNVs	Median CNV Size (kb)	Mean Number of CNVs per Subject
Total	364 (100.0)	211.982 (78.813, 705.031)	22.750
Male	299 (82.1)	213.241 (81.797, 864.947)	23.000
Female	65 (17.9)	132.360 (73.256, 305.926)	21.667
Duplication	211 (58.0)	258.705 (86.903, 708.853)	13.188
Male	163 (77.3)	305.926 (86.903, 1092.277)	12.538
Female	48 (22.7)	133.712 (87.668, 377.979)	16.000
Deletion	153 (42.0)	114.091 (73.256, 656.149)	9.563
Male	136 (88.9)	114.518 (73.256, 693.563)	10.462
Female	17 (11.1)	73.256 (60.021, 135.433)	5.667
A total of 13 males and 3 females.			

Identification potential pathogenic CNVs of ASD

A total of 20 CNVs from 364 CNVs failed to be converted to GRCh37 (hg19); thus, we obtained 72 benign, 65 likely benign, 9 VOUS, 167 likely pathogenic, and 31 pathogenic CNVs (Table 2). We found that more than half CNVs were likely pathogenic or pathogenic.

Table 2
The Classification of CNVs based on ACMG

Classification	Total (%)	Duplication (%)	Deletion (%)
Benign	72 (19.8)	35 (16.6)	37 (24.2)
Likely Benign	65 (17.9)	59 (28.0)	6 (3.9)
VOUS	9 (2.5)	9 (4.3)	0 (0.0)
Likely Pathogenic	167 (45.9)	68 (32.2)	99 (64.7)
Pathogenic	31 (8.5)	28 (13.3)	3 (2.0)
VOUS: variant of unknown significance; A total of 20 CNVs failed to be converted to GRCh37 (hg19), thus, the total proportion was not equal to 100%. ACMG: American College of Medical Genetics guideline			

After we calculated the frequency of each overlapping or non-overlapping CNV in DNA samples from our subjects, 344 CNVs were converted into 115 kinds of CNVs (45 deletions and 70 duplications). All the 115 kinds of CNVs were further classified (benign: 13 kinds; likely benign: 18 kinds; VOUS: two kinds; likely pathogenic: 60 kinds; and pathogenic: 13 kinds) (Supplementary Table 1). The distribution of the 115 kinds of CNVs in chromosome is visualized by circular plot (Fig. 1).

We investigated SFARI, AutismKB, and DECIPHER database to identify potential pathogenic CNVs from the 115 kinds of CNVs, revealing that 22 kinds of CNVs (6 deletions and 16 duplications) were potential pathogenic. The 22 kinds of CNVs were distributed in chromosome 1, 2, 5, 7, 10, 11, 12, 14, 15, 16, 22, and X. Among them, 19 kinds of CNVs were rare (Table 3).

Table 3
Summary of candidate CNVs of ASD

M/F No.	Coordinates, hg18	Cytoband	Size (Kb)	CNV Type	Classification	Rare/Common CNV	Number of genes	Gene Name
1/0	chr1:1179223–2271500	1p36.33	1092.277	Duplication	LP	Rare	42	DVL1, TMEM52
1/0	chr1:5998727–6334157	1p36.31	335.430	Duplication	LP	Rare	9	CHD5
4/0	chr1:225876894–226738916	1q42.13	862.022	Duplication	LP	Rare	26	PRSS38
1/0	chr2:31412158–32712484	2p23.1-p22.3	1300.327	Deletion	P	Rare	12	BIRC6, SPAST, SRD5A2
1/0	chr5:360041–873365	5p15.33	513.324	Duplication	LP	Common	11	AHRR, EXOC3, PDCD6
1/0	chr5:1115468–8452427	5p15.33-p15.2	7336.959	Duplication	P	Rare	49	ADCY2, SLC6A3, TERT
2/0	chr7:524935–1037461	7p22.3	512.526	Duplication	LP	Rare	13	ADAP1, PRKAR1B
1/0	chr7:1037461–2536804	7p22.3-p22.2	1499.343	Duplication	LP	Rare	26	INTS1
1/0	chr7:103622888–104803388	7q22.1-q22.2	1180.501	Deletion	LP	Rare	8	KMT2E, LHFPL3
13/0	chr10:89540133–91524263	10q23.2-q23.31	1984.131	Deletion	LP	Rare	31	PTEN
1/0	chr10:127658856–135254513	10q26.2-q26.3	7595.658	Duplication	P	Rare	58	DOCK1, EBF3, GLRX3
5/2	chr11:498019–2179368	11p15.5	1681.349	Duplication	P	Rare	84	BRSK2, CD151, CTSD, DEAF1, DRD4, HRAS, IGF2, PHRF1, TALDO1
3/0	chr11:132773688–134043707	11q25	1270.019	Duplication	LP	Rare	16	IGSF9B
6/0	chr12:25156062–27414420	12p12.1-p11.23	2258.359	Deletion	LP	Rare	17	KRAS, MED21
1/0	chr14:22086438–22354007	14q11.2	267.569	Deletion	LP	Rare	4	SLC7A7
1/0	chr15:29809025–30298155	15q13.3	489.131	Deletion	P	Rare	1	CHRNA7
0/2	chr16:2021433–2484806	16p13.3	463.373	Duplication	LP	Common	32	PGP, PKD1, RNPS1, SLC9A3R2, TRAF7, TSC2
0/3	chr16:2484806–2747528	16p13.3	262.722	Duplication	LP	Common	13	SRRM2

M/F No.	Coordinates, hg18	Cytoband	Size (Kb)	CNV Type	Classification	Rare/Common CNV	Number of genes	Gene Name
2/0	chr16:61464644–64965235	16q21	3500.591	Duplication	LP	Rare	4	CDH11
4/0	chr22:46395224–49412774	22q13.31-q13.33	3017.550	Duplication	P	Rare	44	CHKB, MAPK12, MAPK8IP2, PANX2, PPP6R2, SBF1, TRABD
2/0	chr22:49412774–49525130	22q13.31-q13.33	112.356	Duplication	P	Rare	3	SHANK3
1/0	chrX:67331017–68768438	Xq12-q13.1	1437.422	Duplication	P	Rare	8	OPHN1

M: Male; F: Female. M/F No. means the number of CNV among male/female. P: Pathogenic; LP: Likely Pathogenic. The genes were reported to be related with ASD.

Identification of potential pathogenic genes with CNVs of ASD

A total of 511 genes from the 22 potential pathogenic CNVs were functionally annotated by GO. The annotated genes were classified into three GO domains (biological processes [BP], cellular component [CC], and molecular function [MF]). For BP, some gene sets were enriched in synaptic-related functions, including modulation of chemical synaptic transmission (GO: 0050804), regulation of trans-synaptic signaling (GO: 0099177), positive regulation of excitatory postsynaptic potential (GO: 2000463), positive regulation of synaptic transmission (GO: 0050806), chemical postsynaptic transmission, (GO: 0099565), modulation of excitatory postsynaptic potential (GO: 0098815), and regulation of postsynaptic membrane potential (GO: 0060078), and in central nervous system related functions (positive regulation of neurological system process [GO: 0031646]). For CC, the top five CC terms included keratin filament (GO: 0045095), myelin sheath (GO: 0043209), Golgi lumen (GO: 0005796), glutamatergic synapse (GO: 0098978), and neuron to neuron synapse (GO: 0098984). For MF, the top five MF terms encompassed catecholamine binding (GO: 1901338), dopamine binding (GO: 0035240), magnesium ion binding (GO: 0000287), insulin receptor binding (GO: 0005158), and lipase activity (GO: 0016298). The top 20 GO functions are presented in Fig. 2 and Supplementary Table 2, 3, 4.

KEGG pathway enrichment analysis showed enriched key pathways, such as dopaminergic synapse (hsa04728), mTOR signaling pathway (hsa04150), GnRH signaling pathway (hsa04912), cholinergic synapse (hsa04725), and MAPK signaling pathway (hsa04010). The top 20 pathways are presented in Fig. 2 and Supplementary Table 5.

We constructed intersections among 511 genes that we found, 363 high-confidence or strong-candidate risk genes of ASD reported in SFARI database, and 228 high-confidence risk genes related to ASD reported in AutismKB database (Fig. 3). After investigating genes in the intersections, we found that cholinergic receptor nicotinic alpha 7 subunit gene (*CHRNA7*) was involved in the regulation of excitatory postsynaptic potential and cholinergic synapse; dopamine receptor D4 gene (*DRD4*) was involved in the regulation of synaptic transmission, dopamine binding, and glutamatergic synapse; HRas proto-oncogene (*HRAS*) played roles in the regulation of excitatory postsynaptic potential, glutamatergic synapse, and mTOR signal pathway; oligophrenin 1 gene (*OPHN1*) correlated with regulated synaptic signal, ionic glutamate receptor binding, and glutamatergic synapse; phosphatase and tensin homolog (*PTEN)* was implicated in the regulation of synaptic signal, neuron differentiation of central nervous system, ionic glutamate receptor binding, sphingolipid signaling, and mTOR signaling; SH3 and multiple ankyrin repeat domains 3 gene (*SHANK3*) was involved in the regulation of synaptic signal, ionic glutamate receptor binding, neuronal synapse, postsynaptic density, and asymmetric synapse; solute carrier family 6 member 3 gene (*SLC6A3*) played roles in dopamine binding, neurotransmitter: sodium cotransporter activity, and neurotransmitter transport activity; and TSC complex subunit 2 gene (*TSC2*) was involved in synapses, postsynaptic density, asymmetric synapses, and mTOR signaling pathways. Scores of all these genes (*CHRNA7*, *DRD4*, *HRAS*, *OPHN1*, *PTEN*, *SHANK3*, *SLC6A3*, and *TSC2*) in AustismKB and corresponding ranks in SFARI are listed in Table 4. *DRD4*, *HRAS*, *OPHN1*, *SHANK3*, *SLC6A3*, and *TSC2* were in the regions of CNVs duplication. *CHRNA7* and *PTEN* were in the regions of CNVs deletion.

Table 4
Summary of candidate genes of ASD

Gene Name	CNV Type	M/F No.	Category of gene in SFARI	Score of gene in AutismKB
CHRNA7	Deletion	1/0	2	#
DRD4	Duplication	5/2	—	30
HRAS	Duplication	5/2	1	20
OPHN1	Duplication	1/0	2	#
PTEN	Deletion	13/0	1	78
SHANK3	Duplication	2/0	1	62
SLC6A3	Duplication	1/0	2	30
TSC2	Duplication	0/2	1	46

M: Male; F: Female. M/F No. means the number of CNV among male/female. —: Not reported as high-confidence or strong candidate autism risk genes in SFARI. #: Not reported as high-confidence autism risk genes in AutismKB.

Identification and Analysis of potential pathogenic miRNAs with CNVs of ASD

We found 50 potential-pathogenic-CNVs-related miRNAs (45 encoded by duplication regions and 5 encoded by deletion regions). According to experimental evidence documenting nervous system dysfunction, we retrieved PubMed, identifying that 12 miRNAs genes were previously reported to be associated with brain or nervous system dysfunction (Table 5).

Table 5
miRNAs with function related to brain or nervous system in CNVs

miRNA ID	CNV Type	Functional relevance	Reference (PMID)
miR-202	Duplication	Depression, Glioma, Neuroblastoma	32425535; 28714009; 21654684; 24337320
miR-210	Duplication	Alzheimer's disease, Epilepsy, Glioblastoma, Glioma, Head and neck paragangliomas, Neuroblastoma, Neuroprotective effects	31092279; 23108914; 21655185; 22977270; 23902947; 25279461; 24729345; 24382515; 25481483; 24930954; 25756397; 29126304; 29362886; 31146085; 32194691; 31896490; 29226333; 30947960; 30746749; 27471387
miR-3178	Duplication	Neuropsychiatric diseases	30766477
miR-339	Duplication	Alzheimer's disease, Glioblastoma, Neuroendocrine neoplasias	32176627; 29983867; 30564636; 24352696; 30176243
miR-4516	Duplication	Glioblastoma	30559405
miR-4717	Duplication	Guillain-Barre Syndrome	27836180
miR-483	Duplication	Alzheimer's disease, Glioma, Neuroblastoma	31938135; 24577456; 22465663
miR-675	Duplication	Glioma	31468534; 28187439; 24466011
miR-6821	Duplication	Alzheimer's disease	27050411
miR-940	Duplication	Glioblastoma, Glioma	31497204; 30906627; 31934283; 29296221; 30431124
miR-107	Deletion	Alzheimer's disease, Amnestic mild cognitive impairment, Bipolar disorder, Brain disorders, Frontotemporal dementia, Glioblastomas, Glioma, Major depression, Neuroblastoma, Neurogenesis, Schizophrenia	31556571; 29258209; 28847283; 26084601; 30543171; 31250578; 27343180; 21625387; 20489155; 28578378; 25662174; 22811466; 20413881; 31778666; 31787850; 29885309; 29671226; 30056425; 30480816; 18234899; 29136645; 23811124; 27143098; 29073742; 25596705; 31605836; 31420923; 23220650; 22594617; 26223576; 23572380; 27501295; 27878295; 32124921; 23962497; 29286086; 21179570; 21111402
miR-558	Deletion	Neuroblastoma	25616966; 27276678

We intersected target genes predicted using miRWalk 2.0 database with the union between SFARI and AutismKB, presenting the interaction networks between CNVs-miRNAs and 219 target genes (Figs. 4 and 5).

We further investigated potential functions of the 219 target genes using GO analysis. For BP, some gene sets were enriched in synaptic-related functions, including synapse organization (GO: 0050808), modulation of chemical synaptic transmission (GO: 0050804), regulation of trans-synaptic signaling (GO: 0099177), synaptic transmission, glutamatergic (GO: 0035249), postsynaptic density organization (GO: 0097106), postsynaptic specialization organization (GO: 0099084) and regulation of glutamatergic synaptic transmission (GO: 0051966), and in central nervous system related functions, including learning or memory (GO: 0007611), cognition (GO: 0050890), and neurotransmitter transport (GO: 0006836). For CC, some gene sets were enriched in synaptic-related cellular components, including synapse membrane (GO: 0097060), postsynaptic specialization (GO: 0099572), and neuron to neuron synapse (GO: 0098984). For MF, some gene sets were enriched in ion-gated channel activity (GO: 0022839), gated channel activity (GO: 0022836), ion channel activity (GO: 0005216), ionotropic glutamate receptor activity (GO: 0004970), and transmitter gated channel activity (GO: 0022835). The top 20 GO functions are presented in Fig. 6 and Supplementary Table 6, 7, 8.

KEGG pathway enrichment analysis showed enriched key pathways, such as glutamatergic synapse (hsa04724), dopaminergic synapse (hsa04728), and Wnt signaling pathway (hsa04310). The top 20 pathways are presented in Fig. 6 and Supplementary Table 9.

Discussion

In the present study, we identified that 22 kinds of CNVs (six deletions and 16 duplications), eight protein-coding genes, and 12 miRNAs genes are associated with ASD risks in northeast Chinese Han from Jilin province, China.

CNVs have repeatedly been found to correlate with ASD risks [41, 42]. In our study, we filtered 22 potential pathogenic CNVs. Individuals with deletions and duplications of 15q13.3 have been found to manifest neuropsychiatric disease and cognitive deficits [43]. In line with the discoveries of Chen *et al.* [44] and Pinto *et al.* [28], we further documented that CNVs at 22q13.33 and 15q13.3 are associated with ASD risks. Autism-related phenotypes are common in patients with deletion or duplication at 22q13.3 [45–48]. Most of the defects are due to haploinsufficiency of *SHANK3* [46]. Chen *et al.* found a deletion at 22q13.3 in two male children with ASD and a duplication at 22q13.31-q13.33 in one male child with ASD from Taiwan, China [44]. In our study, we found a duplication at 22q13.31-q13.33 that overlaps *SHANK3* from two male children with ASD, indicating that the duplication at 22q13.31-q13.33 may play a key role in the etiology of ASD in our population. CNVs at 15q13.3 have been found to be involved in a variety of neuropsychiatric diseases, including intellectual disability/developmental delay, epilepsy, schizophrenia, and ASD [43, 49–51]. The relation between *CHRNA7* at 15q13.3 and neuropsychiatric disorder phenotype has been validated intensively [50]. In accordance with the discovery of Pinto *et al.* [28], we also found that a deletion of *CHRNA7* is associated with ASD risks. Except *CHRNA7* and *SHANK3*, we found CNVs-duplications (*DRD4*, *HRAS*, *OPHN1*, *SLC6A3*, and *TSC2*) and CNVs-deletions (*PTEN*). For *DRD4* and *HARS*, we found seven children with ASD had duplications at 11p15.5, which overlaps *DRD4* and *HARS*. Mutations in *DRD4* are associated with ASD risks [52–54]. The mRNA expression levels of *DRD4* in peripheral blood lymphocytes are higher in people with ASD than that in healthy controls [55, 56]. Herault *et al.* also found positive association between *HRAS* and autism in French-Caucasian [57, 58]. For *OPHN1* at Xq12-q13.1, Celestino-Soper *et al.* found a deletion of exons 7–15 of *OPHN1* at Xq12 in a male child with ASD [59]. In contrast, we found a male child with ASD had a duplication at Xq12-q13.1. For *SLC6A3* at 5p15.33-p15.2, Bowton *et al.* found *SLC6A3* coding variant Ala559Val is related to ASD [60]. We further found a child with ASD had a duplication at 5p15.33-p15.2. For *TSC2* at 16p13.3 and *PTEN* at 10q23.2-q23.31, Bourgeron *et al.* found that mutations in *TSC2* and *PTEN* activate the mTOR/PI3K pathway, associating with ASD risks [61]. We found duplications at 16p13.3 in two female children with ASD. PTEN loss involved in white matter pathology in human with ASD is consistent with in mouse models of Pten loss [62]. We revealed that deletions at 10q23.2-q23.31 overlapping *PTEN* in 13 male children with ASD, rather than 3 female children with ASD. Thus, these eight genes may be implicated in the etiology of ASD.

MiRNAs coded within CNVs are important functional variants, providing a new dimension to recognize the association between genotype and phenotype [63]. MiRNAs play vital roles in governing essential aspects of inhibitory transmission and interneuron development in nervous system [64]. Deletion or duplication of a chromosomal loci changes the levels of miRNAs which further impact on neuronal function and communication [38]. In our study, 12 candidate-susceptible miRNA-coding genes of ASD were identified (ten duplications [*MIR202*, *MIR210*, *MIR3178*, *MIR339*, *MIR4516*, *MIR4717*, *MIR483*, *MIR675*, *MIR6821*, and *MIR940*] and two deletions [*MIR107* and *MIR558*]). *BDNF*, a brain-derived neurotrophic factor and a member of the neurotrophic factor family, is a target gene of miR-202 [65]. Moreover, we further predicted that miR-4717-5p, miR-483-3p, and miR-940 also targeted *BNDF*. Skogstrand *et al.* found that lower BDNF levels in serum correlate with ASD risks [66, 67]. miR-339-5p has been found to be a drug target for Alzheimer's disease, and is low expressed in mature neurons and related to axon guidance [68, 69]. In our study, we found that miR-339-5p targets 42 genes associated with ASD risks. Among these genes, the association of *DIP2A* and ASD risks has been validated by our team [70]; moreover, *Dip2a* knockout mice exhibit autism-like behaviors, including excessive repetitive behavior and social novelty defects [71]. Notably, autism-like behaviours and germline transmission in *MECP2* transgenic monkeys corroborate association between miR-339-5p and *MECP2* [72]. In addition, miR-202-5p, miR-483-3p, and miR-940 also targets *MECP2*. For these reasons, miRNAs coded within CNVs may be implicated in ASD etiology.

Limitations

Our study had some limitations: (1) the sample size in our study may confer limited statistical power to discover significant findings; (2) genetic and environmental factors contribute to ASD risk; however, environmental factors were not available for us; and (3) de novo or inherited of the CNVs were not be classified because of the lack of data from parents.

Conclusions

We identified that 22 kinds of CNVs (six deletions and 16 duplications), eight protein-coding genes, and 12 miRNAs genes are implicated in ASD risks, conferring perception to further reveal ASD etiology.

Abbreviations

ASD Autism spectrum disorder

CNV Copy number variation

Kb kilobase

aCGH Array-based comparative genomic hybridization

miRNAs MicroRNAs

GO Gene Ontology

BP biological processes

CC cellular component

MF molecular function

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of Jilin University.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated during the current study are available from the corresponding authors on reasonable request.

Competing of interests

The authors declare no conflict of interest.

Funding

This study was funded by Graduate Innovation Fund of Jilin University (grant number: 101832018C064); China Scholarship Council (grant number: 201806170230); National Natural Science Foundation of China (grant number: 81973120); the Ministry of Science and Technology of the People´s Republic of China (grant number: 2015DFA31580 and 2018YFC1311600); and Science and Technology Department of Jilin Province (grant number: 20150101130JC).

Author's contributions

Study design: Shuang Qiu, Yi Cheng, and Yawen Liu. Experiment: Shuang Qiu. Data collection, analysis and interpretation: Shuang Qiu, Xianling Cong, Yan Li, Jikang Shi, Yingjia Qiu, Yanbo Guo, Yulu Gu, Yong Li, Xiaojuan Zhu, Yunkai Liu, and Yichun Qiao. Drafting of the manuscript: Shuang Qiu. Critical revision of the manuscript: Yi Cheng and Yawen Liu. Approval of the final version for publication: all co-authors.

Contributor Information

Shuang Qiu, Email: 245412408@qq.com

Xianling Cong, Email: congxl@jlu.edu.cn

Yan Li, Email: 95614837@qq.com

Jikang Shi, Email: 982739713@qq.com

Yingjia Qiu, Email: 2477305445@qq.com

Yanbo Guo, Email: 17519489501@163.com

Yulu Gu, Email: gyluisme@163.com

Yong Li, Email: l_yong@jlu.edu.cn

Xiaojuan Zhu, Email: zhuxj720@nenu.edu.cn

Yunkai Liu, Email: 52993055@qq.com

Yichun Qiao, Email: qiaoyichun@jlu.edu.cn

Yi Cheng, Email: chengyi@jlu.edu.cn

Yawen Liu, Email: ywliu@jlu.edu.cn

Acknowledgments

The authors would like to thank the subjects in this study.

References

1. WHO. Autism spectrum disorders and other developmental disorders: From raising awareness to building capacity. Geneva: World Health Organization; 2013.
2. Sahin M, Sur M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. 350. New York: Science; 2015.
3. Kim JY, Son MJ, Son CY, et al. Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence. The lancet Psychiatry. 2019;6:590–600.
4. Xu Q, Liu YY, Wang X, et al. Autism-associated CHD8 deficiency impairs axon development and migration of cortical neurons. Molecular autism. 2018;9:65.
5. Peñagarikano O, Abrahams BS, Herman EI, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell. 2011;147:235–46.
6. Li Y, Qiu S, Zhong W, et al. Association Between DCC Polymorphisms and Susceptibility to Autism Spectrum Disorder. Journal of autism and developmental disorders. 2020.
7. da Silva Montenegro EM, Costa CS, Campos G, et al. Meta-Analyses Support Previous and Novel Autism Candidate Genes: Outcomes of an Unexplored Brazilian Cohort. Autism research: official journal of the International Society for Autism Research. 2020;13:199–206.
8. Qiu S, Li Y, Bai Y, et al. SHANK1 polymorphisms and SNP-SNP interactions among SHANK family: A possible cue for recognition to autism spectrum disorder in infant age. Autism research: official journal of the International Society for Autism Research. 2019;12:375–83.
9. Bai Y, Qiu S, Li Y, et al. Genetic association between SHANK2 polymorphisms and susceptibility to autism spectrum disorder. IUBMB Life. 2018;70:763–76.
10. Wang J, Gong J, Li L, et al. Neurexin gene family variants as risk factors for autism spectrum disorder. Autism research: official journal of the International Society for Autism Research. 2018;11:37–43.
11. Marui T, Funatogawa I, Koishi S, et al. Association between autism and variants in the wingless-type MMTV integration site family member 2 (WNT2) gene. The international journal of neuropsychopharmacology. 2010;13:443–9.

12. Sandin S, Lichtenstein P, Kuja-Halkola R, et al. The familial risk of autism. Jama. 2014;311:1770–7.
13. Ratto AB, Mesibov GB. Autism spectrum disorders in adolescence and adulthood: Long-term outcomes and relevant issues for treatment and research. Science China Life sciences. 2015;58:1010–5.
14. RK CY, Merico D, Bookman M, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nature neuroscience. 2017;20:602–11.
15. Safari MR, Omrani MD, Noroozi R, et al. Synaptosome-Associated Protein 25 (SNAP25) Gene Association Analysis Revealed Risk Variants for ASD, in Iranian Population. Journal of molecular neuroscience: MN. 2017;61:305–11.
16. Murdoch JD, State MW. Recent developments in the genetics of autism spectrum disorders. Curr Opin Genet Dev. 2013;23:310–5.
17. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36:949–51.
18. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. 305. New York: Science; 2004. pp. 525–8.
19. Xu B, Roos JL, Levy S, et al. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008;40:880–5.
20. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316:445–9.
21. Glessner JT, Li J, Wang D, et al. Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders. Genome medicine. 2017;9:106.
22. Erikson GA, Deshpande N, Kesavan BG, et al. SG-ADVISER CNV: copy-number variant annotation and interpretation. Genetics in medicine: official journal of the American College of Medical Genetics. 2015;17:714–8.
23. Chen CH, Chen HI, Chien WH, et al. High resolution analysis of rare copy number variants in patients with autism spectrum disorder from Taiwan. Scientific reports. 2017;7:11919.
24. Rosti RO, Sadek AA, Vaux KK, et al. The genetic landscape of autism spectrum disorders. Dev Med Child Neurol. 2014;56:12–8.
25. Levy D, Ronemus M, Yamrom B, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron. 2011;70:886–97.
26. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011;70:863–85.
27. Liu X, Takumi T. Genomic and genetic aspects of autism spectrum disorder. Biochem Biophys Res Commun. 2014;452:244–53.
28. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466:368–72.
29. Yin CL, Chen HI, Li LH, et al. Genome-wide analysis of copy number variations identifies PARK2 as a candidate gene for autism spectrum disorder. Molecular autism. 2016;7:23.
30. Marrale M, Albanese NN, Cali F, et al. Assessing the impact of copy number variants on miRNA genes in autism by Monte Carlo simulation. PloS one. 2014;9:e90947.
31. Park H, Kim JI, Ju YS, et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat Genet. 2010;42:400–5.
32. Association AP. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington DC: American Psychiatric Publishing, Inc; 2013.
33. Qiu S, Li Y, Li Y, et al. Association between SHANK3 polymorphisms and susceptibility to autism spectrum disorder. Gene. 2018;651:100–5.
34. Gu Z, Gu L, Eils R, et al. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811–2.
35. Kearney HM, Thorland EC, Brown KK, et al. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genetics in medicine: official journal of the American College of Medical Genetics. 2011;13:680–5.
36. Li L, Li Q, Wang Q, et al. Rare copy number variants in the genome of Chinese female children and adolescents with Turner syndrome. Bioscience reports. 2019;39.
37. Dimassi S, Labalme A, Lesca G, et al. A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. Epilepsia. 2014;55:370–8.

38. Vaishnavi V, Manikandan M, Tiwary BK, et al. Insights on the functional impact of microRNAs present in autism-associated copy number variants. PloS one. 2013;8:e56781.
39. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16:284–7.
40. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nature methods. 2015;12:697.
41. Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nature reviews Genetics. 2008;9:341–55.
42. Kushima I, Aleksic B, Nakatochi M, et al. Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. Cell reports. 2018;24:2838–56.
43. Gillentine MA, Yin J, Bajic A, et al. Functional Consequences of CHRNA7 Copy-Number Alterations in Induced Pluripotent Stem Cells and Neural Progenitor Cells. Am J Hum Genet. 2017;101:874–87.
44. Chen CH, Chen HI, Liao HM, et al. Clinical and molecular characterization of three genomic rearrangements at chromosome 22q13.3 associated with autism spectrum disorder. Psychiatr Genet. 2017;27:23–33.
45. Phelan MC. Deletion 22q13.3 syndrome. Orphanet J Rare Dis. 2008;3:14.
46. Phelan K, McDermid HE. The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). Molecular syndromology. 2012;2:186–201.
47. Soorya L, Kolevzon A, Zweifach J, et al. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. Molecular autism. 2013;4:18.
48. Sarasua SM, Boccuto L, Sharp JL, et al. Clinical and genomic evaluation of 201 patients with Phelan-McDermid syndrome. Human genetics. 2014;133:847–59.
49. Lowther C, Costain G, Stavropoulos DJ, et al. Delineating the 15q13.3 microdeletion phenotype: a case series and comprehensive review of the literature. Genetics in medicine: official journal of the American College of Medical Genetics. 2015;17:149–57.
50. Gillentine MA, Schaaf CP. The human clinical phenotypes of altered CHRNA7 copy number. Biochem Pharmacol. 2015;97:352–62.
51. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455:232–6.
52. Klein M, van Donkelaar M, Verhoef E, et al. Imaging genetics in neurodevelopmental psychopathology. American journal of medical genetics Part B Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics. 2017;174:485–537.
53. Eisenegger C, Knoch D, Ebstein RP, et al. Dopamine receptor D4 polymorphism predicts the effect of L-DOPA on gambling behavior. Biol Psychiatry. 2010;67:702–6.
54. Gadow KD, Devincent CJ, Olvet DM, et al. Association of DRD4 polymorphism with severity of oppositional defiant disorder, separation anxiety disorder and repetitive behaviors in children with autism spectrum disorder. Eur J Neurosci. 2010;32:1058–65.
55. Sener EF, Taheri S, Sahin MC, et al. Altered Global mRNA Expressions of Pain and Aggression Related Genes in the Blood of Children with Autism Spectrum Disorders. Journal of molecular neuroscience: MN. 2019;67:89–96.
56. Emanuele E, Boso M, Cassola F, et al. Increased dopamine DRD4 receptor mRNA expression in lymphocytes of musicians and autistic individuals: bridging the music-autism connection. Neuroendocrinol Lett. 2010;31:122–5.
57. Herault J, Perrot A, Barthelemy C, et al. Possible association of c-Harvey-Ras-1 (HRAS-1) marker with autism. Psychiatry research. 1993;46:261–7.
58. Herault J, Petit E, Martineau J, et al. Autism and genetics: clinical approach and association study with two markers of HRAS gene. Am J Med Genet. 1995;60:276–81.
59. Celestino-Soper PB, Shaw CA, Sanders SJ, et al. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. Human molecular genetics. 2011;20:4360–70.
60. Bowton E, Saunders C, Reddy IA, et al. SLC6A3 coding variant Ala559Val found in two autism probands alters dopamine transporter function and trafficking. Translational psychiatry. 2014;4:e464.
61. Bourgeron T. A synaptic trek to autism. Curr Opin Neurobiol. 2009;19:231–4.
62. Frazier TW. Autism Spectrum Disorder Associated with Germline Heterozygous PTEN Mutations. Cold Spring Harbor perspectives in medicine. 2019;9:a037002.

63. Marcinkowska M, Szymanski M, Krzyzosiak WJ, et al. Copy number variation of microRNA genes in the human genome. BMC Genomics. 2011;12:183.
64. Hsu R, Schofield CM, Dela Cruz CG, et al. Loss of microRNAs in pyramidal neurons leads to specific changes in inhibitory synaptic transmission in the prefrontal cortex. Mol Cell Neurosci. 2012;50:283–92.
65. Xin C, Xia J, Liu Y, et al. MicroRNA-202-3p Targets Brain-Derived Neurotrophic Factor and Is Involved in Depression-Like Behaviors. Neuropsychiatr Dis Treat. 2020;16:1073–83.
66. Skogstrand K, Hagen CM, Borbye-Lorenzen N, et al. Reduced neonatal brain-derived neurotrophic factor is associated with autism spectrum disorders. Translational psychiatry. 2019;9:252.
67. Katoh-Semba R, Wakako R, Komori T, et al. Age-related changes in BDNF protein levels in human serum: differences between autism cases and normal controls. International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience. 2007;25:367–72.
68. Long JM, Ray B, Lahiri DK. MicroRNA-339-5p down-regulates protein expression of β-site amyloid precursor protein-cleaving enzyme 1 (BACE1) in human primary brain cultures and is reduced in brain tissue specimens of Alzheimer disease subjects. J Biol Chem. 2014;289:5184–98.
69. Liu DZ, Ander BP, Tian Y, et al. Integrated analysis of mRNA and microRNA expression in mature neurons, neural progenitor cells and neuroblastoma cells. Gene. 2012;495:120–7.
70. Li Y, Qiu S, Zhong W, et al. rs1007893 and rs2070435 in DIP2A are associated with visual-reaction-behavior phenotype in Chinese Han population with autism spectrum disorder. Research in Autism Spectrum Disorders. 2020;70:101475.
71. Ma J, Zhang LQ, He ZX, et al. Autism candidate gene DIP2A regulates spine morphogenesis via acetylation of cortactin. PLoS Biol. 2019;17:e3000461.
72. Liu Z, Li X, Zhang JT, et al. Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. Nature. 2016;530:98–102.

Figures

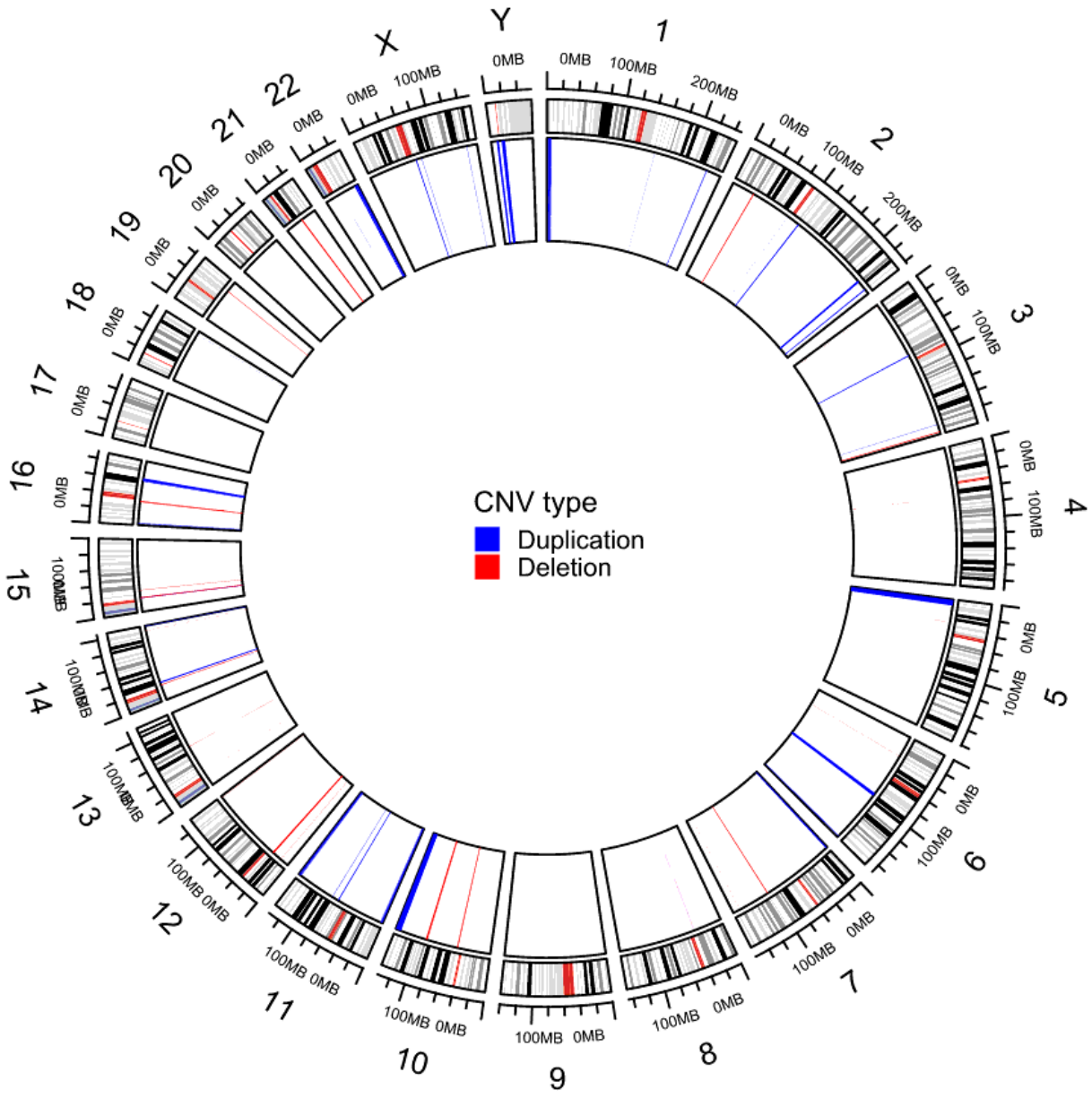

Figure 1

The distribution of CNVs on genome-wide chromosomes

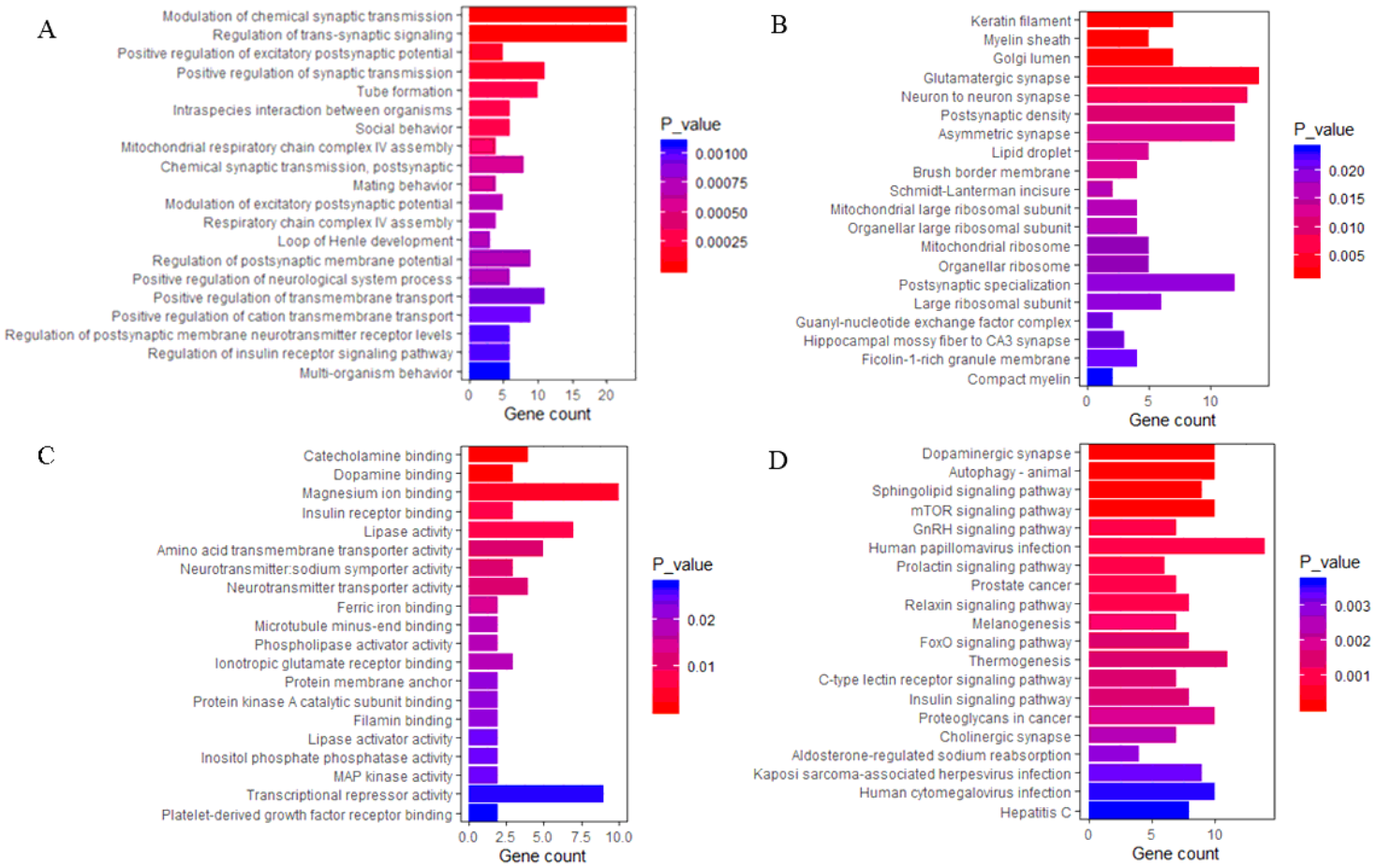

Figure 2

Function and pathway enrichment of the 511 genes from 22 potential pathogenic CNVs Note: Top 20 annotations or pathways ordered by P_value. A: Biological Process; B: Cellular Component; C: Molecular Function; D: Kyoto encyclopedia of genes and genomes pathway. The ordinate represents the gene ontology function, and the abscissa represents the number of genes enriched to the term. P_value indicate the degree of enrichment, with smaller P_value indicating genes that are more likely to play significant functional roles.

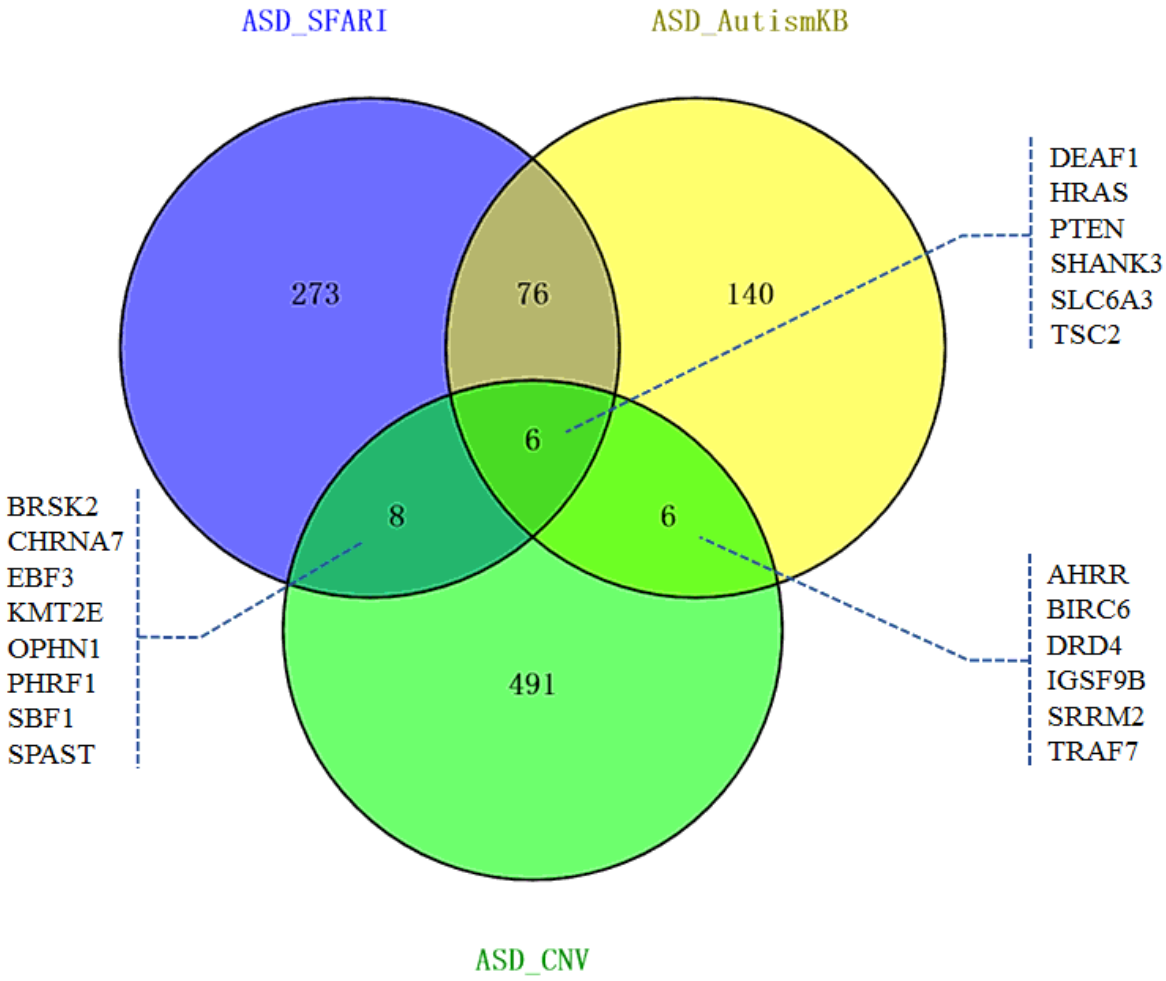

Figure 3

Venn diagram of the number of genes among ASD_SFARI, ASD_AutismKB and ASD_CNV Note: We denote genes in our candidate CNVs for ASD as “ASD_CNV”, the 363 high confidence and strong candidate autism risk genes in SFARI as “ASD_SFARI”, and the 228 high confidence autism related genes in AutismKB as “ASD_AutismKB”.

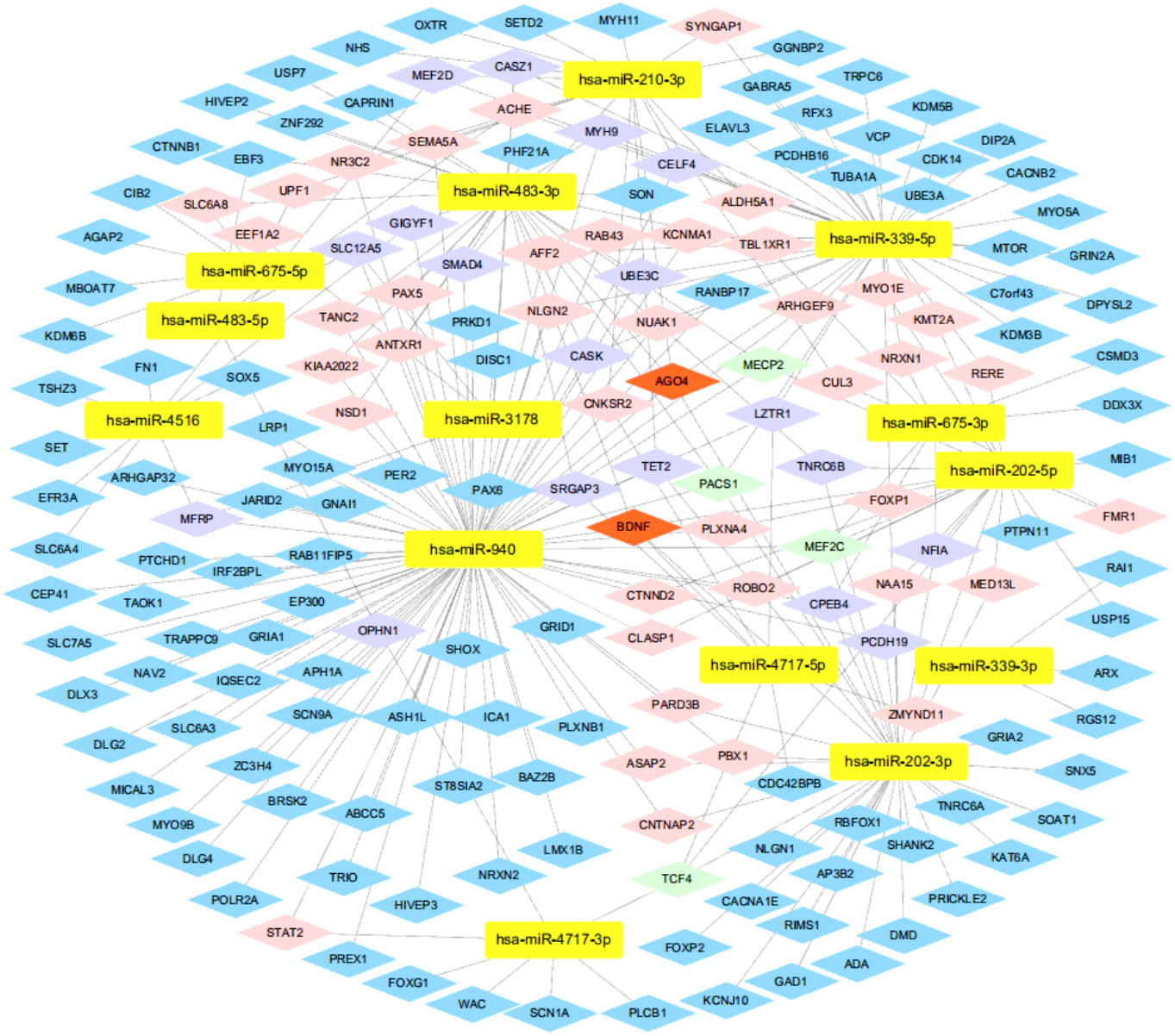

Figure 4

CNVs-miRNAs-target genes interaction network in ASD (duplication) Note: Yellow rectangles represent the miRNAs in pathogenic CNVs regions, while miRNAs-tatget genes are denoted by diamonds. Blue, pink, purple, light green, and red diamonds represent different target genes which are targeted by 1, 2, 3, 4, and 5 miRNAs respectively.

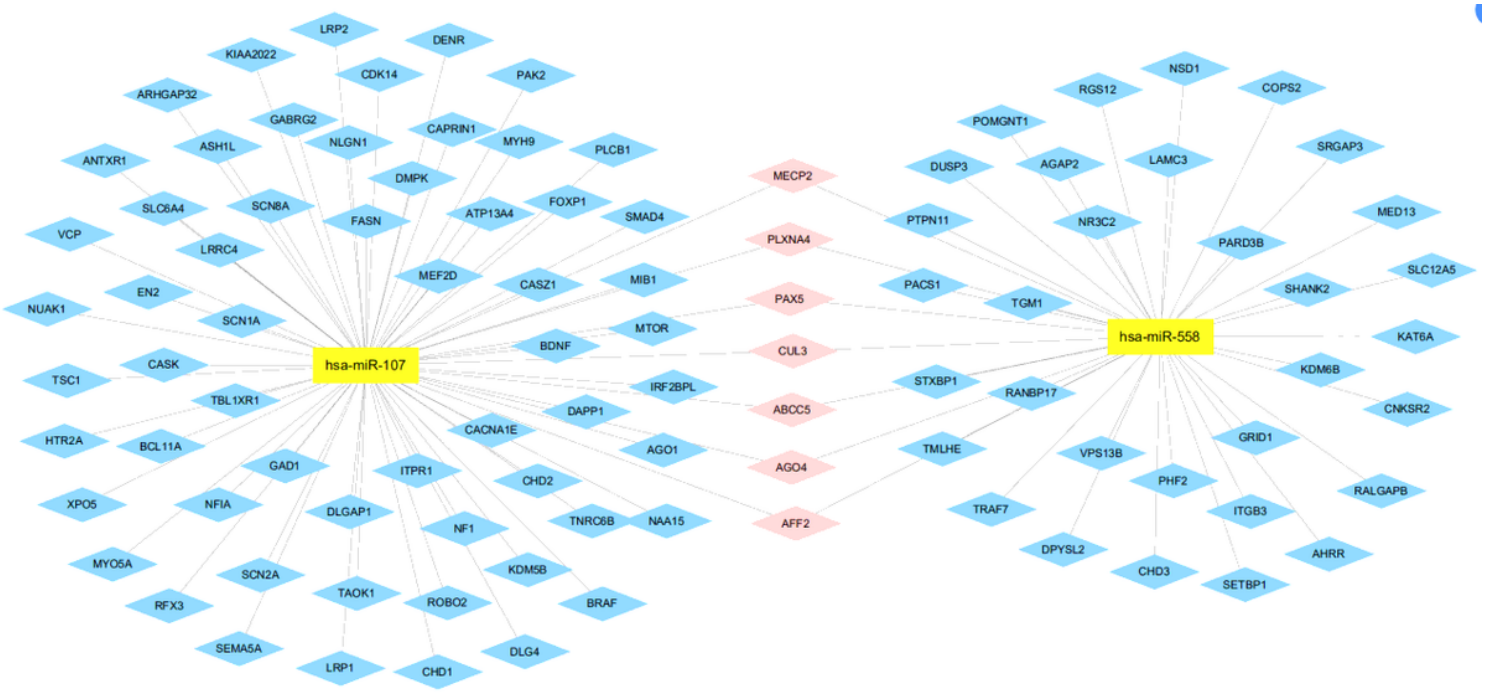

Figure 5

CNVs-miRNAs-target genes interaction network in ASD (deletion) Note: Yellow rectangles represent the miRNAs in pathogenic CNVs regions, while miRNAs-tatget genes are denoted by diamonds. Blue and pink diamonds represent different target genes which are targeted by one and two miRNAs respectively.

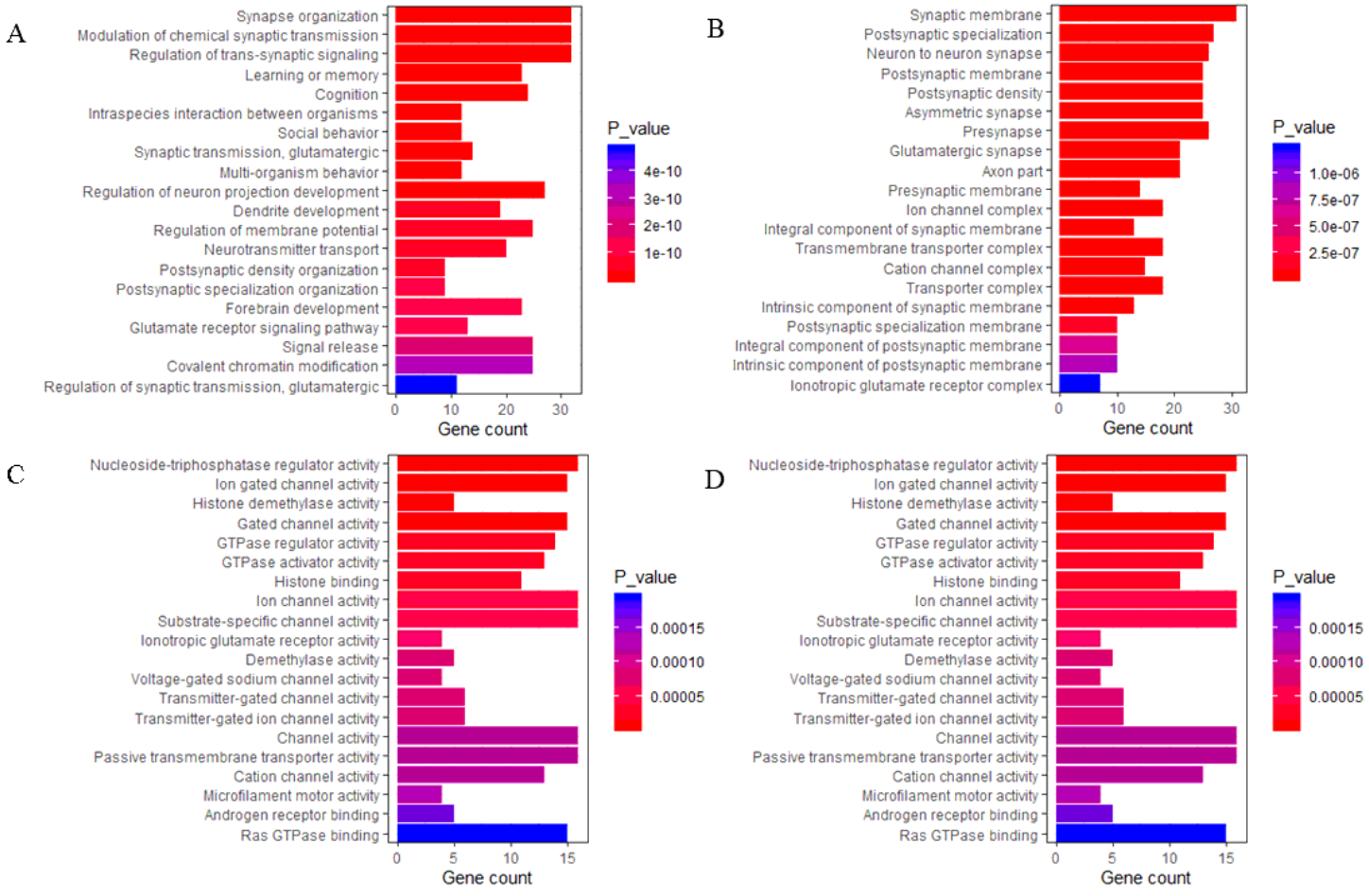

Figure 6

Function and pathway enrichment of the 219 target genes of 16 potential pathogenic miRNAs in CNVs Note: Top 20 annotations or pathways ordered by P_value. A: Biological Process; B: Cellular Component; C: Molecular Function; D: Kyoto encyclopedia of genes and genomes pathway. The ordinate represents the gene ontology function, and the abscissa represents the number of genes enriched to the term. P_value indicate the degree of enrichment, with smaller P_value indicating genes that are more likely to play significant functional roles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.docx