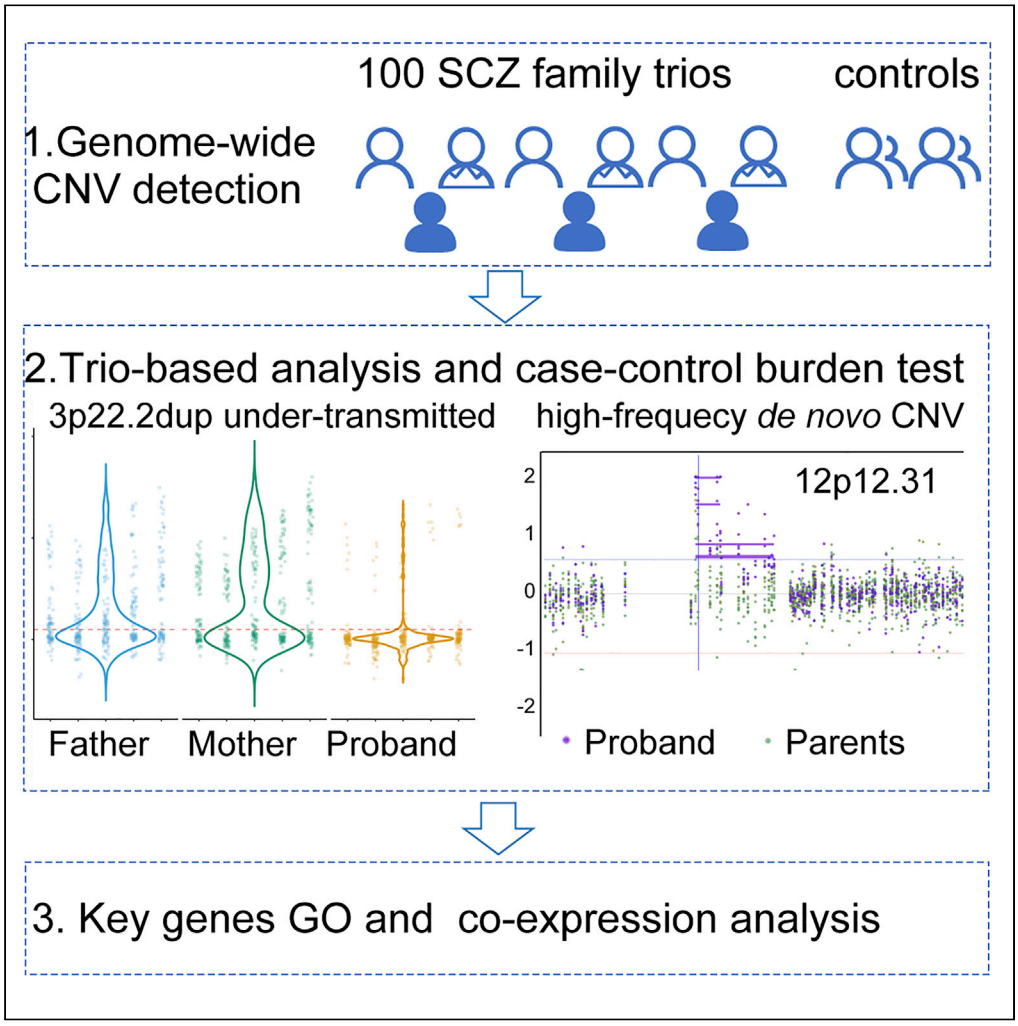
iScience

CellPress
OPEN ACCESS

Article

Genome-wide study of copy number variation implicates multiple novel loci for schizophrenia risk in Han Chinese family trios



Xi Wu, Cong Huai, Lu Shen, ..., Lin He, Chunling Wan, Shengying Qin

helin@sjtu.edu.cn (L.H.)
clwan@sjtu.edu.cn (C.W.)
chinsir@sjtu.edu.cn (S.Q.)

Highlights

Copy number changes of 3p22.2, 7p13, 11p15.1, 12p13, and 6p12 are associated with SCZ

The under-transmission of *CTDSPL* contributes to SCZ susceptibility

Wu et al., iScience *24*, 102894
August 20, 2021
https://doi.org/10.1016/j.isci.2021.102894

iScience

CellPress
OPEN ACCESS

Article

Genome-wide study of copy number variation implicates multiple novel loci for schizophrenia risk in Han Chinese family trios

Xi Wu,[1,5] Cong Huai,[1,5] Lu Shen,[1] Mo Li,[1] Chao Yang,[1] Juan Zhang,[1] Luan Chen,[1] Wenli Zhu,[2] Lingzi Fan,[3] Wei Zhou,[1] Qinghe Xing,[4] Lin He,[1,*] Chunling Wan,[1,*] and Shengying Qin[1,6,*]

SUMMARY

Schizophrenia (SCZ) is a severe neuropsychiatric disorder that affects 1% of the global population. Copy number variations (CNVs) have been shown to play a critical role in its pathophysiology; however, only case-control studies on SCZ susceptibility CNVs have been conducted in Han Chinese. Here, we performed an array comparative genomic hybridization-based genome-wide CNV analysis in 100 Chinese family trios with SCZ. Burden test suggested that the SCZ probands carried more duplications than their healthy parents and unrelated healthy controls. Besides, five CNV loci were firstly reported to be associated with SCZ here, including both unbalanced transmitted CNVs and enriched *de novo* CNVs. Moreover, two genes (*CTDSPL* and *MGAM*) in these CNVs showed significant SCZ relevance in the expression level. Our findings support the crucial role of CNVs in the etiology of SCZ and provide new insights into the underlying mechanism of SCZ pathogenesis.

INTRODUCTION

Schizophrenia (SCZ) is a severe psychiatric disorder that affects approximately 1% of the population worldwide (Howes and Murray, 2014). Both genetic and environmental factors are involved in the pathophysiology of SCZ (Howes and Murray, 2014), and twin studies have revealed a predominantly genetic basis (Boshes et al., 2012). However, SCZ is a complex disease that lots of genetic factors contribute to its development.

As a major kind of contributors to structural genetic diversity, copy number variations (CNVs) are key players in psychiatric and neurodevelopmental disorders, including SCZ, autism spectrum disorder (ASD), and intellectual disability (Kirov et al., 2014; Weiss et al., 2008), and are also associated with neuropsychiatric traits in both disease and general population (Guyatt et al., 2018; Hubbard et al., 2020). Case-control association analysis has revealed the global and large rare CNV burden in SCZ cases (Marshall et al., 2017; Rees and Kirov, 2021) and has discovered numerous CNV hotspots for SCZ susceptibility, including 1q21.1 deletion/duplication (del/dup), 2p16.3 del (*NRXN1*), 3q29 del/dup, 7p36.3 dup (*VIPR2*), 15q13.3 del, 16p11.2 dup, 16p13.1 dup, 17p12 del, and so on (Ingason et al., 2011; Kirov et al., 2009, 2012; Levinson et al., 2011; McCarthy et al., 2009; Szatkiewicz et al., 2014; Yuan et al., 2017). These findings, together with genetic variation studies, outline the possible gene pathways involved in SCZ development, where that anatomical abnormality, dysfunctional neurotransmission, and stress-associated signaling cascade dysregulation are the three main causes of SCZ (Fatani et al., 2017; Marshall et al., 2017; Rees and Kirov, 2021). However, the detailed mechanisms underlying SCZ have yet to be fully understood.

Besides, most of these studies were conducted in Caucasian populations. Given the genetic differences among ethics, studies in Chinese patients with SCZ have validated several known CNVs (e.g., 1q21.1 del, 15q11.2 del, 7q11.23 dup, and 16p11.2 dup) and also identified some new loci, such as 1p36.32 dup, 10p12.1 dup, and 13q13.3 dup (Li et al., 2016; Yuan et al., 2017). These findings indicate that CNVs play an important role in the etiology of SCZ, and more investigations in different populations may help us to understand the complex mechanism.

Unlike case-control studies, family-based analysis can effectively avoid spurious results caused by population stratification and minimize the impact of environmental factors. Based on this method, the importance of *de*

[1]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai, 200030, China

[2]The Fourth People's Hospital of Wuhu, Wuhu, Anhui, 241000, China

[3]Zhumadian Psychiatric Hospital, Zhumadian, Henan, 463000, China

[4]Children's Hospital & Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China

[5]These authors contribute equally

[6]Lead contact

*Correspondence: helin@sjtu.edu.cn (L.H.), clwan@sjtu.edu.cn (C.W.), chinsir@sjtu.edu.cn (S.Q.)

https://doi.org/10.1016/j.isci.2021.102894

Table 1. Global rare CNV burden (>10 kb) in Chinese schizophrenia trios

Dup + Del	SCZ probands (n = 94)	Unaffected parents (n = 159)	Controls (n = 73)
N	420	640	276
Rate	4.468	4.025	3.781
Prop	0.713	0.654	0.863
AVG (kb)	118.4	64.2	52.14
P For rate	0.213	0.404	/
Dup			
N	373	472	156
Rate	3.968	2.969	2.137
Prop	0.692	0.598	0.8219
AVG (kb)	129.7	78.9	43.2
P For rate	**0.0002**	**0.021**	/
Del			
N	47	168	120
Rate	0.50	1.057	1.644
Prop	0.436	0.440	0.247
AVG (kb)	15.02	16.96	190.9
P For rate	0.972	0.806	/

Table shows an analysis of global CNV burden in SCZ probands (n = 94)/unaffected healthy parents (n = 159) vs unaffected controls (n = 73). CNVs were previously filtered with population frequencies <1% according to DGV database. p values were calculated by 1-side empirical significance tests after 10,000 permutations. N, number of total CNV segments. Rate, average CNV numbers of each sample; Prop, proportion of sample with one or more CNV; AVG (kb), average length of each CNV. Significant correlations ($P < 0.05$) are bolded.

novo CNVs has also been revealed with SCZ (Kirov et al., 2012). Therefore, in this study, we performed a family-based genome-wide study of CNVs using 100 SCZ trios of Han Chinese ethnicity. We have verified several previously reported loci and discovered five novel CNV regions related to SCZ, explored the related Gene Ontology (GO) pathways, and also found that *CTDSPL* and *MGAM* in the aforementioned CNVs showed significant SCZ relevance in the expression level. This is the first genome-wide CNV association study with the most trios in the Chinese population so far, and it may be help in the diagnosis and therapy of SCZ.

RESULTS

Rare CNV burden analysis revealed SCZ probands carried more duplications than parents

We recruited 100 Chines family trios with SCZ (SCZ-affected offspring and both of their parents) and 75 controls (without mental diseases) in this study. All DNA samples were applied to CNV detection using Agilent 1 × 1M SurePrint G3 array comparative genomic hybridization (aCGH) microarray. After quality control and kinship analyses, 94 trios with SCZ and 73 controls were involved in the following analysis. The demographic information for all the valid participants is listed in Table S1. There is no kinship between different trios or controls.The data for aCGH microarray has been submitted to ArrayExpress database with accession no. E-MTAB-8075.

A total of 5,940 CNVs were identified from the 282 individuals of the remaining 94 family trios (4,259 from parents and 1,681 from SCZ proband offspring) under strict calling threshold by both aberration detection method 2 (ADM-2) and hidden markov model (HMM) algorithm. The high-frequency CNVs enriched in the SCZ probands were in accordance with the results from previous discoveries of case-control association studies, including duplications on 2p11.2 (Chen et al., 1998), 3p26.1 (Walsh et al., 2008), 2p16.3 (Marshall et al., 2017; Tam et al., 2009), and 11q14.1 (Table S2).

To investigate the global impact of CNVs on susceptibility to SCZ, we compared the rare CNV burdens between the SCZ probands and their unaffected parents. Rare CNVs were filtered by their frequencies (<1%) in Database of Genomic Variants, and unrelated controls were set as the comparison reference to each

Table 2. Large rare CNVs in SCZ probands and parents

FamID	CNV region	Start	Stop	Interval (bp)	CNV type	Origin
In probands						
28	2p11.2 - p11.1	90,265,060	9,190,6643	1,641,583	dup	*De novo*
55	9p13.1 - p11.2	38,768,232	43,836,428	5,068,196	dup	*De novo*
5	9p13.1 - p11.2	39,098,753	47,212,417	8,113,664	dup	*De novo*
61	9p13.1 - p11.2	39,156,895	43,836,428	4,679,533	dup	*De novo*
63	9p13.1 - p11.2	39,156,895	43,836,428	4,679,533	dup	*De novo*
96	9p11.2	43,686,924	47,212,417	3,525,493	dup	*De novo*
In parents						
104	2q32.3	193,664,633	19,490,4714	1,240,081	dup	/
2	9p11.2	43,686,924	47,212,417	3,525,493	dup	/

group, respectively. As shown in Table 1, no significant difference was found in the total CNV burden between both groups. However, SCZ probands carried significantly more duplications than controls (ratio = 1.86, $P = 0.0002$) while the counterpart ratio of parent groups is slightly lower (ratio = 1.39, $P = 0.021$), which means that each SCZ proband has an average of more duplications (proband/parent ratio = 1.47) and fewer deletions (ratio = 0.47) compared with their healthy parents. Besides, six large rare CNVs (>1 Mb in length) were found only in SCZ probands (Table 2).

Four novel candidate SCZ-associated loci were revealed from transmission disequilibrium analyses

As both inherited and uninherited (*de novo*) CNVs may contribute to the pathogenesis of SCZ, we firstly searched for transmission disequilibrium CNVs associated with susceptibility to SCZ between generations in all the 94 trios. In these analyses, 101 well-separated multiclass CNVs (with 1,495 probes) were chosen. In all, 11 of these CNVs (in 10 regions) were significantly associated with the risk for SCZ after Bonferroni correction ($P < 3.3 \times 10^{-5}$; Table 3). The results suggested that five of these CNVs carried by healthy parents were under-transmitted to their SCZ offspring, which means that the transmission of these CNVs between the two generations violated Mendel's laws of genetics, and fewer probands than expected carried these CNVs. So, these five CNVs (including 3p22.2 dup, 7p15.2 dup, 7p13 dup, 9q21.13 dup, and 11p15.1 dup) may have been protective factors against SCZ. The other six CNVs were over-transmitted, which could be considered as risk factors. Of the 10 unbalanced transmitted CNVs, four were novel candidate loci for SCZ, including the under-transmitted 3p22.2 dup (in gene *CTDSPL*), 7p13 dup (in gene *HECW1*), and the over-transmitted 11p15.1 del (in gene *MRGPRX1*) and 12p13.2 del (Table 3, Figure 1). These unbalanced transmissions were also validated by real-time polymerase chain reaction (Figures S1 and S2). In addition, we found a candidate CNV on 7p15.2 spanning a 43 kb region, which was smaller than that of previous reports (Aliyu et al., 2006; Fallin et al., 2003) and contained only one gene, *SKAP2*.

Two new candidate SCZ-associated loci were revealed from de novo mutation analyses

Analyses of *de novo* mutations in all 94 trios revealed that 101 *de novo* CNVs occurred in more than one family trio, and the other 122 *de novo* CNVs occurred only once. Among them, 222 CNVs from 59 trios at 104 CNV regions are rare CNVs.

There were nine high-frequency *de novo* CNVs (Table 4). Besides the known risk CNVs reported previously, a 5.59 kb deletion at 6p12.1 (in gene *MLIP*) was a novel candidate region for risk factors of SCZ. We also narrowed a candidate CNV on 7q34 down to a 23.84 kb region containing only one gene, *MGAM* (Smith et al., 2010). In addition, we observed a *de novo* mutation hotspot 6p21.32 (Figure S3), in which there were 21 CNVs located. CNV breakpoints in 6p21.32 mapping to gene regions of the major histocompatibility complex showed high risk for predisposition to SCZ.

CTDSPL and *MGAM* were predicted as potential pathogenic genes

We then carried out the prediction of SCZ pathogenic potential of the 6 newly discovered candidate genes (*CTDSPL*, *HECW1*, *MRGPRX1*, *SKAP2*, *MLIP*, and *MGAM*) from transmission disequilibrium and

Table 3. CNVs with transmission disequilibrium in Chinese SCZ trios

CNV loci		Position (Mb)	Interval Size (kb)	*P*	CNV type	Transmission[a]	Involved genes	Ref
3	p26.1	4.06–4.15	82.06	1.02E-05	dup	O		(Yu et al., 2017)
3	**p22.2**	**37.98–37.98**	**5.24**	5.39E-06	**dup**	**U**	*CTDSPL*	
3	q29	192.87–192.88	8.13	9.11E-05	del	O		(Li et al., 2016)
5	q31.2	137.181–137.187	6.35	1.15E-06	dup	O		(Baron, 2001)
7	p15.2	26.89–26.94	48.01	2.43E-05	dup	U	*SKAP2*	
7	**p13**	**43.42–43.42**	**9.81**	9.94E-06	**dup**	**U**	*HECW1*	
9	q21.13	78–78.01	6.68	4.09E-07	dup	U		(Chen et al., 2016)
11	**p15.1**	**18.95–18.96**	**8.7**	2.74E-07	**del** **dup**	**O** **U**	*MRGPRX1*	
12	**p13.2**	**11.52–11.54**	**16.51**	1.00E-05	**del**	**O**		
22	q11.22	23.04–23.24	205.07	1.60E-05	dup	O	*mir650*, *mir557*, *IGLL5*	

Transmission disequilibrium CNVs between generations tested by PedCNV. The novelly found risk CNV for SCZ are bolded.
[a]U indicates undertransmission of the CNV from parents to probands, and O indicates overtransmission of the CNV. (dup)/(del) means the type of CNV that over or lower transmitted. Involved genes are genes located in/at the CNV region.

de novo CNV analyses. Based on the results of the weighted gene co-expression network analysis of 11,688 samples sourced from PsychENCODE (Gandal et al., 2018), *CTDSPL* showed a strong correlation with SCZ. *CTDSPL* is in the cluster module geneM3, which contains 19 SCZ risk genes out of 25 hub genes (Figure 2A) and has a significant eigengene association for SCZ (P = 0.00037, false discovery rate = 0.0024, beta = 0.004). The expression mode of *CTDSPL* in patients with SCZ and in non-neuropsychiatric controls was also similar to that of the other SCZ risk genes in the same module (Figure 2B). Besides, *MGAM* was over-expressed in brain tissues of patients with SCZ, while *HECW1* was down-regulated in brain tissues of patients with ASD based on the database, implying that they may play a role in SCZ pathogenesis.

Finally, the 134 genes within the susceptible CNVs discovered by multiple analyses (large-rare CNVs, transmission disequilibrium CNVs, high-frequency *de novo* CNVs, and rare *de novo* CNVs) in this study were applied to GO analysis. The significantly enriched pathways are "morphogenesis of an epithelial sheet", "synapse organization", "stimuli-sensing channels", and four other pathways (Figure 2C).

DISCUSSION

SCZ is a complex disease caused by various polygenetic risks and environmental factors. CNVs, as an important kind of structural variations, may not only affect gene expression within the CNV region in a dose-dependent manner (Chaignat et al., 2011) but also attribute to chromosomal structural changes and affect global gene interaction or gene network (Takumi and Tamada, 2018). A great number of CNVs have been reported to be associated with SCZ risk. However, most studies on the risk CNVs in Chinese population have been found by case-control comparisons (Li et al., 2016; Zhao et al., 2013), which are better for identifying common variations (Pritchard and Cox, 2002). As linkage associations based on family trios may provide supplementary information for identifying and characterizing the full range of variants related to disease susceptibility to disease (Ott et al., 2011), in the present study, we analyzed CNVs in 94 family trios with SCZ to identify potential risk loci associated with SCZ. Unlike previous trio-based studies on Chinese populations (Liu et al., 2005; Wang et al., 2008; Xi et al., 2004; Zhang et al., 2005; Zou et al., 2005) which have mainly focused on candidate genes, this is a comprehensive genome-wide CNV association study with the largest number of Chinese Han family trios so far.

The result of our burden analyses corroborates the importance of large and rare CNVs in the development of SCZ (Kirov et al., 2012; Li et al., 2016; The International Schizophrenia Consortium, 2008; Xu et al., 2008). SCZ probands carried more and larger duplications than their healthy parents, which were mainly *de novo* mutations. However, these SCZ probands lost more than 80% (138/168) of their parents' rare deletions.

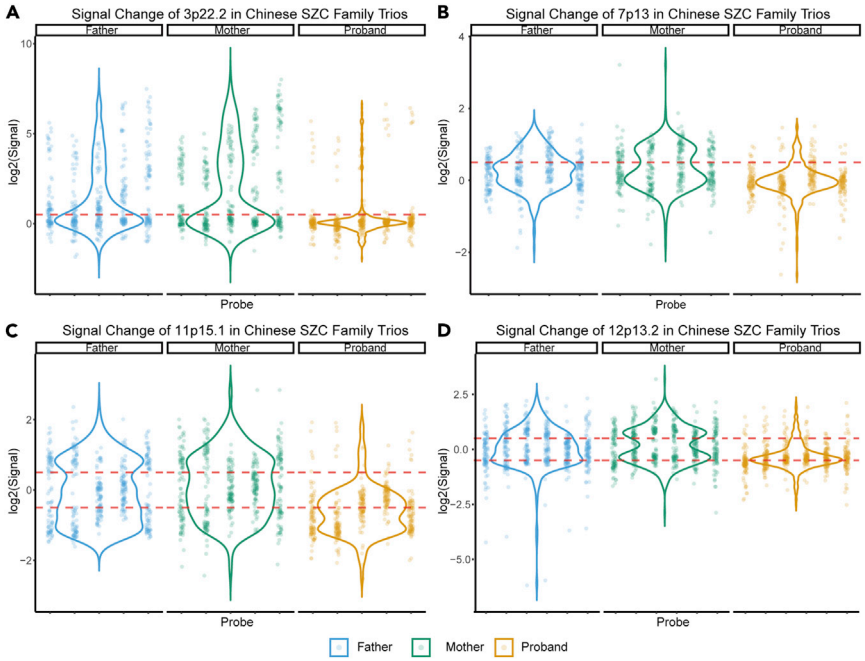


Figure 1. Signal distribution of novel transmission disequilibrium CNVs implicated to SCZ

They are located at 3p22.2 (A), 7p13 (B), 11p15.1(C), and 12p13.2(D) separately. The probe signals have been ADM-2 adjusted and applied log2 transformation. Guidelines indicated log2 signal of ±0.5.

These findings suggest that rare *de novo* duplications may substantially increase the risk for SCZ, while some protective or benign large deletions tend to get lost between generations.

A number of previously identified CNVs for susceptibility to SCZ from case-control studies of various ethnics were validated in our family-based analysis, such as the significantly unbalanced transmitted CNVs 3q29 del (Li et al., 2016; Mulle, 2015) and 5q31.2 dup (Baron, 2001); the enriched *de novo* CNVs 2p16.3 dup (Vrijenhoek et al., 2008) and 12p13.31 dup (Costain et al., 2013); the *de novo* hotspot region 6p21.32 (Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011), *etc.* Besides, several candidate CNVs revealed by this study were in accordance with results from previous GWAS, gene functional studies, and other research studies on biomarkers of SCZ. For example, single nucleotide polymorphisms (SNPs) in the 3p26.1 region were reported to be associated with risk for SCZ in Chinese population in earlier investigations (Walsh et al., 2008; Yu et al., 2017); *FFAR3* at 19q13.12 takes part in the microbiota-gut-brain axis, which regulates the development of SCZ (Clark et al., 2007; Stilling et al., 2016); *MORN1* at 1p36.33 can be alternatively spliced by Gomafu, a long non-coding RNA that is associated with the risk of SCZ (Barry et al., 2014; Ip et al., 2016); and the copy number changes of microRNA-650 at 22q11.22 regulate the production of interleukin 6 (IL-6) (Yun et al., 2015), and the plasma concentrations of IL-6 were significantly elevated in patients with acute-state SCZ compared with healthy controls (Frommberger et al., 1997).

Our transmission disequilibrium and *de novo* CNV analyses successfully revealed a few novel CNVs which may be involved in the development of SCZ, and there are multiple genes located in these CNV regions. Among them, the under-transmission of *CTDSPL* at 3p22.2 is of particular interest. CTDSPL is a carboxy-terminal domain small protein that can be recruited by neuronal genes with RE-1 element and lead to their silencing in non-neuronal cells. The copy number changes of this gene have been reported to be correlated with cognitive and nervous system diseases. One study suggested that the copy number loss of *CTDSPL* was associated with the change in music creativity (Ukkola-Vuoti et al., 2013). Moreover, an 8,510 bp deletion of this gene was associated with a more than two-fold higher risk for a neurological disorder multiple system atrophy in patients than in controls (Hama et al., 2017). Besides, the SNP rs9311168 on the gene body is implicated in structural and functional brain aging (Seshadri et al., 2007). As for SCZ, expression of *CTDSPL* declined in the peripheral blood cells of patients with 22q deletion syndrome (Dantas et al.,

Table 4. High-frequency *de novo* CNVs in Chinese SCZ trios

CNV loci		Position (Mb)	Interval size (kb)	CNV type	Trio counts	Involved genes	Ref
12	p13.31	9.64–9.72	81.52	dup	18		(Costain et al., 2013)
2	p16.3	52.75–52.78	31.74	dup	16		(Vrijenhoek et al., 2008)
19	q13.12	35.85–35.86	9.38	del	15		(Sherwin et al., 2016)
6	p12.1	53.93–53.93	5.59	del	13	*MLIP*	
12	p13.2	11.22–11.25	30.99	dup + del	13	*PRH1-PRR4, PRH1*	(Castellani et al., 2017)
7	q34	141.77–141.79	28.94	dup	12	*MGAM*	(Smith et al., 2010)
17	q21.2	39.42–39.43	7.58	dup	12		
1	p36.33	2.24–2.27	34.44	dup	11	*SKI, MORN1*	
3	q27.3	186.41–186.42	10.18	dup	11		(Bittel et al., 2009)

2019), who often have cognitive deficits and develop SCZ in an estimated 25% of all the cases. Besides, *CTDSPL* co-expressed with many SCZ risk genes, such as *NOTCH2*, *DOCK7*, and *SPON1*, which mainly changed in astrocytes, as shown by PsychENCODE (Gandal et al., 2018; Wang et al., 2018) and our database-based analysis. Furthermore, mir-183-5p, a microRNA that can regulate *CTDSPL*, has also been found to be down-regulated in the blood cell of patients with SCZ (Vachev et al., 2016). All the evidence is consistent with our findings, suggesting that the under-transmission of *CTDSPL* at 3p22.2 from healthy parents to the probands may lead to a decline in *CTDSPL* expression and then play a role in the etiology of SCZ.

There are several lines of evidence supporting the correlations between other newly discovered CNVs and neurodevelopmental disorders. For example, gene *HECW1* at the under-transmitted CNV 7p13 dup encodes an E3 ubiquitin-protein ligase expressed in the brain. The methylation of this gene is slightly higher on the gene body in the striatum of patients with SCZ than in controls (Migdalska-Richards and Mill, 2019; Viana et al., 2017), and it is hypomethylated in patients with autism (Karunakaran et al., 2019; Wang et al., 2014). Gene *MRGPRX1* at the over-transmitted CNV 11p15.1 encodes a G-protein-coupled receptor. This orphan receptor regulates the sensation of pain and is thought to be a marker of autism (Ali, 2017). *MLIP*, which is located in the *de novo* CNV 6p12.1 del, is a transcriptional cofactor that participates in the AKT/mTOR pathway. In a mice model, *MLIP* is co-expression with PDE11A, a gene involved in social deficits of disorders of adulthood (including SCZ) (Hegde et al., 2016). Moreover, the SNP rs9464011 in the intron of *MLIP* was significantly associated with attention-deficit/hyperactivity disorder in a family-based genome-wide quantitative trait loci analysis in a Korean population (Kweon et al., 2018). Our results support the well-known pleiotropy of CNVs and once again demonstrate that some risk loci are shared by different neuropsychiatric disorders.

Moreover, the GO enrichment analysis based on the genes in our susceptible CNVs supports the complex etiology of SCZ, the hypothesis supposes that SCZ is influenced by anatomical abnormality (e.g. epithelial sheet morphogenesis), dysfunctional neurotransmission (e.g. synapse organization), and stress-associated signaling cascades (e.g. stimuli-sensing channels) (Berdenis van Berlekom et al., 2020; Fatani et al., 2017; Maynard et al., 2001), and all these three pathways were enriched in our GO results. Thus, it is hard to simply explain the etiology of SCZ by several genes or CNVs; we can only add more pieces to fill the puzzle.

In conclusion, in this study, we confirmed the CNV burden in SCZ and discovered five new loci and candidate genes (such as *CTDSPL*) for the susceptibility of SCZ by transmission disequilibrium and *de novo* analyses of a large sample of Chinese family trios. These findings shed light on the detailed pathophysiology of SCZ and may promote early diagnosis and personalized therapy for SCZ.

Limitations of the study

This study has several limitations. First, as we used an aCGH assay to detect CNVs, balanced chromosomal abnormalities could not be detected and were not analyzed. Second, as the sample size of our study was relatively small for a complex disease study, validation tests of the novel candidate CNVs

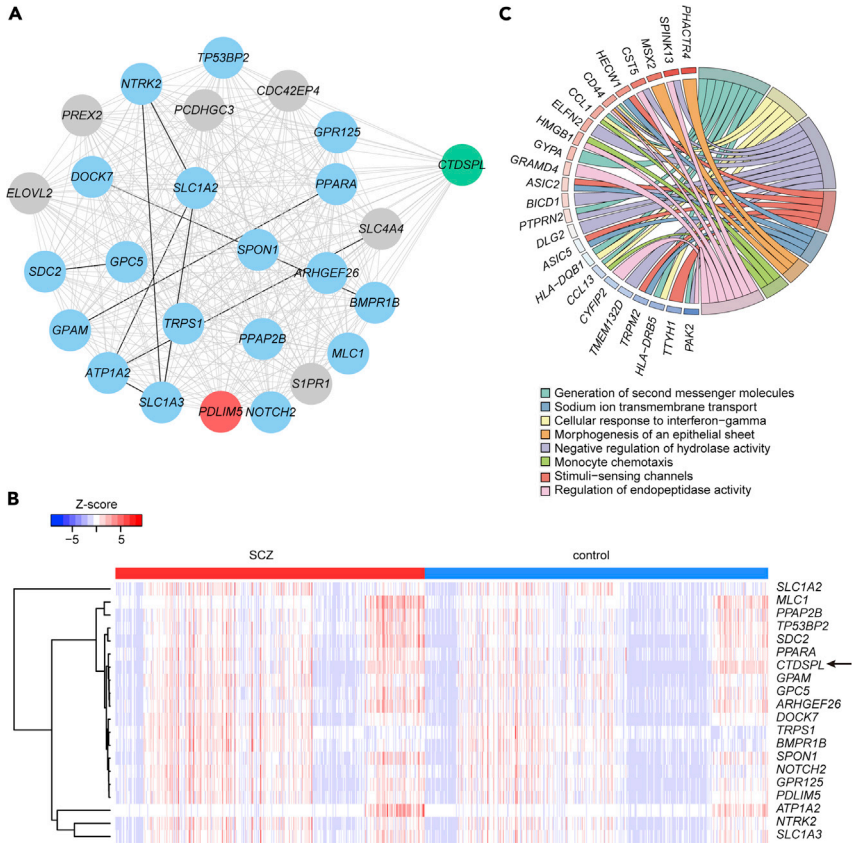


Figure 2. A network of *CTDSPL* and the top 25 nodes (hub genes) in WGCNA module geneM3 (data from PsychENCODE)

(A) The length of each edge is weighted by eigengene connectivity (kME), and edges in gray represent co-expression, and edges in black represent known protein-protein interactions (predicted by STRING). Nodes in blue represent genes differentially expressed in SCZ, and nodes in red represent genes differentially expressed in ASD (false discovery rate <0.05).

(B) Heatmap of *CTDSPL* and SCZ risk gene expression in brain tissues of patients with SCZ (SCZ, n = 454) and in non-neuropsychiatric controls (control, n = 504) (data from PsychENCODE). The expression data underwent *Z* score scaling and hierarchical clustering for each gene (row).

(C) Enrichment heatmap of GO pathways of genes from the large-rare CNVs, transmission disequilibrium CNVs, high-frequency de novo CNVs, and rare de novo CNVs.

are needed by functional studies or association studies in larger population and multiple ethnic populations in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
- METHOD DETAILS
  - Whole-genome array comparative genomic hybridization (aCGH), quality control, and CNV calling

- Statistical analysis of CNVs
- Real-time PCR based CNV copy number detection and calculation
- Pathogenic prediction and gene ontology (GO) annotation

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102894.

ACKNOWLEDGMENTS

We thank Jun Ouyang, Na Zhang, and Qing Zhang for their help with sample management. This work was supported by grants from the National High-tech Research and Development Program of China (grant number 2012AA02A515), National Key Technology R&D Program (grant number 2012BAI01B09), National Natural Science Foundation of China (grant numbers 81773818, 81273596, 30900799, 31701086), National Key Research and Development Program (grant numbers 2017YFC0909303, 2016YFC0905002, 2016YFC1200200), and Shanghai Key Laboratory of Psychotic Disorders (grant number 13dz2260500).

AUTHOR CONTRIBUTIONS

Conceptualization and supervision, S.Q., C.W., L.H., and Q.X.; methodology, C.H., X.W., and L.S.; investigation and validation, X.W., C.H., L.S., M.L., C.Y., and J.Z; formal analysis, C.H., L.S., M.L., W.Z., and L.F.; resources, W.Z. annd L.F.; writing-original draft, C.H. and X.W.; writing-review & editing and visualization, X.W., C.H., L.C., M.L., W.Z., and L.S.; funding acquisition, L.H., S.Q., C.W., and X.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 16, 2021
Revised: June 17, 2021
Accepted: July 19, 2021
Published: August 20, 2021

REFERENCES

Ali, H. (2017). Emerging roles for MAS-related G protein-coupled receptor-X2 in Host Defense peptide, opioid, and neuropeptide-mediated inflammatory reactions. Adv. Immunol. *136*, 123–162. https://doi.org/10.1016/bs.ai.2017.06.002.

Aliyu, M.H., Calkins, M.E., Swanson, C.L., Jr., Lyons, P.D., Savage, R.M., May, R., Wiener, H., McLeod-Bryant, S., Nimgaonkar, V.L., Ragland, J.D., et al. (2006). Project among African-Americans to explore risks for schizophrenia (PAARTNERS): recruitment and assessment methods. Schizophr. Res. *87*, 32–44. https://doi.org/10.1016/j.schres.2006.06.027.

Baron, M. (2001). Genetics of schizophrenia and the new millennium: progress and pitfalls. Am. J. Hum. Genet. *68*, 299–312. https://doi.org/10.1086/318212.

Barry, G., Briggs, J.A., Vanichkina, D.P., Poth, E.M., Beveridge, N.J., Ratnu, V.S., Nayler, S.P., Nones, K., Hu, J., Bredy, T.W., et al. (2014). The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. Mol. Psychiatry *19*, 486–494. https://doi.org/10.1038/mp.2013.45.

Berdenis van Berlekom, A., Muflihah, C.H., Snijders, G., MacGillavry, H.D., Middeldorp, J., Hol, E.M., Kahn, R.S., and de Witte, L.D. (2020). Synapse pathology in schizophrenia: a meta-analysis of postsynaptic elements in postmortem brain studies. Schizophrenia Bull. *46*, 374–386. https://doi.org/10.1093/schbul/sbz060.

Bittel, D.C., Yu, S., Newkirk, H., Kibiryeva, N., Holt, A., 3rd, Butler, M.G., and Cooley, L.D. (2009). Refining the 22q11.2 deletion breakpoints in DiGeorge syndrome by aCGH. Cytogenet. Genome Res. *124*, 113–120. https://doi.org/10.1159/000207515.

Boshes, R.A., Manschreck, T.C., and Konigsberg, W. (2012). Genetics of the schizophrenias: a model accounting for their persistence and myriad phenotypes. Harv. Rev. Psychiatry *20*, 119–129. https://doi.org/10.3109/10673229.2012.694321.

Castellani, C.A., Melka, M.G., Gui, J.L., Gallo, A.J., O'Reilly, R.L., and Singh, S.M. (2017). Post-zygotic genomic changes in glutamate and dopamine pathway genes may explain discordance of monozygotic twins for schizophrenia. Clin. Transl. Med. *6*, 43. https://doi.org/10.1186/s40169-017-0174-1.

Chaignat, E., Yahya-Graison, E.A., Henrichsen, C.N., Chrast, J., Schutz, F., Pradervand, S., and Reymond, A. (2011). Copy number variation modifies expression time courses. Genome Res. *21*, 106–113. https://doi.org/10.1101/gr.112748.110.

Chen, C.H., Shih, H.H., Wang-Wuu, S., Tai, J.J., and Wuu, K.D. (1998). Chromosomal fragile site expression in lymphocytes from patients with schizophrenia. Hum. Genet. *103*, 702–706.

Chen, J., Calhoun, V.D., Perrone-Bizzozero, N.I., Pearlson, G.D., Sui, J., Du, Y., and Liu, J. (2016). A pilot study on commonality and specificity of copy number variants in schizophrenia and bipolar disorder. Transl. Psychiatry *6*, e824. https://doi.org/10.1038/tp.2016.96.

Clark, D., Dedova, I., Cordwell, S., and Matsumoto, I. (2007). Altered proteins of the anterior cingulate cortex white matter proteome in schizophrenia. Proteomics Clin. Appl. *1*, 157–166. https://doi.org/10.1002/prca.200600541.

Costain, G., Lionel, A.C., Merico, D., Forsythe, P., Russell, K., Lowther, C., Yuen, T., Husted, J., Stavropoulos, D.J., Speevak, M., et al. (2013). Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. Hum. Mol. Genet. *22*, 4485–4501. https://doi.org/10.1093/hmg/ddt297.

Dantas, A.G., Santoro, M.L., Nunes, N., de Mello, C.B., Pimenta, L.S.E., Meloni, V.A., Soares, D.C.Q., Belangero, S.N., Carvalheira, G., Kim, C.A., et al. (2019). Downregulation of genes outside the deleted region in individuals with

22q11.2 deletion syndrome. Hum. Genet. *138*, 93–103. https://doi.org/10.1007/s00439-018-01967-6.

Fallin, M.D., Lasseter, V.K., Wolyniec, P.S., McGrath, J.A., Nestadt, G., Valle, D., Liang, K.Y., and Pulver, A.E. (2003). Genomewide linkage scan for schizophrenia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 10q22. Am. J. Hum. Genet. *73*, 601–611. https://doi.org/10.1086/378158.

Fatani, B.Z., Abdullah, R., AldawodFatimah, and Alhawaj, A. (2017). Schizophrenia : etiology, pathophysiology and management : a review. Egypt. J. Hosp. Med. *69*, 2640–2646.

Frommberger, U.H., Bauer, J., Haselbauer, P., Fraulin, A., Riemann, D., and Berger, M. (1997). Interleukin-6-(IL-6) plasma levels in depression and schizophrenia: comparison between the acute state and after remission. Eur. Arch. Psy Clin. Neurosci. *247*, 228–233. https://doi.org/10.1007/Bf02900219.

Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*. https://doi.org/10.1126/science.aat8127.

Guyatt, A.L., Stergiakouli, E., Martin, J., Walters, J., O'Donovan, M., Owen, M., Thapar, A., Kirov, G., Rodriguez, S., Rai, D., et al. (2018). Association of copy number variation across the genome with neuropsychiatric traits in the general population. Am. J. Med. Genet. *177*, 489–502. https://doi.org/10.1002/ajmg.b.32637.

Hama, Y., Katsu, M., Takigawa, I., Yabe, I., Matsushima, M., Takahashi, I., Katayama, T., Utsumi, J., and Sasaki, H. (2017). Genomic copy number variation analysis in multiple system atrophy. Mol. Brain *10*, 54. https://doi.org/10.1186/s13041-017-0335-6.

Hegde, S., Ji, H., Oliver, D., Patel, N.S., Poupore, N., Shtutman, M., and Kelly, M.P. (2016). PDE11A regulates social behaviors and is a key mechanism by which social experience sculpts the brain. Neuroscience *335*, 151–169. https://doi.org/10.1016/j.neuroscience.2016.08.019.

Howes, O.D., and Murray, R.M. (2014). Schizophrenia: an integrated sociodevelopmental-cognitive model. Lancet *383*, 1677–1687. https://doi.org/10.1016/S0140-6736(13)62036-X.

Hubbard, L., Rees, E., Morris, D.W., Lynham, A.J., Richards, A.L., Pardinas, A.F., Legge, S.E., Harold, D., Zammit, S., Corvin, A.C., et al. (2020). Rare copy number variants are associated with poorer cognition in schizophrenia. Biol. Psychiatry. https://doi.org/10.1016/j.biopsych.2020.11.025.

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O.P., Buizer-Voskamp, J.E., Strengman, E., Francks, C., Muglia, P., et al. (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol. Psychiatry *16*, 17–25. https://doi.org/10.1038/mp.2009.101.

Ip, J.Y., Sone, M., Nashiki, C., Pan, Q., Kitaichi, K., Yanaka, K., Abe, T., Takao, K., Miyakawa, T., Blencowe, B.J., et al. (2016). Gomafu lncRNA knockout mice exhibit mild hyperactivity with enhanced responsiveness to the psychostimulant methamphetamine. Sci. Rep. *6*, 27204. https://doi.org/10.1038/srep27204.

Karunakaran, K.B., Chaparala, S., and Ganapathiraju, M.K. (2019). Potentially repurposable drugs for schizophrenia identified from its interactome. Sci Rep. *9*, 1–14.

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., International Schizophrenia Consortium, Wellcome Trust Case Control Consortium, Craddock, N., Owen, M.J., et al. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Human Mol. Genet. *18*, 1497–1503. https://doi.org/10.1093/hmg/ddp043.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153. https://doi.org/10.1038/mp.2011.154.

Kirov, G., Rees, E., Walters, J.T., Escott-Price, V., Georgieva, L., Richards, A.L., Chambert, K.D., Davies, G., Legge, S.E., Moran, J.L., et al. (2014). The penetrance of copy number variations for schizophrenia and developmental delay. Biol. Psychiatry *75*, 378–385. https://doi.org/10.1016/j.biopsych.2013.07.022.

Kweon, K., Shin, E.-S., Park, K.J., Lee, J.-K., Joo, Y., and Kim, H.-W. (2018). Genome-wide analysis reveals four novel loci for attention-deficit hyperactivity disorder in Korean youths. J. Korean Acad. Child Adolesc. Psychiatry *29*, 62–72.

Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., et al. (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am. J. Psychiatry *168*, 302–316. https://doi.org/10.1176/appi.ajp.2010.10060876.

Li, M., Shen, L., Chen, L., Huai, C., Huang, H., Wu, X., Yang, C., Ma, J., Zhou, W., Du, H., et al. (2020). Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. Transl. Psychiatry *10*. https://doi.org/10.1038/s41398-020-0708-y.

Li, Z., Chen, J., Xu, Y., Yi, Q., Ji, W., Wang, P., Shen, J., Song, Z., Wang, M., Yang, P., et al. (2016). Genome-wide analysis of the role of copy number variation in schizophrenia risk in Chinese. Biol. Psychiatry *80*, 331–337. https://doi.org/10.1016/j.biopsych.2015.11.012.

Liu, M., Moon, S., Wang, L., Kim, S., Kim, Y.J., Hwang, M.Y., Kim, Y.J., Elston, R.C., Kim, B.J., and Won, S. (2017). On the association analysis of CNV data: a fast and robust family-based association method. BMC Bioinformatics *18*, 217. https://doi.org/10.1186/s12859-017-1622-z.

Liu, X., Qin, W., He, G., Yang, Y., Chen, Q., Zhou, J., Li, D., Gu, N., Xu, Y., Feng, G., et al. (2005). A family-based association study of the MOG gene with schizophrenia in the Chinese population. Schizophr. Res. *73*, 275–280. https://doi.org/10.1016/j.schres.2004.07.018.

MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992. https://doi.org/10.1093/nar/gkt958.

Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35. https://doi.org/10.1038/ng.3725.

Maynard, T.M., Sikich, L., Lieberman, J.A., and LaMantia, A.S. (2001). Neural development, cell-cell signaling, and the "two-hit" hypothesis of schizophrenia. Schizophr. Bull. *27*, 457–476. https://doi.org/10.1093/oxfordjournals.schbul.a006887.

McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O., et al. (2009). Microduplications of 16p11.2 are associated with schizophrenia. Nat. Genet. *41*, 1223–1227. https://doi.org/10.1038/ng.474.

Migdalska-Richards, A., and Mill, J. (2019). Epigenetic studies of schizophrenia: current status and future directions. Curr. Opin. Behav. Sci. *25*, 102–110.

Mulle, J.G. (2015). The 3q29 deletion confers >40-fold increase in risk for schizophrenia. Mol. Psychiatry *20*, 1028–1029. https://doi.org/10.1038/mp.2015.76.

Ott, J., Kamatani, Y., and Lathrop, M. (2011). Family-based designs for genome-wide association studies. Nat. Rev. Genet. *12*, 465–474. https://doi.org/10.1038/nrg2989.

Pritchard, J.K., and Cox, N.J. (2002). The allelic architecture of human disease genes: common disease-common variant...or not? Hum. Mol. Genet. *11*, 2417–2423.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

Rees, E., and Kirov, G. (2021). Copy number variation and neuropsychiatric illness. Curr. Opin. Genet. Dev. *68*, 57–63. https://doi.org/10.1016/j.gde.2021.02.014.

Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976. https://doi.org/10.1038/ng.940.

Seshadri, S., DeStefano, A.L., Au, R., Massaro, J.M., Beiser, A.S., Kelly-Hayes, M., Kase, C.S., D'Agostino, R.B., Sr, , ., Decarli, C., et al. (2007). Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. BMC Med. Genet. *8*, S15. https://doi.org/10.1186/1471-2350-8-S1-S15.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski,

B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

Sherwin, E., Sandhu, K.V., Dinan, T.G., and Cryan, J.F. (2016). May the force Be with you: the light and Dark sides of the microbiota-gut-brain Axis in neuropsychiatry. CNS Drugs *30*, 1019–1041. https://doi.org/10.1007/s40263-016-0370-3.

Smith, C.L., Bolton, A., and Nguyen, G. (2010). Genomic and epigenomic instability, fragile sites, schizophrenia and autism. Curr. genomics *11*, 447–469. https://doi.org/10.2174/138920210793176001.

Stilling, R.M., van de Wouw, M., Clarke, G., Stanton, C., Dinan, T.G., and Cryan, J.F. (2016). The neuropharmacology of butyrate: the bread and butter of the microbiota-gut-brain axis? Neurochem. Int. *99*, 110–132. https://doi.org/10.1016/j.neuint.2016.06.011.

Szatkiewicz, J.P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J.L., Neale, B.M., Fromer, M., Ruderfer, D., Akterin, S., Bergen, S.E., et al. (2014). Copy number variation in schizophrenia in Sweden. Mol. Psychiatry *19*, 762–773. https://doi.org/10.1038/mp.2014.40.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*. D607–d613. https://doi.org/10.1093/nar/gky1131.

Takumi, T., and Tamada, K. (2018). CNV biology in neurodevelopmental disorders. Curr. Opin. Neurobiol. *48*, 183–192. https://doi.org/10.1016/j.conb.2017.12.004.

Tam, G.W., Redon, R., Carter, N.P., and Grant, S.G. (2009). The role of DNA copy number variation in schizophrenia. Biol. Psychiatry *66*, 1005–1012. https://doi.org/10.1016/j.biopsych.2009.07.027.

The International Schizophrenia Consortium (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature *455*, 237–241. https://doi.org/10.1038/nature07239.

Ukkola-Vuoti, L., Kanduri, C., Oikkonen, J., Buck, G., Blancher, C., Raijas, P., Karma, K., Lahdesmaki, H., and Jarvela, I. (2013). Genome-wide copy number variation analysis in extended families and unrelated individuals characterized for musical aptitude and creativity in music. PLoS One *8*, e56356. https://doi.org/10.1371/journal.pone.0056356.

Vachev, T.I., Popov, N.T., Marchev, D., Ivanov, H., and Stoyanova, V.K. (2016). Characterization of microRNA Signature in peripheral blood of schizophrenia patients using μParaflo miRNA microarray assay. Int. J. Curr. Microbiol. Appl. Sci. *5*, 503–512.

Viana, J., Hannon, E., Dempster, E., Pidsley, R., Macdonald, R., Knox, O., Spiers, H., Troakes, C., Al-Saraj, S., Turecki, G., et al. (2017). Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. Hum. Mol. Genet. *26*, 210–225. https://doi.org/10.1093/hmg/ddw373.

Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Genetic Risk and Outcome in Psychosis (GROUP) Consortium, Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A., et al. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am. J. Hum. Genet. *83*, 504–510. https://doi.org/10.1016/j.ajhg.2008.09.011.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science *320*, 539–543. https://doi.org/10.1126/science.1155174.

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics *31*, 2912–2914. https://doi.org/10.1093/bioinformatics/btv300.

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*. https://doi.org/10.1126/science.aat8464.

Wang, Y., Fang, Y., Zhang, F., Xu, M., Zhang, J., Yan, J., Ju, W., Brown, W.T., and Zhong, N. (2014). Hypermethylation of the enolase gene (ENO2) in autism. Eur. J. Pediatr. *173*, 1233–1244. https://doi.org/10.1007/s00431-014-2311-9.

Wang, Y.C., Chen, J.Y., Chen, M.L., Chen, C.H., Lai, I.C., Chen, T.T., Hong, C.J., Tsai, S.J., and Liou, Y.J. (2008). Neuregulin 3 genetic variations and susceptibility to schizophrenia in a Chinese population. Biol. Psychiatry *64*, 1093–1096. https://doi.org/10.1016/j.biopsych.2008.07.012.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T., et al. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. New Engl. J. Med. *358*, 667–675. https://doi.org/10.1056/NEJMoa075974.

Xi, Z.R., Qin, W., Yang, Y.F., He, G., Gao, S.H., Ren, M.S., Peng, Y.W., Zhang, Z., and He, L. (2004). Transmission disequilibrium analysis of the GSN gene in a cohort of family trios with schizophrenia. Neurosci. Lett. *372*, 200–203. https://doi.org/10.1016/j.neulet.2004.09.041.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nat. Genet. *40*, 880–885. https://doi.org/10.1038/ng.162.

Yu, H., Yan, H., Li, J., Li, Z., Zhang, X., Ma, Y., Mei, L., Liu, C., Cai, L., Wang, Q., et al. (2017). Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. Mol. Psychiatry *22*, 954–960. https://doi.org/10.1038/mp.2016.212.

Yuan, J., Hu, J., Li, Z., Zhang, F., Zhou, D., and Jin, C. (2017). A replication study of schizophrenia-related rare copy number variations in a Han Southern Chinese population. Hereditas *154*, 2. https://doi.org/10.1186/s41065-016-0025-x.

Yun, J.H., Moon, S., Lee, H.S., Hwang, M.Y., Kim, Y.J., Yu, H.Y., Kim, Y., Han, B.G., Kim, B.J., and Kim, J.M. (2015). MicroRNA-650 in a copy number-variable region regulates the production of interleukin 6 in human osteosarcoma cells. Oncol. Lett. *10*, 2603–2609. https://doi.org/10.3892/ol.2015.3581.

Zhang, F., St Clair, D., Liu, X., Sun, X., Sham, P.C., Crombie, C., Ma, X., Wang, Q., Meng, H., Deng, W., et al. (2005). Association analysis of the RGS4 gene in Han Chinese and Scottish populations with schizophrenia. Genes, Brain Behav. *4*, 444–448. https://doi.org/10.1111/j.1601-183X.2005.00167.x.

Zhao, Q., Li, T., Zhao, X., Huang, K., Wang, T., Li, Z., Ji, J., Zeng, Z., Zhang, Z., Li, K., et al. (2013). Rare CNVs and tag SNPs at 15q11.2 are associated with schizophrenia in the Han Chinese population. Schizophr. Bull. *39*, 712–719. https://doi.org/10.1093/schbul/sbr197.

Zhou, Y., Zhou, B., Pache, L., and Chang, M. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.

Zou, F., Li, C., Duan, S., Zheng, Y., Gu, N., Feng, G., Xing, Y., Shi, J., and He, L. (2005). A family-based study of the association between the G72/G30 genes and schizophrenia in the Chinese population. Schizophr. Res. *73*, 257–261. https://doi.org/10.1016/j.schres.2004.01.015.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
QIAamp DNA blood Kits	Qiagen	Cat#51104
SurePrint G3 human CGH microarray Kit, 1 × 1M	Agilent	Cat#G4824A-021529
SureTag Complete DNA labeling Kit	Agilent	Cat#5190-4240
Oligo aCGH/ChIP-on-chip hybridization Kit	Agilent	Cat#5188-5380
Male human reference DNA	Agilent	Cat#5190–4370
Female human reference DNA	Agilent	Cat#5190–3797
SYBR Master Mix	Toyobo	Cat#QPK-201
Deposited data		
aCGH microarray data	This Paper	ArrayExpress:E-MTAB-8075 (https://www.ebi.ac.uk/arrayexpress/)
PsychENCODE, established weighted gene Co-Expression analysis (WGCNA) modules data	(Gandal et al., 2018)	http://resource.psychencode.org
DGV variants	(MacDonald et al., 2014)	http://dgv.tcag.ca/dgv/app/downloads
Oligonucleotides		
Primers for real-time PCR validation of CNV copy numbet, see Table S3	This paper	NA
Software and algorithms		
Feature Extraction software (version 10.7.1.1)	Agilent	https://www.agilent.com/en/product/mirna-microarray-platform/mirna-microarray-software/feature-extraction-software-228496/download-software-feature-extraction-software?productURL=https%3A%2F%2Fwww.agilent.com%2Fen%2Fproduct%2Fmirna-microarray-platform%2Fmirna-microarray-software%2Ffeature-extraction-software-228496
CytoGenomics 5	Agilent	https://www.agilent.com/en/product/cgh-cgh-snp-microarray-platform/cgh-cgh-snp-microarray-software/cytogenomics-software-228500
Nexus copy number 10.0	BioDiscovery	http://pages.biodiscovery.com/cnv-analysis-nexus-copy-number-10.0
PLINK (version 1.9)	(Purcell et al., 2007)	http://pngu.mgh.harvard.edu/purcell/plink/
CopyCaller software (version 2.1)	ThermoFisher	https://www.thermofisher.cn/cn/zh/home/technical-resources/software-downloads/copycaller-software.html
Cytoscape (v3.8.2)	(Shannon et al., 2003)	https://cytoscape.org
STRING v11.0	(Szklarczyk et al., 2019)	https://string-db.org
Metascape	(Zhou et al., 2019)	https://metascape.org/gp/index.html
GOplot package	(Walter et al., 2015)	https://wencke.github.io
PedCNV package	(Liu et al., 2017)	https://github.com/rksyouyou/PedCNV

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to lead contact, Dr. Shengying Qin (chinsir@sjtu.edu.cn)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data for aCGH microarray has been submitted to ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) with accession no. E-MTAB-8075 publicly available as of the date of publication, the accession no have also been referred in KRT.
- This study did not generate program code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

We recruited a total 100 SCZ family trios of Han Chinese ethnicity from Shanghai Mental Health Center. Each trio comprised an offspring affected by SCZ (proband) and both biological parents, 51 of the probands are male and 49 are female. In 71 of the trios, both parents were healthy, without any mental disease. In the other 29 trios, one parent had been diagnosed with SCZ. The average age of parents is 52.8 $\pm$ 9.21, and the average age of probands are 26.8 $\pm$ 9.02.Diagnoses of SCZ in the participants were made by two independent qualified psychiatrists according to *DSM-IV* criteria. Identification of genetic relationship between each parents and children in all family trios were test by kinship analysis based on 18 short tandem repeats loci (including AMEX, D3S1358, D13S317, D7S820, D16S539, Penta E, TPOX, TH01, D2S1338, CSF1PO, D19S433, vWA, D5S818, FGA, D6S1043, D8S1179, D21S11, D18S51) or whole exome sequencing data by our previous study (Li et al., 2020). The DNA sample of 75 controls (without mental disease, 36 males and 38 females) were previously collected and stored in Bio-X Institutes, Shanghai Jiao Tong University, without age records. The detailed information of all human samples was described in Table S1. All procedures of this study were approved by Bio-X Ethical Committee of Human Genetic Resources. All subjects or their legal guardians understood the procedures and provided informed consent prior to participating, according to the Helsinki Declaration of 1975, as revised in 2008. No animals were used in this study.

METHOD DETAILS

Whole-genome array comparative genomic hybridization (aCGH), quality control, and CNV calling

Genomic DNA (gDNA) from the peripheral blood of the subjects were extracted using QIAamp DNA Blood Kits (Qiagen, Valencia, CA, USA) and then performed array comparative genomic hybridization (aCGH) using 1 × 1M SurePrint G3 Human CGH Microarray (Design ID: 021529, hg19; Agilent, Santa Clara, CA, USA) according to the manufacturer's instructions. Agilent sex-matched human DNA was used as a reference. Briefly, we digested 1.5 μg gDNA using *Alu* I and *Rsa* I and labeled them with Cy5 (sample) and Cy3 (reference) by SureTag Complete DNA Labeling Kit (Agilent), respectively. Then the processed samples were hybridized onto microarray slides with Oligo aCGH/ChIP-on-chip Hybridization Kit (Agilent). We acquired images of the slides using an Agilent SureScan Microarray Scanner G2505C and digitized them using Agilent Feature Extraction software (version 10.7.1.1).

Quality control (QC) were performed using CytoGenomics 5 (Agilent). Samples with poor signal-to-noise ratios in the universal chip (derivative log ratio [DLR] >0.35) were excluded from further analyses. From among the valid samples, probes with outlier (OL) flag = 1 were also eliminated.

CNV calling were then performed by two software: CytoGenomics and Nexus Copy Number 10.0 (BioDiscovery, CA, USA). The CytoGenomics call sets were generated using aberration detection method 2 (ADM-2) statistical algorithms at a threshold of 6.0 (Bittel et al., 2009) with a strict aberration filter. GC correction and diploid peak centralization were performed, and replicates were combined to improve the accuracy of signals. CNV aberrations were defined as intervals more than 5,000 bp and greater than three continuous probes with an average $\log_2$ ratio $\geq$0.58 or $\leq$ -1. The Nexus call sets were generated using the FASST2 segmentation algorithm, a kind of hidden markov model (HMM) at the significance threshold of 10^{-7}. CNV aberrations also requiring three or more continuous probes passing the average

log_2 ratio thresholds of ± 0.2. Final CNV calling of each person were the overlapped CNV intervals called by both of these two algorithms.

Statistical analysis of CNVs

Trios with members that did not pass QC were excluded, and the remaining trios were considered valid trios for statistical analysis of CNV. Genome-wide CNV burden analyses was performed with PLINK (Purcell et al., 2007) software to evaluate genetic variation between the SCZ probands and their healthy parents. Rare CNVs with frequencies less than 1% in the Database of Genomic Variants (DGV) were tested in these burden analyses, and large rare CNVs were defined as CNVs larger than 1,000 kb. A total of 10,000 permutations were completed to assess the statistical significance of our results (one-sided tests).

Clusters of overlapping CNVs in different samples were merged into one CNV if they were more than 50% similar to one another (Marshall et al., 2017). We determined whether two CNVs were similar by dividing the length of the overlapped region by the length of either CNV. We calculated the occurrence of CNVs carried by the healthy parents, and high-frequency CNVs were subjected to transmission disequilibrium analysis to test their association with risk for SCZ with a family-based association method using R package *PedCNV* (Liu et al., 2017). All probands and their healthy parents were included in the analyses. The ADM-2 adjusted signal (described previously) of each probe of the high-frequency CNVs was used for association analysis; the first principal component (PC) score method with 3 clusters were set as parameters of the calculation. Sex and SCZ disease state were included as covariates. Analysis of *de novo* CNVs were performed in valid trios, including trios with two healthy parents (unaffected trios) and trios with one healthy parent and one parent affected by SCZ (affected trios). *De novo* CNVs were CNVs the proband had not inherited from either of his or her parents (i.e., different CNVs than any of the parents' CNVs). Enrichment p values were calculated with Fisher's exact test.

Real-time PCR based CNV copy number detection and calculation

gDNA from 24 randomly selected trios were applied to real-time PCR based copy number calling for validate the novel unbalanced transmitted CNVs. We used ViiA 7 Real-Time PCR System (ThermoFisher, CA, USA) and the standard two-steps amplification and melting curve in this system with Tm = 60°C. Each reaction contained 5μL SYBR Master Mix (Toyobo, Japan), 0.2 μM upstream and downstream primers and 0.1–1 ng DNA in a total of 10 μL volume. Genome region *ACTB* was applied as reference of copy number (CN) = 2, and all primer sequences were listed in Table S3. There were three independent replications of amplification for each region in every sample. The copy number were then calculated and predicted by CopyCaller Software (Version 2.1) based on the real-time PCR results, and results were shown as mean $\pm$ s.d..

Pathogenic prediction and gene ontology (GO) annotation

To predict the function of genes in novel found SCZ-related CNVs, we applied all genes to PsychENCODE (http://resource.psychencode.org) to exam their expression and co-expression networks. The analyzed WCGNA results were got from (Gandal et al., 2018). Network of *CTDSPL* and the top 25 nodes (hub genes) in WGCNA were drawn by Cytoscape (v3.8.2) (Shannon et al., 2003), and the length of each edge is weighted by eigengene connectivity (kME). The protein-protein interaction was predicted by STRING (Szklarczyk et al., 2019), and edges between known interacted proteins were shown in black lines.

Raw data of gene expression in brain tissues and population phenotypes were downloaded from PsychENCODE Consortium. There were 454 SCZ samples and 504 control samples. The raw expression data has been transformed to transcripts per million (TPM) and underwent *Z* score scaling and hierarchical clustering for each gene (row). The heatmap were plotted by R package *gplots*.

UCSC genes with the large-rare CNVs, transmission disequilibrium CNVs, high-frequency *de novo* CNVs and rare *de novo* CNVs were selected for GO analysis by online tool Metascape (Zhou et al., 2019) and plotted by R package *GOplot*.

QUANTIFICATION AND STATISTICAL ANALYSIS

The type of statistical test that was performed for each experiment is indicated in the figure or table legends. In burden analyses, p values were calculated by 1-side empirical significance tests after 10,000 permutations by PLINK; in transmission disequilibrium test, *P* were tested by R package *PedCNV*, in *de novo* CNV analyses, the enrichment significance were calculated with Fisher's exact test.