

Identifying Genomic Variations in Monozygotic Twins Discordant for Autism Spectrum Disorder Using Whole-Genome Sequencing

Yan Huang,^{1,4} Yue Zhao,^{2,4} Yue Ren,^{3,4} Ying Yi,² Xiaodan Li,¹ Zhaomin Gao,¹ Xiaolei Zhan,¹ Jia Yu,³ Dong Wang,² Shuang Liang,¹ and Lijie Wu¹

¹Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin 150081, China; ²College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China; ³State Key Laboratory of Medical Molecular Biology, Department of Biochemistry & Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing 100005, China

Autism spectrum disorder (ASD) presents a set of childhood neurodevelopmental disorders with impairments in social communication and restricted, repetitive, and stereotyped patterns of behavior. Here, based on the whole-genome sequencing (WGS) data of three monozygotic twins discordant for ASD, we explored multiple patient-specific genetic variations and prioritized a list of ASD risk genes. Our results identified DVMT (discordant variation in monozygotic twin) observed in at least two twin pairs, including 14,310 SNPs, 2,425 indels, and 16,735 CNVs, referring to a total of 2,174 genes, and 37 of these were covered by all three types of variations. Gene ontology (GO) enrichment analysis of biological processes for 2,174 genes showed that the majority of these genes were related to neurodevelopmental processes. In addition, functional network analysis showed that there was a strong functional relevance between 37 genes covered by all three types of variations. In conclusion, for the first time, we conducted a comprehensive scan of genomic differences between monozygotic twins discordant for ASD, providing researchers with in-depth directions. It may also provide effective strategies for clinical treatment of individuals affected by ASD.

INTRODUCTION

Autism spectrum disorder (ASD) represents a set of childhood neurodevelopmental disorders with impairments in social communication and restricted, repetitive, and stereotyped patterns of behavior.³ It is a lifelong developmental condition with a great degree of heritability and heterogeneity. A recent nationwide population-based study estimated that ASD prevalence was at 2.47% among U.S. children and adolescents in 2014–2016.⁴ Particularly, the prevalence rate for boys is nearly five times higher than that for girls.¹ Although ASD typically occurs in childhood, patients may even need nursing during their whole life, bringing a heavy burden to the whole society. ASD is now considered as one of the world's most urgent public health issues. Since ASD is highly heritable, family studies reconfirm the strong genetic contributions to this disease, declaring it as one of the most heritable neurodevelopmental and neuropsychiatric disorders.^{2,5} Growing evidence shows the importance of complex genetic factors

in ASD (single-nucleotide variants [SNVs], indels [insertions or deletions], and CNVs),^{6,7} and suggests the possible existence of hundreds of autism risk genes.^{8–11} Whole-exome sequencing (WES)^{9,12–14} and whole-genome sequencing (WGS)^{15–17} have been used to detect *de novo* mutations as long as inherited mutations were involved in ASD and other neuropsychiatric disorders, facilitating the discovery of rare pathogenic variations and risk genes. The curation of large genomic datasets has produced a collection of many genes implicated in ASD risk and ASD-linked genes.^{8,10,18} However, even considering all of the data from these complementary approaches, the effect remains incomplete, and up to now, only about 20% of the ASD cases have been explained.^{6,19,20}

Monozygotic or “identical” twins have been widely studied to dissect the relative contributions of genetics and environment to human diseases, and they are widely used in traditional experimental research. While recent high-throughput research only concerned fraternal twins, few studies have empirically tested for genetic differences between monozygotic (MZ) twins discordant for ASDs. MZ twins were thought to originate from a single fertilized zygote and thought to be genetically identical. However, some MZ twins have the same genetic foundation but presented different phenotypes; in this case, MZ twins appear to have discordant phenotypes for ASD. To unearth certain causes of this difference, we conducted a comprehensive scan of genomic differences between MZ twins discordant for ASD based on WGS (Figure 1). In this study, we first collected three pairs of MZ twins discordant for autism and captured all types of genetic variation, including SNVs, small insertions or deletions (indels), and copy number variations (CNVs). Further, for each twin pair, variations

Received 25 May 2018; accepted 21 November 2018;
<https://doi.org/10.1016/j.omtn.2018.11.015>.

⁴These authors contributed equally to this work.

Correspondence: Shuang Liang, Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin 150081, China.

E-mail: liangyouyou2004@163.com

Correspondence: Lijie Wu, Department of Child and Adolescent Health, School of Public Health, Harbin Medical University, Harbin 150081, China.

E-mail: wulijiehyd@126.com



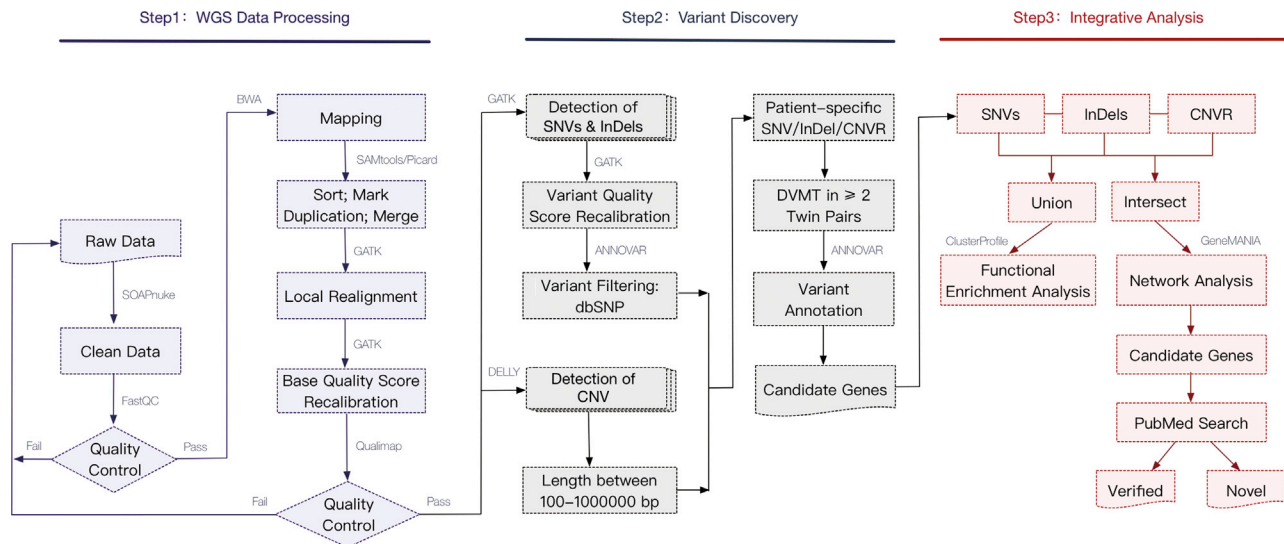


Figure 1. The Bioinformatics Analysis Pipeline

The process includes three steps: the preprocessing of whole-genome sequencing data, the discovery of variants, and the integrative analysis of variant-related genes.

that only presented in individuals with ASD were filtered as discordant variations in MZ twins (DVMTs), and DVMTs in at least two twin pairs were identified for further analysis. At the same time we got a gene set of 2,174 genes involved in all DVMTs, gene ontology (GO) enrichment analysis found many genes were related to the development of autism disease. In addition, functional network analysis showed that there was a strong functional relevance between 37 genes covered by all three types of variations, showing a preference for a diagnosis of autism.

RESULTS

Characteristics of WGS Data

After the filtering of raw data, the basic statistics collection of WGS data for three pairs of MZ discordant twins was concluded (Table S1). In summary, the average number of reads for raw data is 1,151,097,111. After stringent quality control at the sample level, the mean rate of clean data to raw data is 95.11%. Then we computed the genomes from 6 subjects and got an average of 99.53% coverage and 30.85× sequencing depth when mapping the clean data to hg38 with Burrows-Wheeler Aligner (BWA)²¹ (Table 1). Likewise, 99.25% of the genome was covered with at least 4× sequence depth. On average, the mapped reads for each sample were 1,040,389,962, and the mapping rate is 98.30%. The high genome coverage of WGS allows optimal variations calling and potentially captures many clinically relevant mutations.

Identifying DVMTs in Three MZ Discordant Twins

For SNVs and short indels (Figure 2), all SNVs and indels of each individual were first captured by the Genome Analysis Toolkit (GATK).²² At the same time, those variations in Single Nucleotide Polymorphism Database (dbSNP) were filtered out. Second, for each twin pair, variations that only presented in the individuals with ASD were identified,

including 50,644, 42,387, and 49,869 SNVs and 11,118, 9,373, and 10,141 indels. For CNVs (Figure 2), DELLY²³ was performed to identify duplication and deletion, respectively. CNVs with lengths in the range of 100 bp to 1 million bp were retained, including 50,644, 42,387, and 49,869 for duplication and 11,118, 9,373, and 10,141 for deletion for each twin pair. All these variations were observed only in individuals with ASD and were defined as DVMTs.

Further, DVMTs that appeared in at least two twin pairs were identified (Tables S2 and S4). In particular, two CNVs with at least a 50% reciprocal size overlap with each other were the same, and this overlapping region was extracted as a CNVR (copy number variation region). Considering that, 14,310 CNVRs for duplication and 2,425 CNVRs for deletion were extracted (Table S6). Finally, those DVMTs were annotated using the ANNOVAR tool²⁴ (Figures 3A–3D and 4A–4C; Tables S3, S5, and S7). We can see that variations located in the intergenic region account for the biggest proportion, the result of both SNVs and indels, in which non-coding RNA (ncRNA)_intronic has the biggest part of ncRNA. On the other side, no indels located in the exon and 5' UTR compared to SNVs. For the final set of SNVs, we computed the mutation type distribution and found that C > T (G > A) accounted for the largest proportion (Figure 3E). Among those, transition has a higher rate than transversion (Figure 3F). For CNVs, we showed the distribution of duplication and deletion in three twin pairs, respectively (Figure 4B). Finally, to illustrate all kinds of variations, we displayed them in a Circos²⁵ plot (Figure 5).

Prioritization of ASD Risk Genes Based on DVMT

While SNVs and indels usually cover one single gene, CNVs are typically large and may contain multiple genes, preventing the identification of the rightful pathogenic gene. To compile all possible candidate ASD risk genes and evaluate the influence of all variations

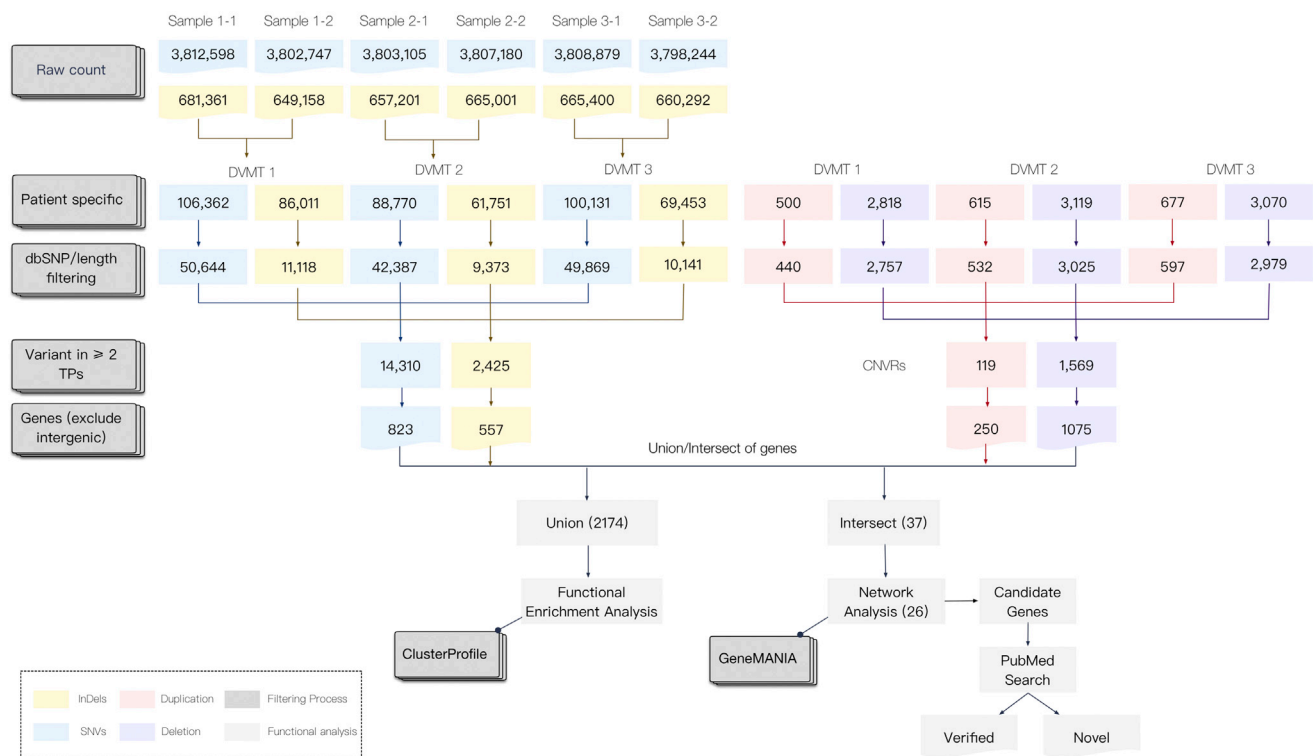
Table 1. The Basic Statistics of Mapping and Quality Control

Twin Pairs	Total Reads	Mapped Reads	Mapping Rate (%)	Average depth (×)	Coverage ≥ 1× (%)	Coverage ≥ 4× (%)
Twin Pair 1						
Twin 1 ^a	1,051,026,934	1,034,220,852	98.40	30.63	99.42	99.15
Twin 2	1,079,153,824	1,062,219,892	98.43	31.34	99.40	99.11
Twin Pair 2						
Twin 1 ^a	1,042,955,285	1,025,555,250	98.33	30.35	99.41	99.12
Twin 2	1,094,044,728	1,070,511,132	97.85	31.87	99.43	99.17
Twin Pair 3						
Twin 1 ^a	1,074,131,972	1,057,584,856	98.46	31.38	99.75	99.48
Twin 2	1,009,031,528	992,247,792	98.34	29.57	99.76	99.49
Mean	1,058,390,712	1,040,389,962	98.30	30.85	99.53	99.25

^aThe twin with autism in the twin pair.

on autism, we focused on a total of overlap 2,174 DVMT-related genes (Table S8), including 823 genes for SNVs, 557 genes for indels, and 1,174 genes for CNVRs (Figure 2). Considering the distribution of genes by ANNOVAR,²⁴ 239 of these genes located on exons. Functional enrichment analysis was performed for these 2,174 candidate genes, showing that the majority of these genes were involved in the regulation of cell morphogenesis, the development of dendrite and synapse, the regulation of synapse, and modulation of chemical synaptic transmission (Figure 6A).

To further explore candidate genes and prioritize important genes, we focused on 37 genes that appeared in all three types of variations (Table S9). We analyzed the functional network of these genes by GeneMANIA.²⁶ Results indicated that 26 of 37 genes were contained in the network (Figure 6B), covering networks of co-expression, co-localization, and shared protein domain and genetic interactions, referring to 97.41%, 1.93%, 0.45%, and 0.21% of all genes, respectively, showing that there is a functional association among them. Referring to the network of 37 genes, several genes are related to

**Figure 2. The Filtration Process of SNVs, Indels, and CNVs**

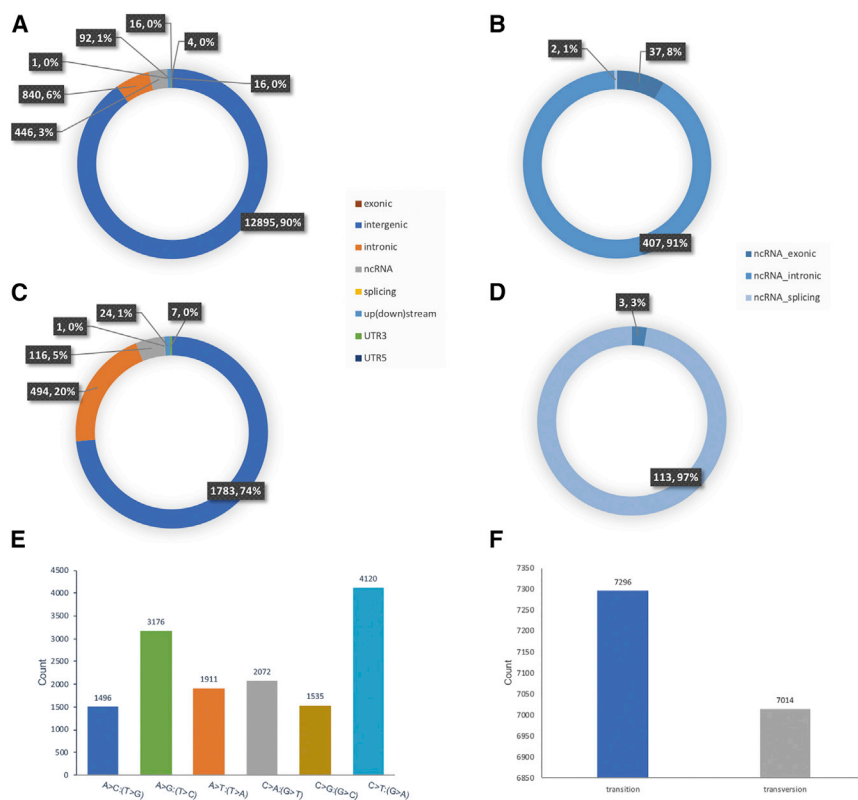


Figure 3. The Detection Results of SNVs and Indels

(A) Annotation result of SNVs. (B) Detailed distribution of ncRNA in SNVs. (C) Annotation result of indels. (D) Detailed distribution of ncRNA in indels. (E) The mutation signature of SNVs. (F) Distribution diagram of transversion and transition.

refine and contribute to the existing genetic architecture of ASD,^{38–40} for the genetic contribution can only explain less than 20% of the patients. In this study, we reported the full detection of genomic differences between MZ twins discordant for autism and aimed to test whether somatic mutations play a role in autism etiology. Three genomic changes dominant in affected individuals were identified by deep sequencing in MZ discordant twin pairs. Our goal was the identification of mutations that were present only in the affected twin, which pointed to novel candidates for autism susceptibility loci.

Careful clinical characterization has revealed that ASD is extremely heterogeneous; variants may be linked to ASDs and other related disabilities, such as intellectual disability, or medical co-morbidities, such as seizures. Further investi-

important functions, which are also core connectors in the network. Among those genes, some are directly associated with ASD, such as Duchenne muscular dystrophy (DMD), a widely studied molecular factor that can be applied in the early diagnosis of autism; it contributes to the organization of autism-associated *trans*-synaptic neurexin-neurologin complexes and to the clustering of synaptic GABA_A receptors.²⁷ Likewise, NRXN3, a type of neurexin gene that is a presynaptic cell-adhesion molecule affecting the function of synapses and mediating the conduction of nerve signals, plays an important role in normal brain development.²⁸ ASMT might be a susceptibility gene for autism as indicated by a large-scale study of Chinese Han individuals, using target sequencing.²⁹ ZNF713 and PTPRN2 are identified as new candidate susceptibility genes.^{30,31} GPC5 suggests a possible role in growth regulation and in cognitive development.^{32,33} Besides, those genes were related to a series of biological processes, such as growth regulation, cognitive development, normal brain development, and conduction of nerve signals. These functions demonstrate potential links in the progression of autism disorder disease.

DISCUSSION

Growing evidence shows the importance of complex genetic factors in ASD (SNVs, indels, and CNVs),^{6,7} suggesting the possible existence of hundreds of autism risk genes.^{8–11} There exists several databases curating general autism risk genes, such as AutDB,³⁵ autismKB,³⁶ and NPdenovo.³⁷ However, more studies are still needed to further

investigation of clinical relevance to these variants is ongoing. To discuss the potential implications of our study, as a more general issue to the use of next-generation sequencing (NGS) for the study of autism and other neuropsychiatric disorders, we integrated evidence from protein-protein interaction (PPI) data curated from the String Database and collected other known neurodevelopmental disease risk genes from public databases (NPdenovo, AutDB, and autismKB) to find highly interactive genes, including a total of 4,434 genes, with an overlapping of 574 genes with our union gene set list (2,174 genes), and they are significantly enriched ($p = 0.014$, hypergeometric test). Based on PPI data, 20 of 37 genes are found to have a PPI with public risk genes. Evidence has shown that genes implicated in ASD also affect several neurodevelopmental disorders by disrupting a wide spectrum of pathways. For example, RBFOX1 plays a crucial role in the regulation of splicing and transcriptional networks in human neuronal development.⁴¹

Genes displayed as core linkers in the network are also closely related to neurodevelopmental diseases (Figure 6B). After a systemic search of papers in PubMed, we found that several genes have been reported to be related to autism. Among those genes, DMD is widely studied in autism disease and shows a preference for a diagnosis of autism; it contributes to the organization of autism-associated *trans*-synaptic neurexin-neurologin complexes and to the clustering of synaptic GABA_A receptors.²⁷ Particularly, males with DMD are at a very high risk of neuropsychiatric disturbance.³⁴ Likewise, NRXN3 is a

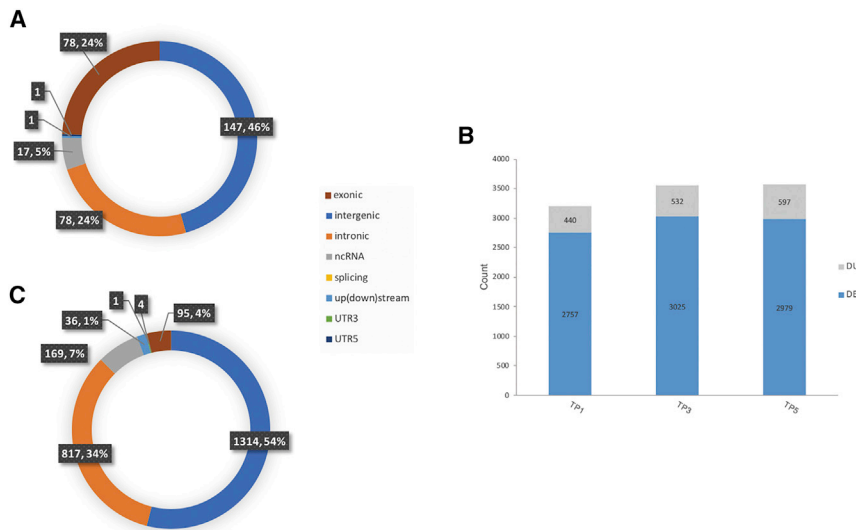


Figure 4. The Detection Results of CNVs

(A) Annotation result of DEL in CNVs. (B) The distribution of DUP and DEL in CNVs for each twin pair. (C) Annotation result of DUP in CNVs. DEL, deletion; DUP, duplication.

type of neurexin gene, and neurexins are presynaptic cell-adhesion molecules that affect the function of synapses and mediate the conduction of nerve signals and that play an important role in normal brain development and become candidate genes for autism.²⁸ Wang et al. found that, among those, ASMT might be a susceptibility gene for autism, in a large-scale study of Chinese Han individuals, using target sequencing.²⁹ ZNF713 and PTPRN2 have been identified as new candidate susceptibility genes.^{30,31} GPC5 and DPP6 suggest a possible role of GPC5 in growth regulation and in cognitive development.^{32,33}

However, we have not yet proven whether these mutations, singly or in combination, contribute to the development of ASD. Given that this is only a pilot study for WGS in ASD and is based on a relatively small number of twins, the replication of our findings is needed to better describe the genetic architecture of ASD in the future. For other mutations, their role in the condition needs to be closely followed in the medical literature. The reported genetic variants also highlighted potential molecular targets for pharmacologic manipulation and open the way for personalized therapeutics in ASD. Promising results, thus far, indicate that WGS could eventually help direct personalized approaches to clinical management in individuals and families affected by ASD. In the future, results of WGS might allow earlier diagnosis of ASD, especially in siblings, for whom recurrence rates are approximately 18%.⁴²

MATERIALS AND METHODS

Sample Selection

Three pairs of ASD-discordant MZ twins were recruited from the Child Development and Behavior Research Center (CDBRC), Harbin Medical University, Heilongjiang Province, China. More than two experienced psychiatrists independently issued ASD diagnoses according to the criteria of the *Diagnostic and Statistical Manual*

of Mental Disorders, 5th edition (DSM-5). All probands were also administered by the Autism Diagnostic Observation Schedule—Generic (ADOS)²⁵ and were all found to meet the criteria for autism. For each patient, we applied the ADOS module according to the level of expressive language. Cases with Rett syndrome, tuberous sclerosis, fragile-X syndrome, and any other neurological conditions suspected to be associated with autism were excluded by clinical examination and molecular genetic tests of the *FMRI* gene.²⁵

Approval from the Ethics Committee of Harbin Medical University (HMUIRB2012008) was obtained before initiating the study.

WGS

WGS was performed with HiSeq 2000 technology (Illumina) at the Beijing Genomics Institute (BGI) in Shenzhen.

Genomic DNA from 3 MZ twin pairs discordant for ASD was isolated from the peripheral blood and used to construct a genomic DNA library for WGS. DNA was extracted and sheared into fragments, which were then purified by gel electrophoresis. DNA fragments were ligated with adaptor oligonucleotides to form paired-end DNA libraries with an insert size of 500 bp. To enrich libraries for sequences with 5' and 3' adaptors, we used ligation-mediated PCR amplification with primers complimentary to the adaptor sequences. The DNA libraries were sequenced to generate 90-bp paired-end reads with at least 30× average genome coverage per sample.

Discovery and Characterization of Various Variants

Raw data were preprocessed by SOAPnuke⁴³ to get clean data, which removed sequences of adaptor, low-quality reads, and unknown reads. Both raw data and clean data were quality controlled by FastQC.⁴⁴ Filtered reads were then mapped to the human reference genome (hg38) with the BWA, v0.7.17.²¹ The BAM file was then sorted by genomic position and indexed using Sequence Alignment/Map tools (SAMtools).⁴⁵ BAM files from the same sample were then merged into a single BAM file using MergeSamFiles in Picard tools. To prevent PCR artifacts from influencing the downstream analysis, we used MarkDuplicates in Picard tools to mark the duplicate reads, which were then ignored. Qualimap⁴⁶ was used to calculate the basic statistics of BAM files.

Local realignment and quality recalibration were performed with the Genome Analysis Toolkit (GATK),v3.8.²² The HaplotypeCaller (HC)

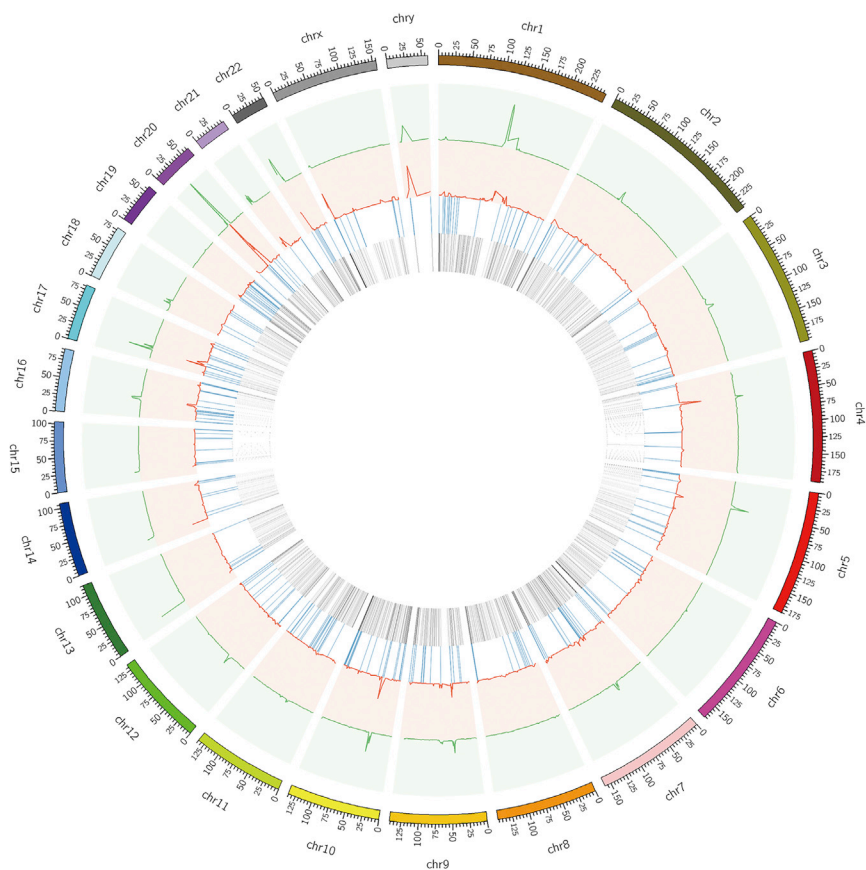


Figure 5. The Display of Circos for All Variations

From the innermost band to the outermost band, CNV (Deletion), CNV (Duplication), the density distribution of indels, and SNVs are indicated.

Variant Filtration and Risk Gene Prioritization

When assessing variants of SNVs and indels, we removed those variants common in the population (in dbSNP142) that were likely to be false positives. For each twin, we also removed variants of the normal sibling from the proband, and then we retained variants found in at least two pairs of twins. GATK SelectVariants was used to get intersect or union set from Variant Call Format (VCF) files. As for CNVs, sizes between 100 bp and 1,000,000 bp were preserved, since it is an interval region of genome; we considered CNVs with at least 50% reciprocal size overlap with each other to be the same and only retained the overlapped region for the following analysis, considered as the CNVR. We used BEDtools⁴⁷ to achieve this goal.

The effects (e.g., missense, nonsense, or frame-shift mutations) and classifications (e.g., in exonic, intronic, or intergenic regions) of variants across the genome were annotated by ANNOVAR.²⁴ Finally, we got a gene set annotated from each type of variation and computed the union or intersect set of those genes.

was used to detect SNVs and small indels (< 65 nt). We then used the GATK Variant Quality Score Recalibration (VQSR) to filter spurious SNVs and indels due to sequencing errors and mapping artifacts. We performed CNV analysis with DELLY.²³ CNV calls with a p value < 0.05 and calls with >50% of q0 (zero mapping quality) reads within the CNV regions were removed (q0 filter). All variations are displayed in a Circos plot.²⁵

annotated from each type of variation and computed the union or intersect set of those genes.

Functional Analysis

For gene union set from three types of variation, we conducted a functional enrichment analysis of biological process by clusterProfiler.⁴⁸ For gene intersect set from three types of

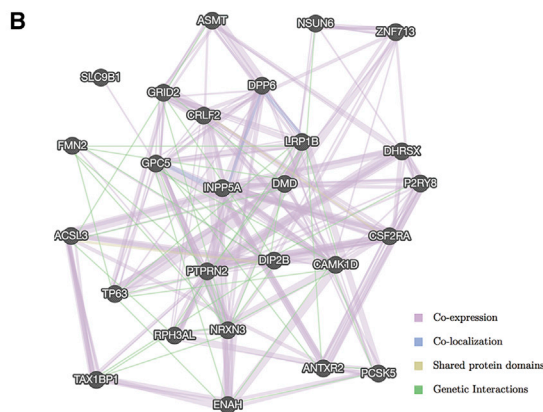
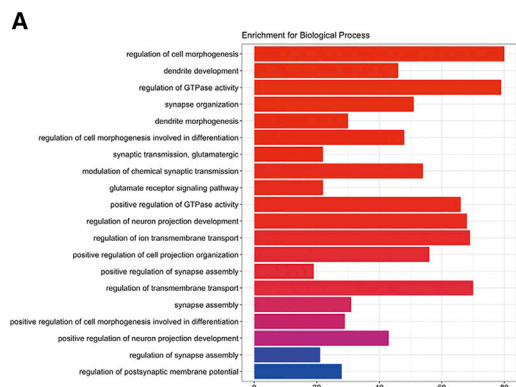


Figure 6. The Functional Analysis of Genes

(A) Enrichment of the biological process for 2,174 genes. (B) Network analysis of 37 genes by GeneMANIA.

variation, we explored the functional network of these genes by GeneMANIA.²⁶

DATA AND SOFTWARE AVAILABILITY

The sequencing data reported in this paper have been deposited at the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>; SRA: PRJNA471799).

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine tables and can be found with this article online at <https://doi.org/10.1016/j.omtn.2018.11.015>.

AUTHOR CONTRIBUTIONS

Y.H., Y.Z., Y.R., Y.Y., and J.Y. performed the analysis of sequencing data. Y.H., X.L., Z.G., and X.Z. collected the samples and performed the clinical examination and molecular genetic tests. Y.H., Y.Z., and Y.Y. wrote the initial manuscript. S.L. and D.W. improved the thesis of the section of the paper. L.W. designed the experiment. All authors reviewed the manuscript and provided comments that were incorporated in the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

We thank all the patients and their parents for their support and participation. This study was funded by grants from the National Natural Science Foundation of China (81202221 and 81273094), the China Postdoctoral Science Foundation (2014M561375), the Scientific Research Fund of Heilongjiang Provincial Education Department (12521232), and the Scientific Research Fund of Heilongjiang Postdoctoral Program (LBH-Z14153).

REFERENCES

1. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators; Centers for Disease Control and Prevention (2012). Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill. Summ.* 61, 1–19.
2. Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., and Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* 25, 63–77.
3. de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* 22, 345–361.
4. Xu, G., Strathearn, L., Liu, B., and Bao, W. (2018). Prevalence of autism spectrum disorder among US children and adolescents, 2014–2016. *JAMA* 319, 81–82.
5. Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., et al. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* 68, 1095–1102.
6. Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* 22, 229–237.
7. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* 86, 749–764.
8. Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* 1380, 42–77.
9. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242–245.
10. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368–372.
11. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863–885.
12. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241.
13. O’Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246–250.
14. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285–299.
15. Yuen, R.K.C., Merico, D., Bookman, M., Howe, J.L., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20, 602–611.
16. Jiang, Y.H., Yuen, R.K.C., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., et al. (2013). Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* 93, 249–263.
17. Yuen, R.K.C., Thiruvahindrapuram, B., Merico, D., Walker, S., Tammimies, K., Hoang, N., Chrysler, C., Nalpathamkalam, T., Pellicchia, G., Liu, Y., et al. (2015). Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* 21, 185–191.
18. Basu, S.N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* 37, D832–D836.
19. Christian, S.L., Brune, C.W., Sudi, J., Kumar, R.A., Liu, S., Karamohamed, S., Badner, J.A., Matsui, S., Conroy, J., McQuaid, D., et al. (2008). Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol. Psychiatry* 63, 1111–1117.
20. Ma, D., Salyakina, D., Jaworski, J.M., Konidari, I., Whitehead, P.L., Andersen, A.N., Hoffman, J.D., Slifer, S.H., Hedges, D.J., Cukier, H.N., et al. (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.* 73, 263–273.
21. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
22. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
23. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
24. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.
25. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Jr., Leventhal, B.L., DiLavore, P.C., Pickles, A., and Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30, 205–223.
26. Zuberi, K., Franz, M., Rodriguez, H., Montojo, J., Lopes, C.T., Bader, G.D., and Morris, Q. (2013). GeneMANIA prediction server 2013 update. *Nucleic Acids Res.* 41, W115–W122.

27. Miranda, R., Nagapin, F., Bozon, B., Laroche, S., Aubin, T., and Vaillend, C. (2015). Altered social behavior and ultrasonic communication in the dystrophin-deficient mdx mouse model of Duchenne muscular dystrophy. *Mol. Autism* 6, 60.
28. Wang, J., Gong, J., Li, L., Chen, Y., Liu, L., Gu, H., Luo, X., Hou, F., Zhang, J., and Song, R. (2018). Neurexin gene family variants as risk factors for autism spectrum disorder. *Autism Res.* 11, 37–43.
29. Wang, L., Li, J., Ruan, Y., Lu, T., Liu, C., Jia, M., Yue, W., Liu, J., Bourgeron, T., and Zhang, D. (2013). Sequencing ASMT identifies rare mutations in Chinese Han patients with autism. *PLoS One* 8, e53727.
30. Lionel, A.C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A.R., Howe, J.L., Wang, Z., et al. (2011). Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci. Transl. Med.* 3, 95ra75.
31. Metsu, S., Rainger, J.K., Debacker, K., Bernhard, B., Rooms, L., Grafodatskaya, D., Weksberg, R., Fombonne, E., Taylor, M.S., Scherer, S.W., et al. (2014). A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families. *Hum. Mutat.* 35, 1295–1300.
32. Kannu, P., Campos-Xavier, A.B., Hull, D., Martinet, D., Ballhausen, D., and Bonafé, L. (2013). Post-axial polydactyly type A2, overgrowth and autistic traits associated with a chromosome 13q31.3 microduplication encompassing miR-17-92 and GPC5. *Eur. J. Med. Genet.* 56, 452–457.
33. Maussion, G., Cruceanu, C., Rosenfeld, J.A., Bell, S.C., Jollant, F., Szatkiewicz, J., Collins, R.L., Hanscom, C., Kolobova, I., de Champfleury, N.M., et al. (2017). Implication of LRRC4C and DPP6 in neurodevelopmental disorders. *Am. J. Med. Genet. A.* 173, 395–406.
34. Ricotti, V., Mandy, W.P.L., Scoto, M., Pane, M., Deconinck, N., Messina, S., Mercuri, E., Skuse, D.H., and Muntoni, F. (2016). Neurodevelopmental, emotional, and behavioural problems in Duchenne muscular dystrophy in relation to underlying dystrophin gene mutations. *Dev. Med. Child Neurol.* 58, 77–84.
35. Peraanu, W., Larsen, E.C., Das, I., Estévez, M.A., Sarkar, A.A., Spring-Pearson, S., Kollu, R., Basu, S.N., and Banerjee-Basu, S. (2018). AutDB: a platform to decode the genetic architecture of autism. *Nucleic Acids Res.* 46 (D1), D1049–D1054.
36. Xu, L.-M., Li, J.-R., Huang, Y., Zhao, M., Tang, X., and Wei, L. (2012). AutismKB: an evidence-based knowledgebase of autism genetics. *Nucleic Acids Res.* 40, D1016–D1022.
37. Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., Li, X., Shao, Q., Ran, X., Li, Z., et al. (2016). Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol. Psychiatry* 21, 290–297.
38. Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., Hu, Y., Xu, L., Li, E., and Wang, D. (2018). MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res.* 46 (D1), D371–D374.
39. Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., Liu, L., Hou, P., Cui, T., Tan, P., et al. (2017). RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.* 45 (D1), D115–D118.
40. Zhang, T., Tan, P., Wang, L., Jin, N., Li, Y., Zhang, L., Yang, H., Hu, Z., Zhang, L., Hu, C., et al. (2017). RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.* 45 (D1), D135–D138.
41. Lee, H., Deignan, J.L., Dorrani, N., Strom, S.P., Kantarci, S., Quintero-Rivera, F., Das, K., Toy, T., Harry, B., Yourshaw, M., et al. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 312, 1880–1887.
42. Ozonoff, S., Young, G.S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L.J., Constantino, J.N., Dobkins, K., et al. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 128, e488–e495.
43. Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7, 1–6.
44. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
46. Okonechnikov, K., Conesa, A., and Garcia-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294.
47. Quinlan, A.R. (2014). BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* 47, 11.12.1–11.12.34.
48. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.