Supplemental Materials

Supplemental Figures and Figure Legends

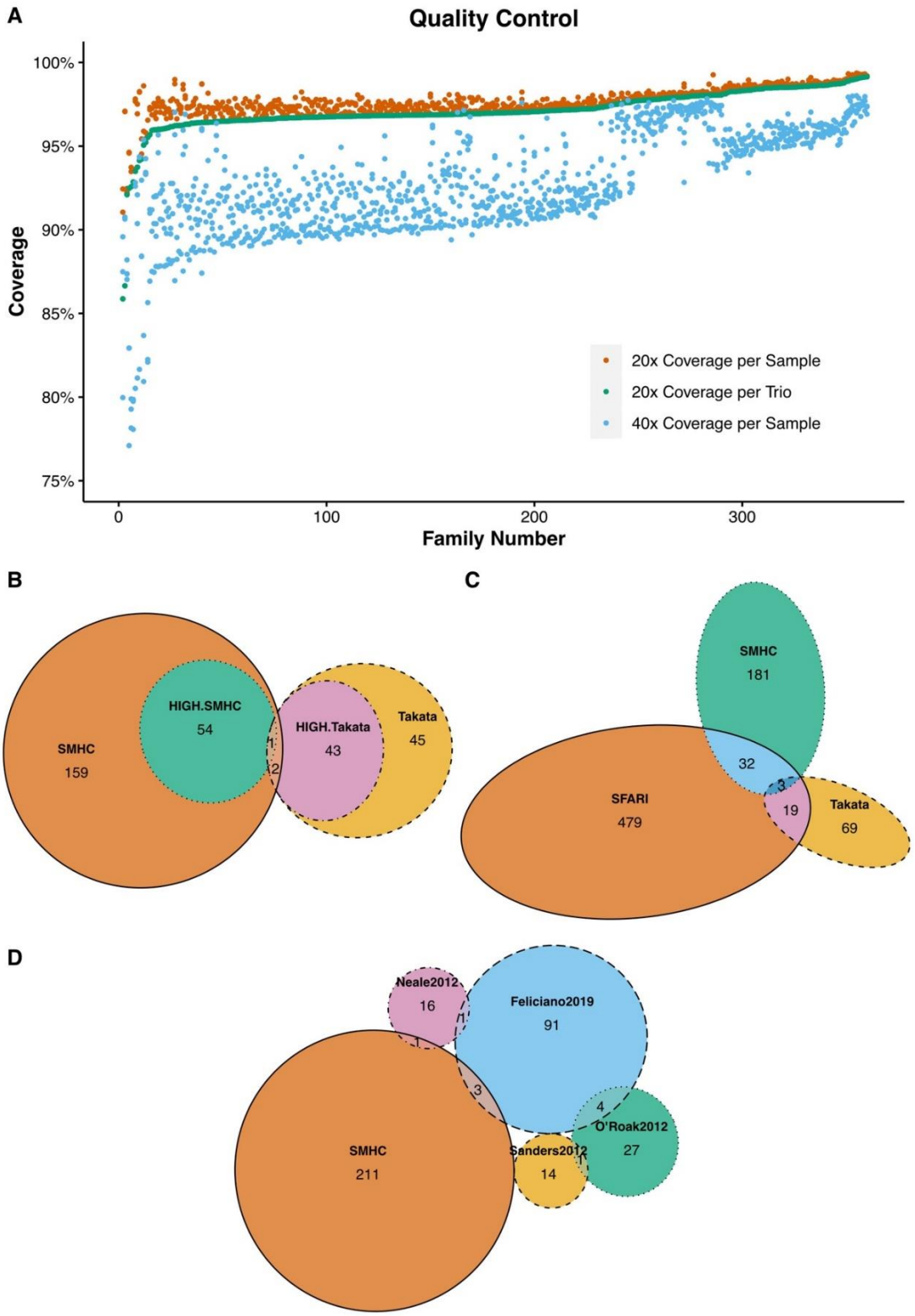

Fig. S1 Quality control results and enrichment analyses of various functional types of *de novo* mutations in our and published ASD cohorts. **A** Sequencing coverage performance. The proportions of the target exome region covered with the indicated numbers (≥20× or 40×) of reads. Data are sorted in the order of the 20× trio rank (trios with the smallest proportion covered with ≥20× reads on the left and the largest on the right) (orange, 20× individual coverage; blue, 40× individual coverage). **B** The overlap of ASD risk genes between the Chinese ASD cohort (SMHC) and the Japanese ASD cohort (Takata) (dark yellow circle, genes with DNMs in the Chinese ASD cohort; light yellow circle, genes with DNM mutations confirmed by Sanger sequencing in the Japanese ASD cohort; green and magenta circles, genes with HIGH-impact DNMs in Chinese and Japanese ASD cohorts, respectively). **C** The overlap of ASD risk genes between the Chinese ASD cohort (SMHC), the the Japanese ASD cohort (Takata), and the SFARI gene list (including Cat S, 1, 2) (green and yellow circles, genes with DNMs in Chinese and Japanese ASD cohorts). **D** The overlap of ASD risk genes between the Chinese ASD cohort (SMHC) with four other studies (samples sizes from 200 to 400 ASD trios). All genes with DNMs were analyzed.

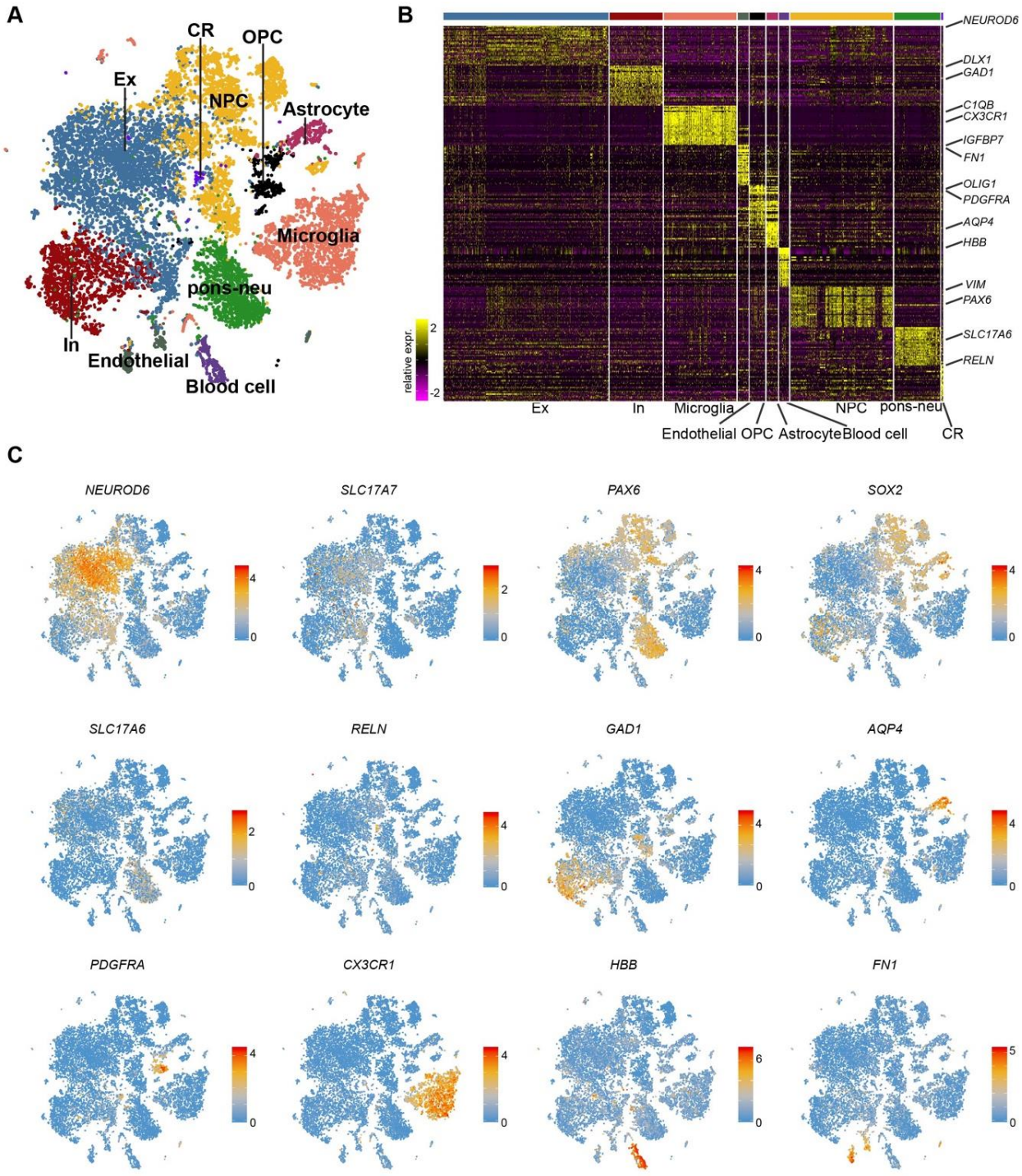

Fig. S2 Overview of transcriptional heterogeneity among cells from diverse brain regions. **A** Visualization of major cell types by UMAP. **B** Heatmap illustrating the differently-expressed genes among major cell types. Well-known marker genes for each cell type are labeled on the left (high, yellow; low, purple). **C** UAMP showing the expression of marker genes. *NEUROD6* and *SLC17A7* for excitatory neurons; *PAX6* and *SOX2* for NPCs; *SLC17A6* for neurons from the pons; *RELN* for Cajal-

Retzius cells; *GAD1* for GABAergic interneurons; *AQP4* for astrocytes; *PDGFRA* for OPCs; *CX3CR1* for microglia; *HBB* for blood cells, and *FN1* for endothelial cells.

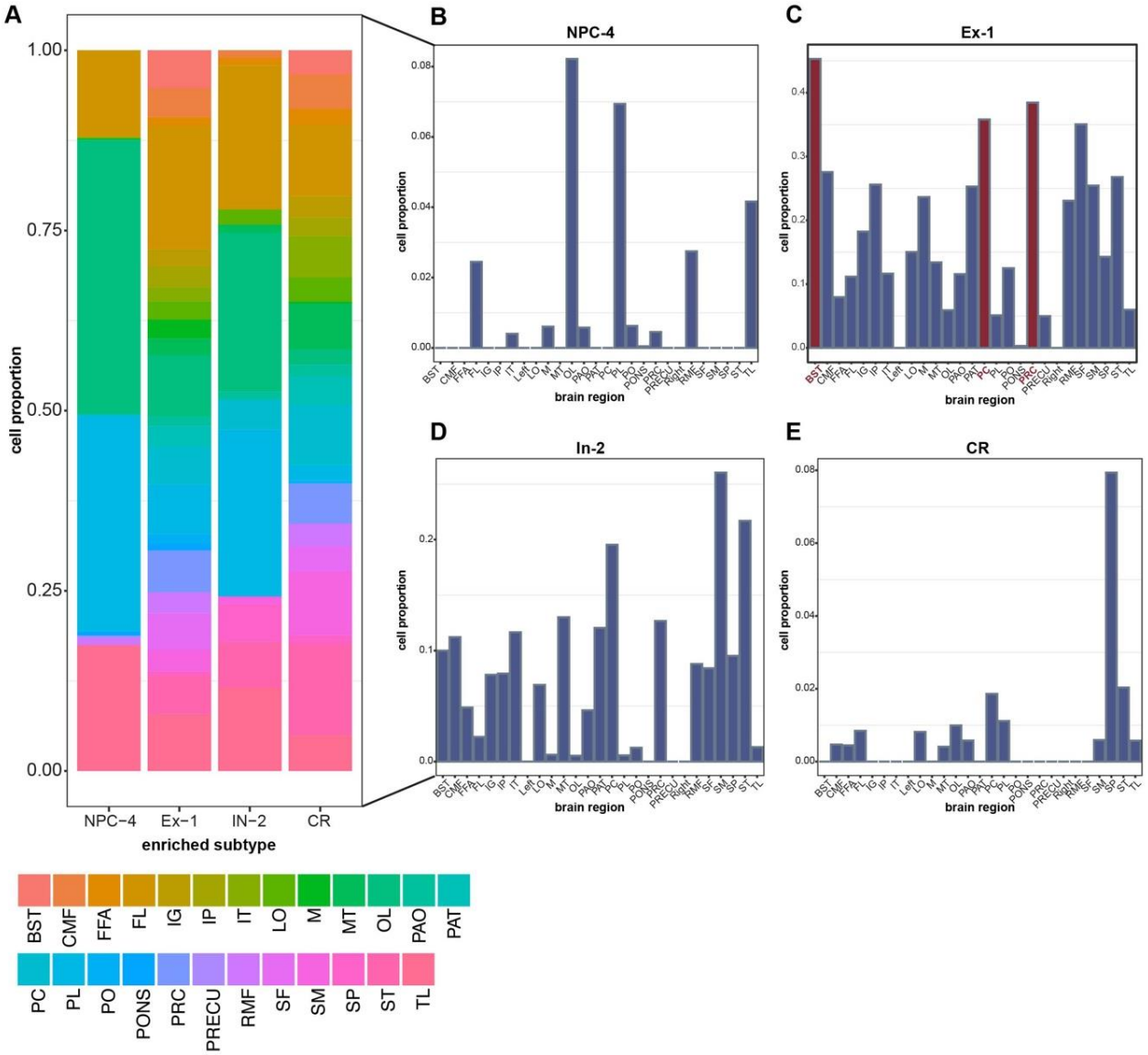

Fig. S3 Regional distribution of cells from ASD risk gene-enriched subtypes. **A** Barplot displaying the regional composition of 4 ASD risk gene-enriched subtypes. **B–D** Histograms showing the proportion of cells for subtypes NPC-4 (**B**), Ex-1 (**C**), In-2 (**D**), and CR (**E**) from each brain region.

Supplemental Tables

Table S1 All genes with *de novo* mutations found in the Chinese ASD cohort (see Excel version).

Table S2 TADA-Denovo analysis of genes with *de novo* mutations (see Excel version).

Table S3 List of chromosomal segments with *de novo* copy number variations which includes genes present in the SFARI gene list (see Excel version).

Table S4 Demographic information of ABIDE-I participants.

		VBM	resting-state fMRI
ASD	*n*	330	265
	Sex (M)	286	232
	Age (Mean±SD)	17.77 ± 8.7	17.81 ± 8.1
	Medication	43	58
	FIQ	105.95	107.74
Healthy Control	*n*	437	379
	Sex (M)	351	300
	Age	17.15 ± 7.2	17.46 ± 7.67
	Medication (Other illness)	2	1
	FIQ	110.96	111.41

Table S5 Seed-based functional connectivity (left hemisphere).

The table lists brain regions that show significant FC differences between ASD and healthy control groups, where the TFCE value represents the T value adjusted by the non-threshold enhancement technique, and the Corrected *P* value represents the *P* value adjusted by FWE correction. Cluster size

indicates the size of the cluster. Only clusters significantly in the permutation test and bigger than 20 voxels (3×3×3×20 = 540 mm^3) are shown.

Left BST							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Cuneus/Precuneus	-6	-78	30	-4.79	609.549	0.0048	11151
Superior Temporal/Transverse Temporal (R)	60	-24	12	-5.50	567.863	0.008	2349
Superior Temporal/Heschl (R)	57	-6	3	-4.27	461.693	0.0206	297
Precuneus (R)	24	-81	48	-4.23	480.983	0.016	999
Left PC							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Fusiform (L)	-45	-48	-21	-4.80	535.973	0.0108	2889
Fusiform (R)	45	-51	-24	-4.48	484.363	0.017	1026
Precuneus/Cuneus/Lingual (L)	-15	-72	21	-5.37	780.266	0.0006	26649
Middle Temporal (L)	-48	-48	9	-4.64	502.022	0.0144	1647
Insula (R)	36	-21	15	-6.18	688.107	0.0022	7074
Inferior Parietal (L)	-42	-33	24	-4.79	522.421	0.0126	2106
Postcentral	3	-36	57	-6.38	1359.3	0.0002	88614
Left PRC							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Precuneus (R)	18	-63	15	-5.38	582.154	0.0072	10125
Insula/Heschl (R)	33	-21	15	-5.16	607.654	0.0058	5238
Middle Temporal (L)	-63	-48	12	-4.33	477.922	0.0194	1647
Middle Temporal (R)	48	-39	0	-4.74	485.143	0.0188	459
Inferior Parietal (L)	-39	-39	21	-4.11	476.775	0.0194	1134
Precentral/Postcentral	6	-30	57	-5.41	834.612	0.0004	60156

Table S6 Seed-based functional connectivity (right hemisphere).

The table lists brain regions that show significant FC differences between ASD and healthy control groups, where the TFCE value represents the T value adjusted by the non-threshold enhancement technique, and the Corrected *P* value represents the *P* value adjusted by FWE correction. Cluster size indicates the size of the cluster. Only clusters significant in the permutation test and bigger than 20 voxels (3×3×3×20 = 540 mm^3) are shown.

Right BST							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Cuneus (L)	-9	-78	27	-5.75	1011.66	0.0002	116289
Superior Temporal/Insula (R)	63	-24	9	-6.05	953.017	0.0002	32616
Superior Frontal (L)	-18	54	-15	-3.73	443.367	0.0202	864
Superior Frontal (L)	-21	63	0	-3.42	433.862	0.0228	459
Superior Frontal (R)	27	57	12	-3.72	441.42	0.0202	486
Middle Cingulate	-6	-21	42	-3.45	450.212	0.0188	1080
Superior Parietal (R)	33	-51	54	-4.06	482.928	0.0128	2970
Middle Cingulate/Paracentral (R)	6	-33	45	-4.19	462.463	0.0154	2565
Precentral (L)	-45	-12	57	-4.61	516.693	0.0082	1944
Superior Frontal (R)	27	-3	69	-4.52	520.732	0.0082	3753
Postcentral (L)	-21	-27	72	-4.17	465.641	0.0152	918
Superior Parietal (L)	-30	-48	72	-3.79	442.636	0.0202	297
Right PC							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Paracentral/Postcentral (R)	12	-39	63	-5.17	763.177	0.001	120771

Lingual (R)	12	-66	-3	-3.29	456.093	0.0218	513
Superior/Middle Temporal Lobe (R)	48	-36	3	-4.69	512.012	0.0124	1161
Right PRC							
Area	MNI coordinates			T value	TFCE value	Corrected *P* value	Cluster size (mm^3)
	x	y	z				
Precuneus/Cuneus (L)	-12	-72	21	-5.93	700.318	0.0022	9774
Precentral (R)	48	-3	30	-4.86	504.085	0.0144	2484
Postcentral (L)	-42	-21	39	-5.43	651.465	0.0042	30186

Table S7 List of abbreviation for brain regions.

Yuan et al. Table S7

Nr.	brain region	abbr.
1	rostral-middle-frontal	RMF
2	pars orbitalis	PAO
3	pars triangulars	PAT
4	superior frontal	SF
5	caudal-middle-frontal	CMF
6	pars opercularis	PO
7	pre-central	PRC
8	frontal lobe	FL
9	occipital lobe	OL
10	parietal lobe	PL
11	temporal lobe	TL
12	post-central	PC
13	supra-maginal	SM
14	superior parietal	SP
15	inferior parietal	IP
16	precuneus	PRECU
17	inferior temporal	IT
18	middle temporal	MT
19	superior temporal	ST
20	bank superior temporal	BST
21	fusiform	FFA
22	lateral occipital	LO
23	insular gyrus	IG
24	medulla	M
25	pons	pons
26	right hemisphere	right
27	left hemisphere	left

Supplemental Materials and Methods

CNV Detection

CNVs were called with GATK PreprocessIntervals, CollectReadCounts, AnnotateIntervals, FilterIntervals, DetermineGermlineContigPloidy, GermlineCNVCaller, IntervalListTools, and PostprocessGermlineCNVCalls based on a cohort mode pipeline detecting germline copy number variants. All CNVs were annotated to GRCh38/hg38 by VEP and AnnotSV [1].

Real-time Quantitative PCR Validation

To confirm the *de novo* CNVs detected by WES, we applied quantitative PCR (qPCR) using DNA from probands, their parents, and controls. We used the comparative CT method (delta-delta CT method) for relative quantification, with data normalized against an endogenous control sequence (glyceraldehyde 3-phosphate dehydrogenase, GAPDH) with two normal copies. Genomic DNA was amplified using SYBR Green (Thermo Fisher Scientific) and qPCR was performed on a StepOnePlus™ Real-Time PCR System (Applied Biosystems). Data were analyzed using R version 3.6.3 and the PCR package [2]. The primers used for qPCR were *TBR1* Forward GGGATGACGAATCAGTCAGA, *TBR1* Reverse TGGCTGGACTGAGAGAGGAG, *RAI1* Forward TCTCCAGGCCAGAAAGAAAA, *RAI1* Reverse TGAATGCCTGGAATGAATGA, *SHANK3* Forward TGCCTCACGGAGTTTTCTCT, *SHANK3* Reverse ATGCGGGACTTTATGCAAAC, *MECP2* Forward CACGGAAGCTTAAGCAAAGG, *MECP2* Reverse TCAAGCACACCTGGTCTCAG, *GAPDH* Forward ATCAAGAAGGTGGTGAAGCA, *GAPDH* Reverse TGACAAAGTGGTCGTTGAGG.

Statistical Analyses

We statistically evaluated the number of dDNMs in each gene using TADA-Denovo [3]. We included

IMPACT HIGH and Possible-damaging missense mutations in the TADA-Denovo analysis. Parameters for this analysis were determined by following the TADA manual. Per-gene mutation rates for LOF and Possible-damaging missense DNMs were obtained from mirDNMR based on sequence context [4].

Single-cell RNA-seq Data Processing

To infer the gene expression pattern of ASD risk genes in the human brain, two publicly available scRNA-seq datasets were downloaded [5, 6] and assembled. The assembled gene-by-cell count matrix was then imported into the R package Seurat [7] for downstream analysis. Briefly, the assembled count matrix was first loaded into the function CreateSeuratObject to create a Seurat object followed by a Log-normalization process using the function NormalizeData. Variable genes were then identified with the function FindVariableFeatures. Next, principal component analysis was applied by the function RunPCA. Unbiased clustering was done with the functions FindNeighbors and FindClusters followed by Uniform manifold approximation and projection (UMAP) dimension reduction analysis with the function RunUMAP.

Identification of Differentially Expressed Genes

Differentially expressed genes were computed with the Seurat function FindAllMarkers by setting the parameter only.pos = TRUE. Genes with adjusted *P*-values <0.05 were selected as differentially expressed genes.

Enrichment Analysis of ASD Risk Genes

To explore the expression pattern of ASD risk genes, gene-set enrichment was analyzed using R package AUCell [8] independently for the 55 High-impact genes and the 165 Moderate-impact genes on the assembled scRNA-seq dataset. The Z-score of the gene-set enrichment score was then computed

with the R function scale and visualized by UMAP.

Ranking of Cells by Enrichment Score of ASD Risk Genes

To assess the expression pattern of ASD risk genes among different cell types and brain regions, we computed the averaged gene-set enrichment scores for each cell type and brain region and then ranked the cell types and brain regions in descending order. The averaged gene-set enrichment scores and rankings were visualized with Heatmap.

The Proportion of Highly Expressed Genes Among Subtypes and Brain Regions

To calculate the proportion of highly expressed genes in each of the two gene sets (the 55 High-impact genes and the 165 Moderate-impact genes), for each gene set, we counted the number of genes that were expressed in at least 25% of the cells from each subtype or brain region. The proportion of highly expressed genes was then computed by dividing the number of highly expressed genes in each subtype or brain region by the total number of genes in each gene set.

Acquisition of Brain Imaging Data

Imaging data were obtained from the Autism Brain Imaging Data Exchange (ABIDE-I, http://fcon_1000.projects.nitrc.org/indi/abide/) [9], a publicly available database released in 2012. 1112 subjects (539 ASDs, 573 age-matched healthy controls) from 16 international imaging sites underwent anatomical and resting-state functional MRI scans, and a series of out-of-scanner phenotypic information was also collected. The scanning parameters of each site can be found online.

Preprocessing of Imaging Data

All the preprocessed BOLD time series data and voxel-based morphometry (VBM) data used in this study were available on the Preprocessed Connectomes Project (PCP; http://preprocessed-

connectomes-project.org/). In this paper, resting-fMRI data and T1-weight MRI data from ABIDE were preprocessed by the Data Processing Assistant for Resting-State fMRI (DPARSF, http://rfmri.org/dpabi) [10] pipeline, one of the four preprocessing pipelines adopted by the ABIDE Preprocessed repository; further information can be found in the literature of Craddock *et al.* [11].

Voxel-based Morphometry

For structural MRI, preprocessing included segment, spatial normalization, modulation, and spatial smoothness. The raw T1w images of each subject were segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissue classes by calling the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/) [12] and transformed from individual native space into the Montreal Neurological Institute (MNI) space using the DARTEL normalization method [13]. Then, normalized GM data were modulated to obtain the GM volume (GMV) measure. Finally, GMV data were smoothed with an 8 mm Full-width-Half-Maximum (FWHM) Gaussian kernel. Total brain volume (TBV = GM + WM) was calculated for use as a co-variable in further analysis. According to the quality control information given in the phenotype file provided by the PCP, all scans rated as low quality (qc_anat_rater_1="fail" or qc_anat_rater_2="fail") were excluded from analysis.

We abstracted the GMV of 6 ROIs for statistical analysis: bilateral BST, PC, and PRC. All ROIs were defined by Desikan parcellation [14], and the GMV of each ROI was obtained by taking an average of all voxels located in the ROI.

Seed-based Functional Connectivity

For resting-state fMRI, preprocessing included removing the first 4 volumes, slice timing correction and head motion correction, regression of co variables, low-frequency filtering, spatial registration, spatial normalization, and smoothing [10]. We used a rigid body transformation to realign the time series of images for each subject. To reduce the respiratory and cardiac effects, signals from the WM and CSF were regressed out from the BOLD signal as co variables. Since the global signal is

considered to reflect respiratory, cardiac, and motion information which may lead to spurious correlations [15, 16], we applied global signal regression to remove these potential confounders. Then, linear and quadratic trends were also regressed out, and band-pass filtering (0.01–0.1 Hz) was applied. The transformation information acquired from T1w image normalization was used to perform spatial normalization for resting-state images. Finally, we smoothed using a 6 mm FWHM kernel. All the subjects with low quality (qc_func_rater_1="fail" or qc_func_rater_2=="fail") or large head motion (mean framewise displacement (FD) >0.2 mm) were excluded.

Seed-based FC of the bilateral BST, PC, and PRC was calculated. For each participant, we extracted averaged time series within all ROIs and calculated their temporal coupling with all other brain voxels using Pearson correlation, resulting in 6 FC maps.

Statistics

For ROI-level GMV, we examined the differences between the ASD group and the healthy control group using a linear mixed model (LMM, using the MatLab built-in function *fitlme*), which has been shown to effectively control for site effects in previous large-scale neuroimaging studies. For the GMV of each ROI, we set the model as follows:

$$GMV \sim Diagnosis + Sex + Age + TBV + (1|Site)$$

where the site effect is put into the model as a random intercept, while diagnosis, age, and sex are used as fixed effects. The *t*-value and *P*-value of the regression coefficient of *Diagnosis* were obtained from the LMM to reflect the GMV difference between ASD and HC. For all statistical results, we applied a significance threshold of $P < 0.05$ (two-tailed) and corrected for multiple comparisons using the Family-Wise Error (FWE) method (number of comparisons = 6).

For seed-based FC, we used a two-sample t-test to assess inter-group differences. Age, sex, site (coded as dummy variables), and head motion (mean FD) were regressed out as covariates before inter-group comparison. The t statistical maps were enhanced using threshold-free cluster enhancement (TFCE)

[17] with nonparametric permutation testing, which has been demonstrated to effectively increase reproducibility and control the false positive rate in voxel-level neuroimaging analysis [18, 19]. The TFCE was applied based on the statistical module of the DPARSF toolbox. The number of permutations was set to 5000 and the cluster-extent threshold was $Z = 2.3$. We applied a threshold of $P < 0.05$ (two-tailed) and corrected for multiple comparisons (FWE) based on the results of a permutation test using the TFCE value.

Additional Analysis

To confirm that our results were not due to medication use, we excluded all subjects on medication and repeated the analysis. Besides, we repeated our analyses using full-scale intelligence quotient (FIQ) as a covariate to control for the potential confounding effects of the subjects' intelligence. Intelligence in the Abide-I data set is mainly measured by Wechsler instruments (Wechsler Adult Intelligence Scale (WAIS) [20] and Wechsler Intelligence Scale for Children (WISC) [21], while some other series scales are also used; refer to http://fcon_1000.projects.nitrc.org/indi/abide/ for more information.

Supplemental References

1. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, *et al*. AnnotSV: An integrated tool for structural variations annotation. Bioinformatics 2018, 34: 3572–3574.

2. Ahmed M, Kim DR. Pcr: An R package for quality assessment, analysis and testing of qPCR data. PeerJ 2018, 6: e4473.

3. He X, Sanders SJ, Liu L, de Rubeis S, Lim ET, Sutcliffe JS, *et al*. Integrated model of *de novo* and inherited genetic variants yields greater power to identify risk genes. PLoS Genet 2013, 9: e1003671.

4. Jiang Y, Li Z, Liu Z, Chen D, Wu W, Du Y, *et al*. mirDNMR: A gene-centered database of background *de novo* mutation rates in human. Nucleic Acids Res 2017, 45: D796–D803.

5. Fan X, Fu Y, Zhou X, Sun L, Yang M, Wang M, *et al*. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. Sci Adv 2020,

6: eaaz2978.

6. Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L, *et al*. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell Res 2018, 28: 730–745.

7. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, *et al*. Comprehensive integration of single-cell data. Cell 2019, 177: 1888–1902.e21.

8. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, *et al*. SCENIC: Single-cell regulatory network inference and clustering. Nat Methods 2017, 14: 1083–1086.

9. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, *et al*. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 2014, 19: 659–667.

10. Yan. DPARSF: A MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. Front Syst Neurosci 2010: 13.

11. Cameron C, Yassine B, Carlton C, Francois C, Alan E, András J, *et al*. The Neuro Bureau Preprocessing Initiative: Open sharing of preprocessed neuroimaging data and derivatives. Front Neuroinform 2013, 7: 41.

12. Ashburner J, Friston KJ. Unified segmentation. NeuroImage 2005, 26: 839–851.

13. Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage 2007, 38: 95–113.

14. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, *et al*. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006, 31: 968–980.

15. Chang C, Glover GH. Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. NeuroImage 2009, 47: 1448–1459.

16. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage 2014, 84: 320–341.

17. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 2009, 44: 83–98.
18. Chen X, Lu B, Yan CG. Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. Hum Brain Mapp 2018, 39: 300–318.
19. Noble S, Scheinost D, Constable RT. Cluster failure or power failure? Evaluating sensitivity in cluster-level inference. Neuroimage 2020, 209: 116468.
20. Wechsler D. Wechsler adult intelligence scale: Manual. 1st ed. New York: Psychological corporation, 1955.
21. Wechsler D. Wechsler intelligence scale for children: Manual. 1st ed. New York: Psychological corporation, 1949.