Molecular Psychiatry (2017) 22, 1282–1290

www.nature.com/mp

IMMEDIATE COMMUNICATION

Targeted sequencing and functional analysis reveal brain-size-related genes and their networks in autism spectrum disorders

Jinchen Li[1,8], Lin Wang[1,8], Hui Guo[1], Leisheng Shi[2], Kun Zhang[2], Meina Tang[1], Shanshan Hu[2], Shanshan Dong[1], Yanling Liu[1], Tianyun Wang[1], Ping Yu[2], Xin He[3], Zhengmao Hu[1], Jinping Zhao[4], Chunyu Liu[1,5], Zhong Sheng Sun[2,6] and Kun Xia[1,7]

Autism spectrum disorder (ASD) represents a set of complex neurodevelopmental disorders with large degrees of heritability and heterogeneity. We sequenced 136 microcephaly or macrocephaly (Mic–Mac)-related genes and 158 possible ASD-risk genes in 536 Chinese ASD probands and detected 22 damaging *de novo* mutations (DNMs) in 20 genes, including *CHD8* and *SCN2A*, with recurrent events. Nine of the 20 genes were previously reported to harbor DNMs in ASD patients from other populations, while 11 of them were first identified in present study. We combined genetic variations of the 294 sequenced genes from publicly available whole-exome or whole-genome sequencing studies (4167 probands plus 1786 controls) with our Chinese population (536 cases plus 1457 controls) to optimize the power of candidate-gene prioritization. As a result, we prioritized 67 ASD-candidate genes that exhibited significantly higher probabilities of haploinsufficiency and genic intolerance, and significantly interacted and co-expressed with each another, as well as other known ASD-risk genes. Probands with DNMs or rare inherited mutations in the 67 candidate genes exhibited significantly lower intelligence quotients, supporting their strong functional impact. In addition, we prioritized 39 ASD-related Mic–Mac-risk genes, and showed their interaction and co-expression in a functional network that converged on chromatin remodeling, synapse transmission and cell cycle progression. Genes within the three functional subnetworks exhibited distinct and recognizable spatiotemporal-expression patterns in human brains and laminar-expression profiles in the developing neocortex, highlighting their important roles in brain development. Our results indicate some of Mic–Mac-risk genes are involved in ASD.

Molecular Psychiatry (2017) **22,** 1282–1290; doi:10.1038/mp.2017.140; published online 25 July 2017

INTRODUCTION

Autism spectrum disorder (ASD) describes a group of neurodevelopmental diseases with high heritability and is characterized by social communication impairment, repetitive behaviors and restricted interests.[1,2] Genome-wide association studies identified a handful of genomic loci, including 1p13.2 in the Han Chinese population we previous reported,[3] but continued to fall short in their reproducibility.

Progress in whole-exome sequencing (WES)[4–9] and whole-genome sequencing (WGS)[10–14] allowed their preferential employment in thousands of ASD subjects to detect *de novo* mutations (DNMs) involved in ASD[15,16] and other neuropsychiatric disorders,[17,18] and enabled detection of rare pathogenic variations and risk-genes. In addition, several studies employed protein–protein interaction (PPI) and co-expression networks for ASD-risk genes to investigate biological pathways they involve,[8,19–22] and found that candidate genes were mainly implicated in chromatin remodeling, Wnt signaling, transcriptional regulation and synaptic functions.[8,16] Moreover, previous studies revealed that ASD-risk genes exhibited dramatic spatiotemporal expression patterns in human brain,[19,20,23,24] accelerating the understanding of ASD pathology. However, known ASD risk-genes can explain only a small fraction of the probands.[8,22] Therefore, prioritizing novel and effective ASD risk-genes remains a critical issue.

Owing to the extremely high heterogeneity and complex pathophysiology of ASD,[16,25] investigating gene panels exhibiting similar functional effects in large cohorts will greatly assist in identifying pathogenic variations. Recently, targeted sequencing studies[26–29] identified several high-confidence candidate genes,[16,26,30] including *CHD8* and *PTEN* associated with macrocephaly (Mac); *CTNNB1* and *DYRK1A* associated with microcephaly (Mic). Abnormal head circumference (HC) is known to be highly comorbid with ASD,[26,31,32] which is caused by incorrect head growth and brain development during prenatal period and early childhood. Some Mic-associated genes actually interact with Mac-associated genes.[16] For example, Mac-associated gene, *CHD8* negatively regulates β-catenin targeted genes including Mic-associated gene *CTNNB1* via WNT pathway, influences neuronal

---

[1]The State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; [2]Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China; [3]Department of Human Genetics, University of Chicago, Chicago, IL, USA; [4]Mental Health Institute, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China; [5]Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA; [6]Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China and [7]Key Laboratory of Medical Information Research, Central South University, Changsha, Hunan, China. Correspondence: Professor K Xia or Professor C Liu, The State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan 410078, China or Professor ZS Sun, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China.
E-mail: liuchunyu@sklmg.edu.cn or xiakun@sklmg.edu.cn or sunzs@biols.ac.cn

[8]These authors contributed equally to this work.

Received 30 October 2016; revised 31 March 2017; accepted 19 May 2017; published online 25 July 2017

migration, polarity and synaptogenesis.[33] With this interaction, ASD proband with *CHD8* truncating variations have significant Mac, while ASD probands with *CTNNB1* truncating variations exhibited Mic.[16] Currently, more than 100 genes and loci have been found to be associated with Mic–Mac. However, no work to our knowledge has systematically elucidated the link between Mic–Mac-risk genes and ASD.

In this study, we hypothesize that some Mic–Mac-risk genes are involved in ASD; they might be important candidate genes to include in searching for mutations in ASD. With that, we sequenced 294 candidate genes including 136 Mic–Mac-risk genes and 158 possible ASD-risk genes, in 536 Chinese ASD patients. We then integrated the genetic and clinical data from the Simons Simplex Collection (SSC) and Autism Sequencing Consortium (ASC) for following aims (1) to prioritize additional ASD-risk genes and ASD-related Mic–Mac-risk (AMM) genes, (2) to characterize the relationship between AMM and ASD genes and (3) to investigate their functional networks and expression patterns in developing human brains.

MATERIALS AND METHODS

Targeted sequencing of subjects with ASD

A total of 536 unrelated ASD trios (probands and their unaffected parents) from the Autism Clinical and Genetic Resources in China (ACGC)[28] were recruited. Genomic DNA (2 μg) from 504 patients with ASD were used to construct a DNA library prior to targeted sequencing, using Agilent SureSelect target enrichment (Agilent Technologies, Santa Clara, CA, USA). In addition, 8 μg of DNA from 32 ASD trios (96 samples) was used to construct a genomic DNA library for WGS. For all DNA libraries, the Illumina Hiseq2000 sequencing system (Illumina, San Diego, CA, USA) was employed to generate the paired-end raw data (Figure 1). This study was approved by the Institutional Review Board of the State Key Laboratory of Medical Genetics, School of Life Sciences at Central South University, Changsha, Hunan, China. All subjects who participated in this study completed informed consent before the original sample collection.

We manually collected 105 Mic and 31 Mac-risk genes from PubMed, with each gene reportedly associated with Mic or Mac in at least one study (Supplementary Table 1). In addition, 158 possible ASD-risk genes were chosen according to the following criteria: (1) they harbored damaging DNMs including single-nucleotide variants (SNVs), insertions and deletions (indels), or located inside copy number variations in ASD or other neuropsychiatric disorders; (2) they were associated with ASD based on SFARI Gene database;[34] and (3) they have been detected damaging variations in the 32 Chinese ASD proband by WGS. This study mainly focused on prioritizing additional ASD and AMM risk-genes, and investigated their functional pathways and expression patterns. The targeted sequencing data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under accession numbers PRJCA000393.

Mutation detection and annotation

Burrows-Wheeler Aligner[35] was employed to align clean reads to the human reference genome (hg19), Sequence Alignment/Map tools[36] was used to remove duplicate reads and generate position-sorted files, and the Genome Analysis Toolkit HaplotypeCaller[37] was employed to detect SNVs and small indels. ANNOVAR[38] was then employed to annotate mutations as described previously.[39,40] Only loss-of-function (LoF) mutations (stop-loss, stop-gain, splicing-site SNVs and frameshift indels) and deleterious missense mutations (mutations predicted by ⩾9 out of 14 methods to be deleterious) with minor-allele frequencies < 0.001 (based on ExAC[41]) were retained and validated by Sanger sequencing (Figure 1).

Risk-gene prioritization

The Transmission and *De Novo* Association (TADA) program[42] was employed to prioritize risk genes. Damaging variations in the 294 sequenced genes from 1457 controls without neuropsychiatric disorders or related diseases from an in-house Chinese exome database were imported to the TADA program to improve the power of risk-gene prioritization. In-house Chinese exome database provides background variations in the Chinese population that were successfully used in several studies.[43–45] The DNMs of 4167 probands and rare inherited mutations (RIMs) of 2377 probands, as well as those of 1786 siblings by WES/WGS, were combined with genetic data from the Chinese cohort to acquire the best possible power for prioritizing ASD-risk genes (Figure 1). Publicly available DNMs were mainly sourced from SSC and ASC, which can also been downloaded from the NPdenovo database (www.wzgenomics.cn/NPdenovo),[39] and RIMs of individuals in SSC were obtained from a published study.[46] The intelligence quotients (IQs) and HCs of ASD patients were sourced from SSC.

Functional networks and spatiotemporal expression patterns

Human PPI networks were obtained from STRING database (http://string.embl.de/). The spatiotemporal transcriptomes of the human brain across anatomically different regions and developmental periods were downloaded from BrainSpan (http://www.brainspan.org/). We calculated Pearson correlation coefficients between any two genes based on their expression levels in all 524 tissue samples. Gene pairs with $|R| \geqslant 0.8$ were regarded as being co-expressed and defined as a connection in the human brain as previous study.[47] The permutation test was performed similar to a previous study[47] to evaluate the functional connections of candidate genes (Figure 1). In brief, we compared the number of co-expressed genes within the gene set or with known ASD-risk genes and their connections with those of 1 000 000 random iterations. The permutation test has took gene length and expected gene-level DNM rate[48] into consideration.

We then merged the co-expression networks derived from the candidate genes with the PPI networks for the same genes. The connections between any two genes were defined by information from the PPI and/or co-expression networks. AMM genes were regarded as ‘seed genes’ that were directly connected and used to form an interconnected functional network. Known Mic–Mac- or ASD-risk genes based on the SFARI Gene database[34] that were connected with at least two of AMM genes were added to the above network (Figure 1). Network figures were created using Cytoscape 3.1.0 (http://www.cytoscape.org/). Biological process of Gene Ontology (GO) was used to report the biological pathway in the functional network. Furthermore, human brain co-expression data from HBT[49] (http://hbatlas.org/) and PPI data from InWeb_IM[50] (http://www.intomics.com/inbio/map) were used to replicate the functional network. InWeb_IM was a scored experimental PPI network recently released, with less than one third of interaction overlapped with STRING.

Signed hybrid-weighted gene co-expression network analysis (WGCNA)[51] was performed across all tissue samples from BrainSpan using the standard method with a power of six to cluster the spatiotemporal-expression patterns and prenatal laminar-expression profiles[52] of a given gene set.

RESULTS

Targeted sequencing and quality control

Targeted sequencing of 294 genes was employed in a cohort of 504 ASD probands from the ACGC (Figure 1). On average, we obtained 293.39 MB of clean data for each sample, about two-thirds of which were mapped to target regions, resulting in ~ 120× sequence depth (Supplementary Table 2). In addition, an average of 97.90% and 95.52% of the target regions covered at least 10 and 20 reads, respectively. In another project, we employed WGS (> 30×) in another 32 ASD trios to detect all classes of mutations in the genome. Here, to maximize the

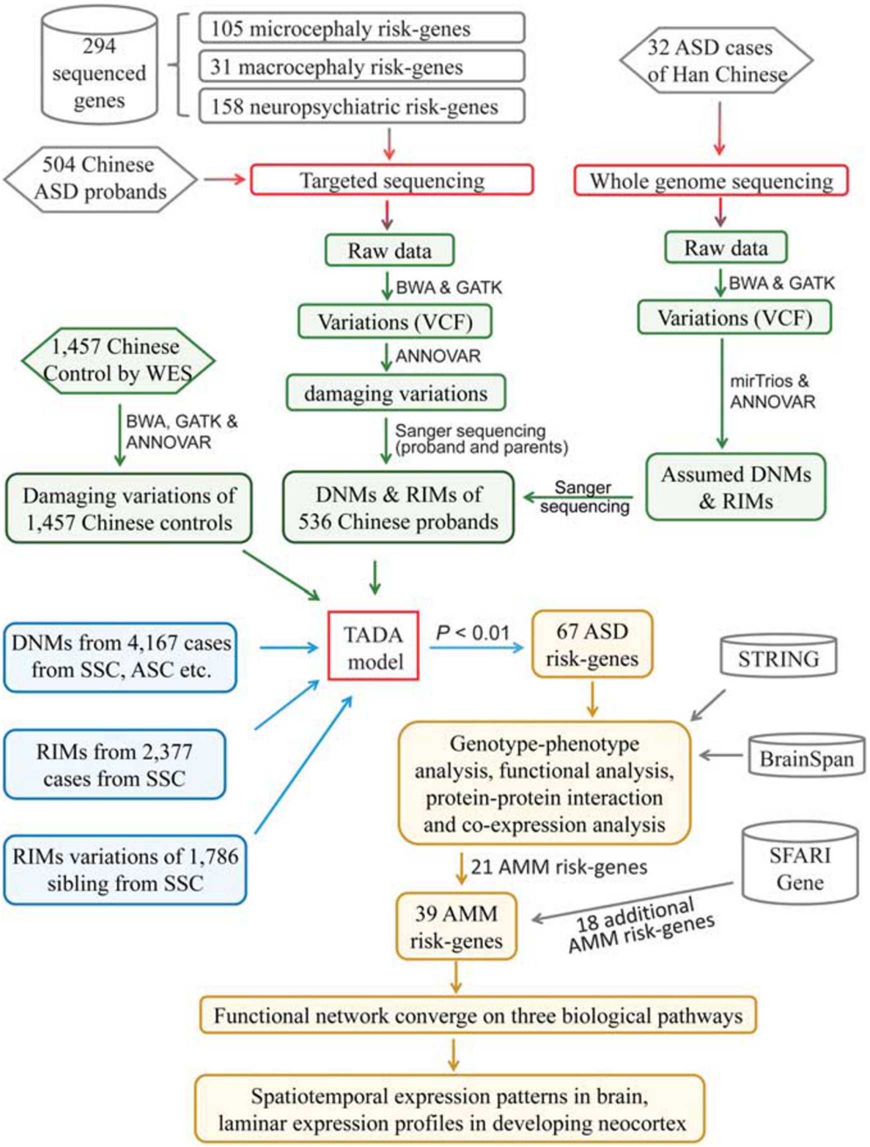

Figure 1. Workflow of data analysis. Targeted sequencing of 294 genes as employed in 504 probands, WGS as employed in 32 ASD trios. Rare damaging mutations in the 294 genes from 536 samples were validated by Sanger sequencing. Genetic variations in the 294 genes from publicly available WES/WGS studies (4167 probands plus 1786 controls) along with our Chinese population (536 cases plus 1457 controls) were used to prioritize risk-genes using the TADA model. Then, we employed genotype-phenotype, genetic functional, PPI and gene co-expression analysis to investigate the functional impact of candidate genes. Finally, we developed a functional network that converged on three pathways for AMM risk-genes and characterized their spatiotemporal-expression patterns in brain and laminar-expression profiles in developing neocortex. AMM, ASD-related Mic–Mac-risk; ASD, autism spectrum disorder; Mac, macrocephaly; Mic, microcephaly; PPI, protein–protein interaction; TADA, Transmission and *De Novo* Association; WES, whole-exome sequencing; WGS, whole-genome sequencing.

Chinese sample size, we added genetic variations found in 294 genes from 32 probands to those from 504 probands (Figure 1). Together, variation data from 536 Chinese ASD trios were used for further analysis.

De novo mutations and X-linked variations

We detected and verified a total of 496 LoF or deleterious missense mutations with minor-allele frequency < 0.001 in the 294 genes (Supplementary Table 3). After confirming the state of these mutations in parents by Sanger sequencing, we detected 22 DNMs, including four nonsense mutations, four frameshift indels and 14 deleterious missense mutations, in 20 genes (Table 1). Based on previously reported expected DNM rates,[48] we observed significantly higher DNM burden for the 294 sequenced genes (observed $n = 22$ vs expected $n = 3.98$; odds ratio $= 5.53$; binomial test $P = 9.23 \times 10^{-11}$). In our recently developed NPdenovo database,[39] nine of the 20 genes also harbored damaging DNMs in published WES/WGS results, including two genes (*CHD8* and *SCN2A*) with recurrent DNMs and *POGZ*, *MECP2* and *DYRK1A* with LoF DNMs. We also detected a novel frameshift mutation in *NRXN2* (c.808dupG, p.A270fs), a known causative gene associated with ASD.

In our previous study,[28] we sequenced 189 ASD-candidate genes using single-molecule molecular-inversion probes (smMIPs), with 37 genes including 11 Mic–Mac-associated genes from 504 samples overlapping with this study. Here, we recaptured the five DNMs identified in the previous study, as well as two additional DNMs that were missed in the smMIP study (Table 1). For the

Table 1. Detected DNMs and hemizygous variations

Chr.	*Position*	*Ref*	*Alt*	*Gene symbol*	*GenBank No.*	*Effect*	*Nucleotide change*	*AA. alteration*	*DNMs in SSC or ASC*	*PH (percentile)*	*RVIS (percentile)*
DNMs											
chr14	21862265[a]	–	T	*CHD8*	NM_001170629	Frameshift	c.5688dupA	p.R1897fs	7 LoF, 2 Dmis	0.637 (10.3%)	−2.337 (1.2%)
chr14	21862296	AG	–			Frameshift	c.5657delAG	p.T1886fs			
chr21	38853069[a]	G	T	*DYRK1A*	NM_001396	Stop-gain	c.457G > T	p.E153X	4 LoF	0.658 (9.6%)	−0.422 (25.6%)
chrX	153296471[b]	G	A	*MECP2*	NM_001110792	Stop-gain	c.844C > T	p.R282X	1 Dmis	0.359 (24.3%)	−0.885 (10.5%)
chr11	64457918	–	G	*NRXN2*	NM_015080	Frameshift	c.808dupG	p.A270fs	–	0.197 (43.7%)	−2.164 (1.4%)
chr1	151402109[a]	G	A	*POGZ*	NM_207171	Stop-gain	c.379C > T	p.Q127X	3 LoF	0.192 (44.5%)	−1.527 (3.4%)
chr2	166179813[a]	C	T	*SCN2A*	NM_001040143	Stop-gain	c.1819C > T	p.R607X	4 LoF, 7 Dmis	0.566 (12.7%)	−1.990 (1.8%)
chr2	166170189[a]	C	T			Missense	c.1094C > T	p.T365M			
chr9	3248164	TA	–	*RFX3*	NM_001282116	Frameshift	c.1835delTA	p.L612fs	–	0.934 (2.5%)	−0.799 (12.5%)
chr4	114275103[b]	G	C	*ANK2*	NM_001148	Missense	c.5329G > C	p.V1777L	4 LoF, 4 Dmis	0.953 (2.0%)	−3.329 (0.4%)
chr8	1877564	G	A	*ARHGEF10*	NM_014629	Missense	c.3034G > A	p.A1012T	1 Dmis	0.192 (44.6%)	−1.016 (8.1%)
chr9	136913560	C	T	*BRD3*	NM_007371	Missense	c.731G > A	p.R244Q	–	0.437 (18.8%)	−0.705 (14.8%)
chr15	49048380	T	C	*CEP152*	NM_001194998	Missense	c.3065 A > G	p.Q1022R	–	0.139 (57.8%)	0.972 (90.2%)
chr1	240072174	A	T	*CHRM3*	NM_000740	Missense	c.1423 A > T	p.I475F	1 Dmis	0.177 (47.5%)	−0.136 (43.8%)
chr1	53676893	T	C	*CPT2*	NM_000098	Missense	c.1547 T > C	p.F516S	–	0.115 (66.5%)	0.514 (80.3%)
chr5	127702120	T	G	*FBN2*	NM_001999	Missense	c.2252 A > C	p.E751A	–	0.625 (10.7%)	−1.868 (2.0%)
chr12	52635307	A	G	*KRT7*	NM_005556	Missense	c.745 A > G	p.M249V	–	0.250 (35.4%)	1.067 (91.7%)
chr18	48241495	G	A	*MAPK4*	NM_001292040	Missense	c.593G > A	p.R198Q	–	0.136 (58.7%)	−0.511 (21.7%)
chr16	14340619	C	T	*MKL2*	NM_014048	Missense	c.1502C > T	p.S501F	–	0.316 (27.9%)	−1.549 (3.3%)
chr22	36697021	C	T	*MYH9*	NM_002473	Missense	c.2714G > A	p.R905H	1 Dmis	0.259 (34.3%)	−1.994 (1.7%)
chr9	134396829	C	T	*POMT1*	NM_001136114	Missense	c.1444C > T	p.R482W	–	0.134 (59.6%)	−0.301 (32.3%)
chr9	71849443	T	C	*TJP2*	NM_001170415	Missense	c.1772 T > C	p.L591S	–	0.302 (29.1%)	1.459 (95.2%)
Hemizygous variations											
chrX	76814250	C	T	*ATRX*	NM_138270	Missense	c.6280G > A	p.V2094I	–	NA	−0.926 (9.8%)
chrX	18622983	C	A	*CDKL5*	NM_003159	Missense	c.1939C > A	p.P647T	–	0.432 (19.1%)	−0.667 (15.9%)
chrX	153594754	C	T	*FLNA*	NM_001110556	Missense	c.1150G > A	p.V384M	–	0.752 (7.1%)	−3.240 (0.4%)
chrX	13771497	G	C	*OFD1*	NM_003611	Missense	c.1066G > C	p.E356Q	1 LoF, 1 Dmis	0.106 (70.4%)	−0.176 (40.6%)
chrX	153694155	A	T	*PLXNA3*	NM_017514	Stop-gain	c.2497 A > T	p.K833X	1 LoF	0.114 (66.9%)	−2.494 (0.9%)
chrX	153698849	C	T			Missense	c.5051C > T	p.S1684L			
chrX	154493543	G	A	*RAB39B*	NM_171998	Missense	c.31C > T	p.L11F	–	0.432 (19.1%)	0.013 (54.6%)
chrX	152959387	C	A	*SLC6A8*	NM_001142805	Missense	c.C1139A	p.P380Q	1 LoF	0.415 (20.2%)	−0.737 (13.9%)

Abbreviations: ASC, Autism Sequencing Consortium; ASD, autism spectrum disorder; Dmis, deleterious missense mutations; DNM, *de novo* mutation; LoF, loss-of-function mutations; NA, not applicable; PH, probability of haploinsufficiency and percentile rank; RVIS, residual variation intolerance score and percentile rank; smMIP, single-molecule molecular inversion probes; SSC, Simons Simplex Collection. We totally detect 22 extreme DNMs in 20 genes from 536 Han Chinese ASD probands.). All these mutations have been verified by Sanger sequencing. DNMs of SSC and ASC were sourced from NPdenovo database we developed (http://www.wzgenomics.cn/NPdenovo/index.php). [a]DNMs also been detected in our previous studies using smMIPs. [b]DNMs were missed in smMIPs study but detected in this study.

p.R282X site in *MECP2*, 96 reads covered here as compared with only two reads in the smMIP study (Supplementary Figure 1). The p.V1777L in *ANK2* was missed because its inheritance not having been validated in previous study. In addition, 11 genes harboring DNMs in only Chinese patients were detected in present study, including a frameshift deletion (p.L612fs) in *RFX3* that is essential for brain development.[53] Furthermore, for the overlapped 37 genes in our previous study,[28] additional 20 DNMs in eight genes were identified in another independent Chinese cohorts, including two Mic-associated genes (*CDKL5*, *DYRK1A*) and two Mac-associated genes (*CHD8*, *MECP2*) (Supplementary Table 4).

Among the 496 damaging variations, eight in seven risk genes (*ATRX*, *CDKL5*, *FLNA*, *OFD1*, *PLXNA3*, *RAB39B* and *SLC6A8*; Table 1) were inherited X-linked hemizygous variations. *PLXNA3*, which plays important roles in axon guidance in the developing nervous system,[54] harbored one nonsense (c.2497 A > T, p.K833X) and one missense (c.5051C > T, p.S1684L) variation in two unrelated male patients. Meanwhile, a splicing-site DNM of this gene in a Caucasian male proband was detected (c.1928+1G > A) by WES,[46] suggesting that *PLXNA3* may play important roles in ASD molecular pathology.

To evaluate the functional impact of DNMs and hemizygous variations, we sourced gene-level haploinsufficiency rates[55] and residual variation intolerance scores (RVIS)[56] from original studies. Compared with all RefSeq genes (background), these genes exhibited significantly higher probabilities of haploinsufficiency ($P = 0.0098$; Supplementary Figure 2a) and lower RVIS ($P = 0.0021$; Supplementary Figure 2b). Specifically, 21 (77.78%) and 22 (81.48%) of 27 genes ranked in the top 50% of haploinsufficiency and RVIS, respectively (Table 1), indicating that they are likely to be intolerant of damaging mutations.

Prioritization of ASD-candidate genes and their functional impact

In addition to DNMs, RIMs play important roles in ASD etiology.[46,57,58] Integrating datasets from multiple independent studies offers insight into ASD-candidate genes.[8,23,59,60] We then employed TADA program,[42] to prioritize candidate genes. Based on genetic variations in 536 patients with ASD and 1457 Chinese

Table 2. Prioritized of ASD candidate genes

Classes	*Mic risk-genes* (n = 15)	*Mac risk-genes* (n = 6)	*Non-MIC and Mac risk-genes* (n = 46)	
$P_{TADA} \leqslant 0.0001$ (Strong, $n = 10$)	*DYRK1A, RELN*	*CHD8, PTEN*	*SCN2A, ANK2, POGZ, KATNAL2, SUV420H1, SYNGAP1*	
$0.0001 < P_{TADA} \leqslant 0.001$ (Suggested, $n = 10$)	*NRXN1*		*DYNC1H1, FCRL6, GABRB3, MFRP, MUC4, NR3C2, SPP2, TBR1, ZNF626*	
$0.001 < P_{TADA} \leqslant 0.005$ (Positive, $n = 17$)	*CHMP1A, TCF4, ASPM, CTNNB1*	*MECP2*	*ASXL3, C15orf62, CHRM3, PRPF39, RAD21L1, SLC25A39, STXBP1, TSPAN17, COL28A1, CPT2, RFX3, SCP2, PDCD1*	
$0.005 < P_{TADA} < 0.01$ (Possible, $n = 23$)	*CEP135, GNAS, TRAPPC9, VIP, ATP6V0A2, POMT1*		*APH1A, ETFB, KRT26, CD276, NRXN2, PAFAH1B2, RAB2A, RNF38, SLC7A7, ST3GAL6, TM4SF19, DNAH3, KCNK7, LILRB2, ZNF548, CACNA1D*	
X-linked genes ($n = 7$)	*ATRX, CDKL5*	*RAB39B, SLC6A8, OFD1*	*FLNA, PLXNA3*	
Chinese probands with DNMs ($n = 536$)	2 (0.37%)	3 (0.56%)	8 (1.19%)	Subtotal: 13 (2.42%)
Chinese probands with RIMs ($n = 536$)	36 (6.71%)	3 (0.56%)	83 (15.49%)	Subtotal: 122 (22.76%)
Published probands with DNMs ($n = 4167$)	26 (0.62%)	15 (0.36%)	78 (1.87%)	Subtotal: 117 (2.81%)
Published probands with RIMs ($n = 2377$)	101 (4.25%)	15 (0.63%)	205 (8.62%)	Subtotal: 318 (13.38%)
All probands with DNMs ($n = 4703$)	28 (0.60%)	18 (0.38%)	86 (1.83%)	Subtotal: 130 (2.76%)
All probands with RIMs ($n = 2913$)	137 (4.70%)	18 (0.62%)	288 (9.89%)	Subtotal: 440 (15.10%)

Abbreviations: DNMs, *de novo* mutations; MAC, macrocephaly; MIC, microcephaly; RIMs, rare-inherited mutations; TADA, Transmission and *De Novo* Association; WES, whole-exome sequencing; WGS, whole-genome sequencing. The TADA model combines information of DNMs and RIMs in different functional classes with proper weightings to prioritize candidate genes. Candidate genes were classified into different groups based on the P_{TADA} in Chinese population and/or the combination of genetic data from published WES/WGS studies as shown in Supplementary Table 6.

controls, as well as DNMs from 4167 probands, RIMs from 2377 probands and 1786 controls from SSC or ASC (Supplementary Table 5), we prioritized 60 candidate genes ($P_{TADA} < 0.01$; Table 2 and Supplementary Table 6). Compared with random expectations ($n = 3$), there was significant enrichment of genes with $P < 0.01$ ($n = 60$) corresponding to a false-discovery rate of 0.05 (Supplementary Figures 3a–b). In addition, seven genes carrying hemizygous variations were automatically listed as candidates due to their involvement in ASD, intellectual disability (ID) or other X-linked neurodevelopmental disorders. Together, 67 ASD-candidate genes were finally prioritized, and their DNMs and RIMs were detected in 2.76% (130 of 4703) and 15.10% (440 of 2913) of cataloged probands, respectively (Table 2). We also compared the 67 candidate genes to two previous TADA gene sets identified from large cohorts of WES data.[8,22] Stephan *et al.* paper identified 65 genes,[22] of which 16 were included in our panel and 15 of 16 genes were prioritized in present study. Another study by Silvia *et al.* identified 108 genes,[8] we prioritized 21 of 25 genes that were included in our panel (Supplementary Figure 4). We consider this high reproducibility.

We found that the 67 candidate genes exhibited significantly higher probabilities of haploinsufficiency ($P = 0.00025$; Supplementary Figure 2a) and lower RVIS ($P = 7.39 \times 10^{-6}$; Supplementary Figure 2b). In addition, compared to all ASD patients, probands with DNMs in candidate genes exhibited a significantly lower full-scale IQ ($P = 0.00023$), verbal IQ ($P = 0.0081$) and non-verbal IQ ($P = 0.00021$) (Supplementary Figures 5a–c). Probands with RIMs exhibited moderately lower IQ ($P < 0.05$), consistent with previous findings.[46] Furthermore, permutation test was performed as our previous study[39] to investigate the relationships of the 67 candidates based on co-expression networks in the human brain and PPI data. It revealed significantly co-expression (Supplementary Figure 6a–d) and interaction (Supplementary Figures 7a–d) between candidates, as well as with known ASD-risk genes. Since the 294 sequenced genes including the 136 Mic–Mac genes were significantly interacted and co-expressed than random expectation ($P < 0.01$), we re-analysis the connectedness of the 67 candidate genes by regarding the 294 sequenced genes as background in the progress of random sampling. As a result, candidate genes were consistently co-expressed and interacted more than by chance ($P < 0.05$, Supplementary Table 7). To replicate the strong interconnectedness of candidate genes, we sourced brain co-expression data from HBT,[49] and human PPI data from InWeb_IM[50] recently developed, and found the candidates were significantly interacted and co-expressed than random expectations (Supplementary Figures 8a–d). These results indicated that the 67 candidates are more likely to be associated with ASD as compared with random expectations.

Mic–Mac-candidate genes in ASD

We noted that 21 of 67 candidate genes were shared between Mic–Mac and ASD (Table 2). These included both previously well-studied genes, such as *DYRK1A*, *CHD8* and *PTEN*, and several novel genes. We detected four frameshift indels (p.V531fs, p.L891fs, p.G1937fs and p.Q2594fs) and three damaging missense variations (p.R490C, p.E1132G and p.R1667H) of *ASPM* in nine unrelated patients, which encodes a centrosomal protein that regulates the Wnt pathway during neurogenesis. In addition, 1 missense DNM (p.C1557R) and 15 RIMs of *ASPM* were detected in ASD patients from the SSC,[46] indicating that the Mic-risk gene *ASPM* is associated with ASD ($P_{TADA} = 0.0013$).

We found that DNMs and RIMs in the 21 AMM genes were detected in 0.98% (46 of 4703) and 5.32% (155 of 2913) of ASD patients, respectively (Table 2). We calculated standardized HC (*Z*-score) of probands from SSC based on age and gender as in previous study[26] and it turned out that probands with DNMs or

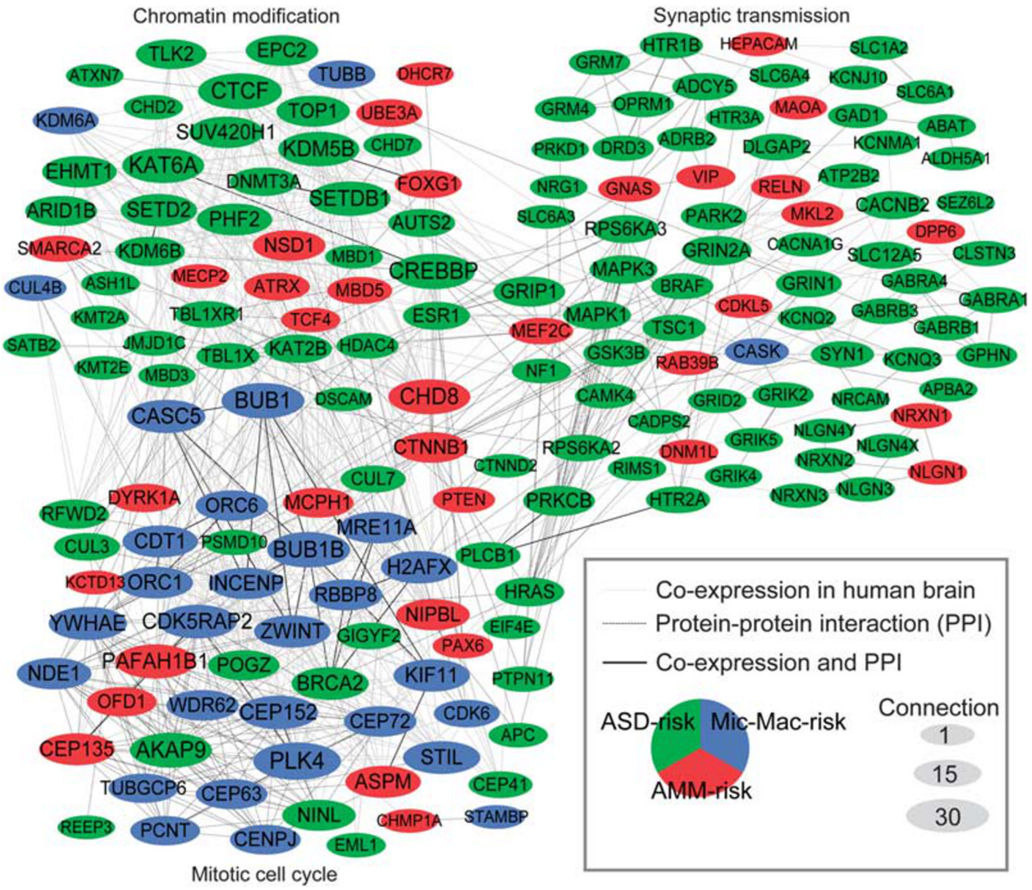

Figure 2. Co-expression and PPI networks. AMM genes form a biological network involved in chromatin modification, synaptic transmission and the mitotic cell cycle. Only genes implicated in the above three functional subnetworks were shown. Biological process of GO was used to report the biological pathway. AMM genes were interconnected and co-expressed with one another, as well as with known ASD or Mic–Mac-risk genes. All nodes are sized based on the degree of connectivity. AMM, ASD-related Mic–Mac-risk; GO, Gene Ontology; Mac, macrocephaly; Mic, microcephaly; PPI, protein–protein interaction.

RIMs in ASD-related MIC-risk genes exhibited significantly lower HC, and those in ASD-related macrocephaly-risk genes exhibited significantly higher HC (both $P < 0.05$; Supplementary Figure 9). In addition, we found Chinese probands harboring DNMs in *CHD8* and *MECP2* show macrocephaly and probands with DNMs of *CDKL5* and *DYRK1A* show MIC, consistent with previous reports.[26,61]

Convergent networks of AMM genes

In addition to the 21 prioritized AMM genes, we cataloged 18 additional genes listed as both Mic–Mac- and ASD-risk genes, including *DPP6*, *FOXG1*, *MBD5*, *UBE3A* and *VPS13B*, based on the SFARI Gene database (Supplementary Table 8). Together, 39 AMM risk-genes were used as ‘seed genes’ to develop a functional network based on PPI and co-expression connections. Known Mic–Mac- or ASD-risk genes that co-expressed or interconnected with at least two AMM genes were merged into this network, resulting in a functional network comprising 178 genes, converging on chromatin modification ($n = 51$), synaptic transmission ($n = 84$) and the mitotic cell cycle ($n = 56$) (Figure 2 and Supplementary Table 9). The three modules included 12, 17 and 11 AMM risk-genes, respectively.

We noted that *DYRK1A*, a MIC causative gene in ASD, was clustered in the subnetwork associated with the mitotic cell cycle, which is involved in brain growth and neurogenesis.[16] Variations that disrupt genes involving mitotic cell cycle progression could lead to dysfunctional mitosis of neural cells and scarcity in the central nervous system pool, highlighting the important roles of cell cycle genes in the molecular pathology of ASD and Mic–Mac. In addition, we found that different subnetworks were connected by some hub genes (such as *CTNNB1*) that regulate multiple functional pathways. To replicate these results of AMM genes in other independent data, we constructed a functional network based on data from InWeb_IM and HBT and found that network genes enriched in nervous system development, synapse organization, chromatin organization, cell cycle and regulation progression (Supplementary Table 10).

Expression patterns involved in brain development

To characterize the spatiotemporal-expression patterns of the 178 genes in convergent networks during brain development, we performed WGCNA in the tissue samples from BrainSpan and identified three independent modules (M1–M3) comprising 146 genes (Figure 3a and Supplementary Table 9). We found that genes within M1 ($n = 72$) were highly expressed in the human brain during the embryonic and early-to-middle fetal periods (8–26 post-conceptual weeks (pcw)), but gradually decreased and bottomed out at the age of 2-year old. In contrast, the expression levels of M2 genes ($n = 46$) were minimal during the embryonic period and gradually increased during the prenatal period, reaching a stable level after birth. M3 genes ($n = 28$) were highly expressed during 8 pcw to 9 pcw in some subregions, including

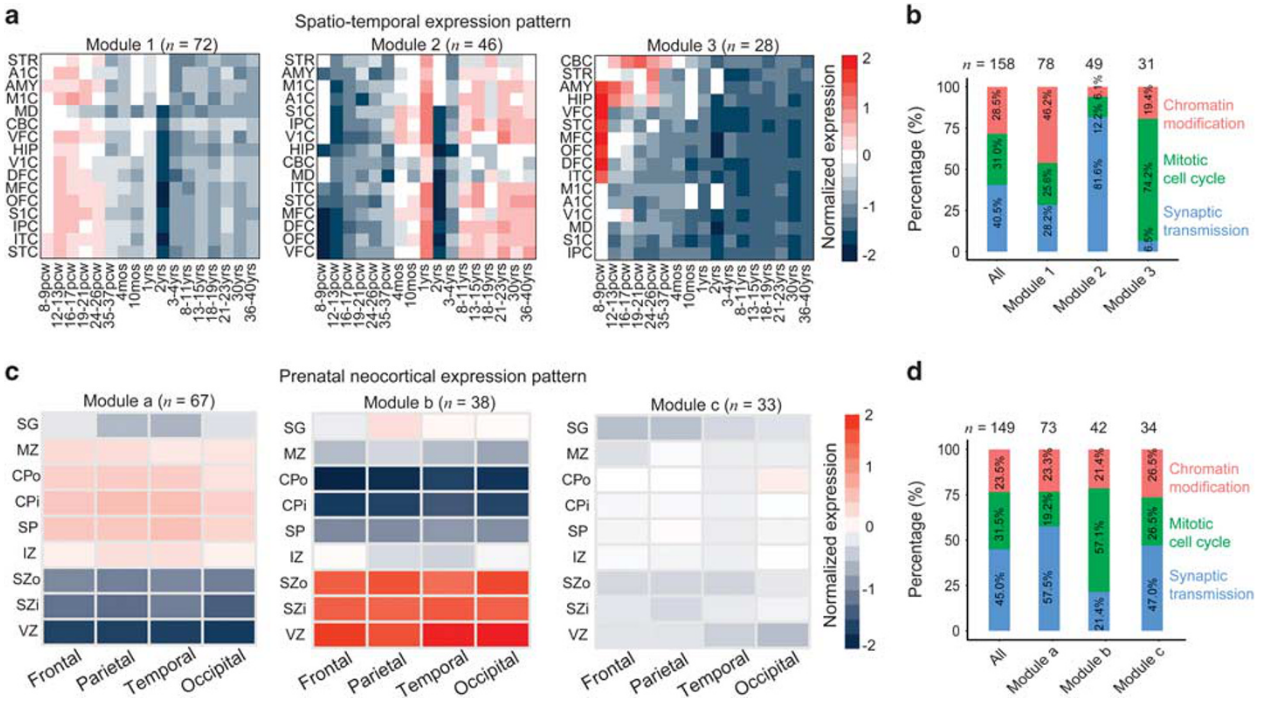

Figure 3. Expression patterns in the human brain. (**a**) The three spatiotemporal expression patterns (M1, M2 or M3) for network genes based on RNAseq data from BrainSpan, and corresponding to 17 developmental stages across 16 subregions. (**b**) Genes in different functional subnetworks overlapped with M1, M2 or M3, with the number of genes and percentages indicated. Biological process of GO was used to report the biological pathway. (**c**) Characterization of neocortical expression profiles (Ma, Mb and Mc) for network genes. Four subregions of the developing neocortex, delineating nine layers per subregion, were analyzed. (**d**) For each functional subnetwork, the number of genes and their percentage of overlap with each model are shown. Expression levels are normalized according to the $\log_2$-fold of the average expression level of each gene. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. A1C, primary auditory cortex; AMY, amygdaloid complex; CPo, outer cortical plate; CPi, inner cortical plate; DFC, dorsolateral prefrontal cortex; GO, Gene Ontology; HIP, hippocampus; IPC, posteroinferior parietal cortex; ITC, inferolateral temporal cortex; M1C, primary motor cortex; MD, mediodorsal nucleus of thalamus; MFC, anterior cingulate cortex; MZ, marginal zone; OFC, orbital frontal cortex; STC, posterior superior temporal cortex; SG, subpial granular zone; SP, subplate zone; IZ, subplate zone; STR, striatum; SZo, outer subventricular zone; SZi, inner subventricular zone; S1C, primary somatosensory cortex; V1C, primary visual cortex; VFC, ventrolateral prefrontal cortex; VZ, ventricular zone.

the frontal cortex, hippocampus and amygdaloid complex, followed by a sharp decrease in expression and stabilization at low levels. We replicated above recognizable spatiotemporal-expression patterns based on HBT expression data (Supplementary Figure 10). Furthermore, we found that the three modules significantly overlapped with the three subnetworks previously (Figure 3b). Specifically, M1, M2 and M3 genes were significantly enriched in genes from subnetworks involving chromatin modification (hypergeometric test, $P = 4.71 \times 10^{-6}$), synapse transmission ($P = 0.00039$) and the mitotic cell cycle, respectively ($P = 2.20 \times 10^{-14}$).

Since expression levels of candidate genes were dramatically fluctuated in the human cortex during the fetal period as report above, we then characterized prenatal neocortical-expression profiles based on laser-microdissection data from 526 prenatal neocortical samples[52] (15–21 pcw). Interestingly, we identified three co-expression modules (Ma, Mb and Mc) having distinct laminar signatures (Figure 3c and Supplementary Table 9). The expression of genes within Ma ($n = 67$) was high in the middle layers but lower in marginal to the intermediate zone, whereas that of Mb ($n = 38$) displays opposite trends, which was lower in the middle layers and high in the deep cortical layers and the outer subventricular to the ventricular zone. Moreover, we found that Mb significantly overlapped with cell cycle genes ($P = 8.82 \times 10^{-9}$), indicating that intricate regulations of the mitotic cell cycle play important roles in the developing cortex (Figure 3d).

DISCUSSION

Recent large-scale WES/WGS studies illuminated the essential roles of DNMs and RIMs in ASD and accelerated the understanding of the neurobiological and genetic basis of ASD.[4–14] Here, we detected several pathogenic DNMs, including recurrences in known strong-risk genes, in ASD patients from a Chinese cohort. Generally, DNMs have strong effects on the function of genes, especially for the genes with high probabilities of haploinsufficiency and genic intolerance. We combined genetic data from Chinese and Caucasian populations and prioritized 67 ASD-risk genes. Probands with DNMs or RIMs involving risk genes exhibited significantly lower IQ, demonstrating the functional effects of these genes. Although pathogenesis and mechanisms of only a small fraction of ASD risk-genes have been validated by *in vitro* or *in vivo* experiments, such as *ANK3*,[62] *CTNND2*,[63] *DYRK1A*,[34] *GLRA2*,[64] *PTEN*,[65,66] *PRICKLE2*,[67] *SHANK3*,[68] *SLC6A3*[69] and *TRPC6*,[70] our study provided a meaningful reference for further studies of ASD-risk genes. The fact that DNMs of these candidate genes contributed to a small fraction of ASD patients suggested more resequencing studies were needed to identify more strong candidates and decipher genetic mechanisms. Of interest, we also prioritized several risk-genes that only detected DNMs in Chinese cohorts and these genes warranted further studies.

Though ASD is highly clinical and genetic heterogeneous, some ASD candidate genes with DNMs in fact can define clinical subtype of ASD with characterized phenotypes.[71] For example, individuals with DNMs in *CHD8* were characterized by ID, macrocephaly, distinct faces and gastrointestinal complaints,[72] while DNMs in *DYRK1A* is associated with ASD, ID, MIC, febrile seizures in infancy, impaired speech, hypertonia and a specific facial gestalt.[30] The approach investigating the relationship between phenotypes and genotypes not only deepens the understanding of ASD etiology but also assists in clinical diagnose and intervention. In present study, we prioritized 39 genes associated with both Mic–Mac and ASD. Patients having DNMs or RIMs in these genes exhibited abnormal HC, highlighting the shared genetic mechanisms underlying ASD and Mic–Mac. Systematic discovery of genetic variants associated with both Mic–Mac and ASD may be helpful to test the pathogenic effects of AMM genes and in determining certain phenotypes.

We also performed PPI and co-expression analyses to investigate the related functional pathways, with results converging on chromatin modification, synaptic transmission and the mitotic cell cycle. Previous studies demonstrated that chromatin and synaptic genes play important roles in ASD pathophysiology,[8,22] and our study highlighted cell cycle genes involved in ASD and Mic–Mac. Of interest, cell cycle and Wnt-signaling pathway exhibit complex interaction and both of them were involved in regulation of cell fate, growth and proliferation.[73] ASD patients with DNMs in the β-catenin/Wnt-signaling network have been observed to be associated with both macrocephaly and MIC.[71] Consequently, we found that 14 AMM genes (*ASPM*, *CHD8*, *CDKL5*, *CTNNB1*, *DYRK1A*, *FOXG1*, *HEPACAM*, *OFD1*, *PAX6*, *PTEN*, *MECP2*, *MEF2C*, *NIPBL* and *TCF4*) were involved in the Wnt pathway. Of note, a recent study demonstrated that *CHD8* regulates cortical development by stimulating transcription of cell cycle genes,[74] providing new insight into the roles of the cell cycle and Wnt signaling in ASD and Mic–Mac.

We characterized spatiotemporal-expression patterns of genes in networks and correlated these patterns to functional pathways. The expression profiles of genes in the M1 and M2 clusters were similar to previous reports related to ASD.[23] However, we identified novel patterns of M3 genes highly expressed during the early fetal period and significantly overlapping with cell cycle subnetworks. Characterization of their cerebral cortex laminar features revealed that network genes were highly expressed in different cortical layers, corresponding to the three models. As shown in module Mb, the cell cycle genes were highly expressed in deep cortical layers highlighting their essential roles during brain development. Genes involved in similar molecular functions usually exhibited similar expression patterns in human brain, providing new insights in understanding ASD etiology.

In summary, through sequencing of MIC and macrocephaly-related genes, and previously known ASD candidate genes in Chinese ASD cohorts, we identified new DNMs and RIMs in these genes. We further prioritized novel risk genes, and identified their associated brain expression features and biological networks. Our results shed light on deciphering etiologies of ASD and Mic–Mac and facilitate discovery of information necessary for accurate genetic counseling and testing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the patients and their families for their participation in this study. The project was funded by the National Basic Research Program of China (2012CB517900) and the National Natural Science Foundation of China (31571301, 81330027, 31400919 and 31671114). HG was supported by the Young Talent Lifts Project of CAST.

REFERENCES

1 de la Torre-Ubieta L, Won HJ, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* 2016; **22**: 345–361.
2 Lai MC, Lombardo MV, Baron-Cohen S. Autism. *Lancet* 2014; **383**: 896–910.
3 Xia K, Guo H, Hu Z, Xun G, Zuo L, Peng Y *et al.* Common genetic variants on 1p13.2 associate with risk of autism. *Mol Psychiatry* 2014; **19**: 1212–1219.
4 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
5 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
6 O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 2012; **485**: 246–250.
7 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
8 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
9 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
10 Jiang Y-h YuenRK, Jin X, Wang M, Chen N, Wu X *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Human Genet* 2013; **93**: 249–263.
11 Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N *et al.* Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med* 2015; **21**: 185–191.
12 Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I *et al.* Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am J Hum Genet* 2016; **98**: 58–74.
13 Yuen RK, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B *et al.* Genome-wide characteristics of *de novo* mutations in autism. *NPJ Genomic Med* 2016; **1**: 160271–1602710.
14 C Yuen RK, Merico D, Bookman M, L Howe J, Thiruvahindrapuram B, Patel RV *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* 2017; **20**: 602–611.
15 Ku C, Polychronakos C, Tan E, Naidoo N, Pawitan Y, Roukos D *et al.* A new paradigm emerges from the study of *de novo* mutations in the context of neurodevelopmental disease. *Mol Psychiatry* 2013; **18**: 141–153.
16 Krumm N, O'Roak BJ, Shendure J, Eichler EE. A *de novo* convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014; **37**: 95–105.
17 McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y *et al. De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol Psychiatry* 2014; **19**: 652–658.
18 Kataoka M, Matoba N, Sawada T, Kazuno AA, Ishiwata M, Fujii K *et al.* Exome sequencing for bipolar disorder points to roles of *de novo* loss-of-function and protein-altering mutations. *Mol Psychiatry* 2016; **21**: 885–893.
19 Willsey AJ, Sanders SJ, Li MF, Dong S, Tebbenkamp AT, Muhle RA *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 2013; **155**: 997–1007.
20 Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 2013; **155**: 1008–1021.
21 Cristino AS, Williams SM, Hawi Z, An JY, Bellgrove MA, Schwartz CE *et al.* Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol Psychiatry* 2014; **19**: 294–301.
22 Sanders SJ, Xin H, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 2015; **87**: 1215–1233.
23 Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Mol Psychiatry* 2013; **18**: 1054–1056.
24 Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D. Genotype to phenotype relationships in autism spectrum disorders. *Nat Neurosci* 2015; **18**: 191–198.
25 Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci* 2016; **19**: 1454–1462.
26 O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.

27 O'Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C *et al*. Recurrent *de novo* mutations implicate novel genes underlying simplex autism risk. *Nat Commun* 2014; **5**: 5595.

28 Wang T, Guo H, Bo X, Stessman H, Wu H, Coe BP *et al*. *De novo* genic mutations among a Chinese autism spectrum disorder cohort. *Nat Commun* 2016; **7**: 13316.

29 Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M *et al*. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet* 2017; **49**: 515–526.

30 van Bon BW, Coe BP, Bernier R, Green C, Gerdts J, Witherspoon K *et al*. Disruptive *de novo* mutations of DYRK1A lead to a syndromic form of autism and ID. *Mol psychiatry* 2016; **21**: 126–132.

31 Fombonne E, Roge B, Claverie J, Courty S, Fremolle J. Microcephaly and macrocephaly in autism. *J Autism Develop Disorders* 1999; **29**: 113–119.

32 Sacco R, Gabriele S, Persico AM. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. *Psychiatry Res* 2015; **234**: 239–251.

33 Thompson BA, Tremblay V, Lin G, Bochar DA. CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* 2008; **28**: 3894–3904.

34 Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA *et al*. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013; **4**: 36.

35 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.

36 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

37 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.

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

39 Li J, Cai T, Jiang Y, Chen H, He X, Chen C *et al*. Genes with *de novo* mutations are shared by four neuropsychiatric disorders discovered from NP *denovo* database. *Mol Psychiatry* 2016; **21**: 290–297.

40 Li J, Jiang Y, Wang T, Chen H, Xie Q, Shao Q *et al*. mirTrios: an integrated pipeline for detection of *de novo* and rare inherited mutations from trios-based next-generation sequencing. *J Med Genet* 2015; **52**: 275–281.

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

42 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.

43 Cai T, Yang L, Cai W, Guo S, Yu P, Li J *et al*. Dysplastic spondylolysis is caused by mutations in the diastrophic dysplasia sulfate transporter gene. *Proc Natl Acad Sci USA* 2015; **112**: 8064–8069.

44 Yu P, Yang W, Han D, Wang X, Guo S, Li J *et al*. Mutations in WNT10B are identified in individuals with oligodontia. *Am J Hum Genet* 2016; **99**: 195–201.

45 Guo S, Yang L, Liu H, Chen W, Li J, Yu P *et al*. Identification of novel compound mutations in PLA2G6-Associated neurodegeneration patient with characteristic MRI imaging. *Mol Neurobiol* 2016.

46 Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K *et al*. Excess of rare, inherited truncating mutations in autism. *Nat Genet* 2015; **47**: 582–588.

47 Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S *et al*. Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013; **154**: 518–529.

48 Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM *et al*. A framework for the interpretation of *de novo* mutation in human disease. *Nat Genet* 2014; **46**: 944–950.

49 Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu XM, Li MF *et al*. Spatio-temporal transcriptome of the human brain. *Nature* 2011; **478**: 483–489.

50 Li TB, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G *et al*. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* 2017; **14**: 61–64.

51 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559.

52 Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A *et al*. Transcriptional landscape of the prenatal human brain. *Nature* 2014; **508**: 199–206.

53 Magnani D, Morle L, Hasenpusch-Theil K, Paschaki M, Jacoby M, Schurmans S *et al*. The ciliogenic transcription factor Rfx3 is required for the formation of the thalamocortical tract by regulating the patterning of prethalamus and ventral telencephalon. *Human Mol Genetics* 2015; **24**: 2578–2593.

54 Low LK, Liu XB, Faulkner RL, Coble J, Cheng HJ. Plexin signaling selectively regulates the stereotyped pruning of corticospinal axons from visual cortex. *Proc Natl Acad Sci USA* 2008; **105**: 8136–8141.

55 Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. *PloS Genet* 2010; **6**: e1001154.

56 Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PloS Genet* 2013; **9**: e1003709.

57 Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG *et al*. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 2013; **77**: 235–242.

58 Toma C, Torrico B, Hervas A, Valdes-Mas R, Tristan-Noguero A, Padillo V *et al*. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. *Mol Psychiatry* 2014; **19**: 784–790.

59 Hoischen A, Krumm N, Eichler EE. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* 2014; **17**: 764–772.

60 Liu L, Sabo A, Neale BM, Nagaswamy U, Stevens C, Lim E *et al*. Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PloS Genet* 2013; **9**: e1003443.

61 Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. *Trends Genet* 2013; **29**: 575–584.

62 Durak O, de Anda FC, Singh KK, Leussis MP, Petryshen TL, Sklar P *et al*. Ankyrin-G regulates neurogenesis and Wnt signaling by altering the subcellular localization of beta-catenin. *Mol Psychiatry* 2015; **20**: 388–397.

63 Turner TN, Sharma K, Oh EC, Liu YFP, Collins RL, Sosa MX *et al*. Loss of delta-catenin function in severe autism. *Nature* 2015; **520**: 51–56.

64 Pilorge M, Fassier C, Le Corronc H, Potey A, Bai J, De Gois S *et al*. Genetic and functional analyses demonstrate a role for abnormal glycinergic signaling in autism. *Mol Psychiatry* 2016; **21**: 936–945.

65 Tilot AK, Bebek G, Niazi F, Altemus JB, Romigh T, Frazier TW *et al*. Neural transcriptome of constitutional Pten dysfunction in mice and its relevance to human idiopathic autism spectrum disorder. *Mol Psychiatry* 2016; **21**: 118–125.

66 Frazier TW, Embacher R, Tilot AK, Koenig K, Mester J, Eng C. Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism. *Mol Psychiatry* 2015; **20**: 1132–1138.

67 Sowers LP, Loo L, Wu Y, Campbell E, Ulrich JD, Wu S *et al*. Disruption of the non-canonical Wnt gene PRICKLE2 leads to autism-like behaviors with evidence for hippocampal synaptic dysfunction. *Mol Psychiatry* 2013; **18**: 1077–1089.

68 Durand CM, Perroy J, Loll F, Perrais D, Fagni L, Bourgeron T *et al*. SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. *Mol Psychiatry* 2012; **17**: 71–84.

69 Hamilton PJ, Campbell NG, Sharma S, Erreger K, Hansen FH, Saunders C *et al*. *De novo* mutation in the dopamine transporter gene associates dopamine dysfunction with autism spectrum disorder. *Mol Psychiatry* 2013; **18**: 1315–1323.

70 Griesi-Oliveira K, Acab A, Gupta AR, Sunaga DY, Chailangkarn T, Nicol X *et al*. Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons. *Mol Psychiatry* 2015; **20**: 1350–1365.

71 Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. *Cell* 2014; **156**: 872–877.

72 Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O *et al*. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 2014; **158**: 263–276.

73 Niehrs C, Acebron SP. Mitotic and mitogenic Wnt signalling. *EMBO J* 2012; **31**: 2705–2713.

74 Durak O, Gao F, Kaeser-Woo YJ, Rueda R, Martorell AJ, Nott A *et al*. Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat Neurosci* 2016; **19**: 1477–1488.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)