Identification of primary copy number variations reveal enrichment of Calcium, and MAPK pathways sensitizing secondary sites for autism

Swati Agarwala, Avinash M. Veerappa and Nallur B. Ramachandra[*]

Abstract

Background: Autism is a neurodevelopmental condition with genetic heterogeneity. It is characterized by difficulties in reciprocal social interactions with strong repetitive behaviors and stereotyped interests. Copy number variations (CNVs) are genomic structural variations altering the genomic structure either by duplication or deletion. De novo or inherited CNVs are found in 5–10% of autistic subjects with a size range of few kilobases to several megabases. CNVs predispose humans to various diseases by altering gene regulation, generation of chimeric genes, and disruption of the coding region or through position effect. Although, CNVs are not the initiating event in pathogenesis; additional preceding mutations might be essential for disease manifestation. The present study is aimed to identify the primary CNVs responsible for autism susceptibility in healthy cohorts to sensitize secondary-hits. In the current investigation, primary-hit autism gene CNVs are characterized in 1715 healthy cohorts of varying ethnicities across 12 populations using Affymetrix high-resolution array study. Thirty-eight individuals from twelve families residing in Karnataka, India, with the age group of 13–73 years are included for the comparative CNV analysis. The findings are validated against global 179 autism whole-exome sequence datasets derived from Simons Simplex Collection. These datasets are deposited at the Simons Foundation Autism Research Initiative (SFARI) database.

Results: The study revealed that 34.8% of the subjects carried 2% primary-hit CNV burden with 73 singleton-autism genes in different clusters. Of these, three conserved CNV breakpoints were identified with *ARHGAP11B*, *DUSP22*, and *CHRNA7* as the target genes across 12 populations. Enrichment analysis of the population-specific autism genes revealed two signaling pathways—calcium and mitogen-activated protein kinases (MAPK) in the CNV identified regions. These impaired pathways affected the downstream cascades of neuronal function and physiology, leading to autism behavior. The pathway analysis of enriched genes unravelled complex protein interaction networks, which sensitized secondary sites for autism. Further, the identification of miRNA targets associated with autism gene CNVs added severity to the condition.

Conclusion: These findings contribute to an atlas of primary-hit genes to detect autism susceptibility in healthy cohorts, indicating their impact on secondary sites for manifestation.

Keywords: Autism subgenome, CNVs, Autism-IPA, Autism pathway analysis, Primary-hit

* Correspondence: nbruom@gmail.com
Department of Studies in Genetics and Genomics, Genetics and Genomics
Lab, University of Mysore, Mysore, Karnataka, India

Background

Autism is a genetic and neurodevelopmental condition with difficulties in reciprocal social interactions, abnormalities in verbal and nonverbal communication, strong repetitive behaviors, and stereotyped interests [1]. The most exclusive autism comorbidities are hypersensitivity, mood swings, impulsivity, agitation, and impairment in cognitive functions at different levels [2, 3]. The prevalence of these deficits in one or more functional domains result in autism onset, mostly before the age of three [3, 4]. Various studies have been conducted on autism starting from linkage studies, genome-wide association studies, single-nucleotide polymorphism (SNP) genotyping to present-day next-generation sequence analysis. In addition to these approaches, copy number variation (CNV) is one of the most promising studies, which adds another dimension to autism research. CNVs refer to the genomic structural variations with more than 1000 bases to many million bases in terms of size with alteration to the gene dosage. These variations can cause functional loss by disrupting regulatory elements, generating fusion proteins, or through position effect variegation. CNV occurrence can be limited to a single gene or a contiguous set of genes in a dosage-sensitive nature. Hence, the presence of these CNVs in genes can contribute to human phenotypic variability, complex behavioral traits, and disease susceptibility [5].

Various studies have addressed the impact of CNVs on autism. The first familial CNV-based study in autism identified de novo CNVs in 10% of the cases [6]. The conceptualization of CNV studies on autism has identified significant common and rare variants. These variants conferred differential effects on autism risk in the general population [7, 8]. Seventeen different loci, localized across 11 chromosomes, proposed a multigene model for CNV pathogenesis [6, 9–17]. Rare CNVs resulted in increased risk for autism by up to a 20-fold increment [17]. More than 40 recurrent autism CNVs have been identified [18]. CNV correlation has been established for multiple loci with significant autism genes namely *SHANK2, SHANK3, NRXN1, NLGN4, PCDH10, DIA1, NHE926*, and *PARK2* [19]. Notably, 1q21.1 and 16p13.11 duplication/deletions, 15q11–q13 duplication, and 16p11.2 deletion have been the important contributors for recurrent autism CNVs [20].

CNV analysis in healthy cohorts acts as a frontier for disease susceptibility, which is evident in various research studies conducted over the last two decades. These findings highlight the role of the CNV burden in healthy groups with added contributions from other factors. It aids in the development of biomarkers for the diagnosis and prognosis of neurodevelopmental phenotypes [21, 22]. Besides, various researchers have reported a significantly higher burden of rare CNVs involving functional genes in diseases [7, 8, 17, 23]. It is hypothesized that the CNVs might not be the initiating event in the pathogenesis, and additional preceding mutations may be necessary to induce the condition [23]. A study conducted by Girirajan et al. [23] has put forth a two-hit model for disease manifestation with two promising findings. Affected individuals with a microdeletion on chromosome 16p12.1 are more likely to have additional significant CNVs than healthy individuals. The second finding of 16p12.1 (CDR2) deletion, 3q29 (DLG1) duplication, and rare copy number variants in affected individuals indicates an association of CNVs with the occurrence of severe intellectual disability and neuropsychiatric diseases due to a variable set of outcomes [23]. These outcomes are described as primary and secondary hits in the two-hit model. Susceptible gene variants present in a healthy individual, which predisposes them to disease susceptibility, are described as primary-hit. These primary-hits might/might not result in disease manifestation. For a definite progression of the disease, the occurrence of another gene variant in the individual is necessary, which are stated as secondary-hits. Combinatorial effect of a primary-hit and secondary-hit in an individual results in disease development [7]. This two-hit model could be applied to autism as well, owing to its genetic heterogeneity.

The current investigation is aimed at identifying primary autism gene CNVs in the healthy cohorts. The identification of the primary-hit CNVs in the healthy cohorts would help in uncovering the autism susceptibility loci. These primary-hit CNVs would act as molecular biomarkers for recognition of secondary-hits to minimize disease progression.

Methods

The study included a sample cohort of 1715 normal healthy individuals belonging to different ethnicities. Firstly, it included 270 HapMap samples with 30 both-parent-and-adult-child trios from Yoruba people in Ibadan, Nigeria (HapMap YRI) as well as CEPH/Utah Collection (HapMap CEU), and 45 unrelated HapMap individuals, from Tokyo Japan (HapMap) as well as Han Chinese in Beijing Japanese (HapMap CHB) populations [24]. Secondly, 155 Chinese and an equal number of 472 each, from Ashkenazi Jews replicate 1 (AJI), as well as Ashkenazi Jews, replicate 2 (AJII) populations were selected. Thirdly, 184 individuals from Taiwan, 41 from the New World population (Totonacs and Bolivians), 53 from Australia, and 31 Tibetan samples were recruited [25]. These sample datasets were obtained from the Array Express Archive of the European Bioinformatics Institute. Case registries were referred for the exclusion of subjects, wherein the samples with pre-diagnosed autism and autistic symptoms were excluded.

Thirty-eight individuals from twelve families residing in Karnataka, India, with an age group of 13–73 years

were selected for the comparative CNV analysis. Ethical approval was obtained by the Institutional Human Ethics Committee (IHEC) of the University of Mysore, Karnataka, India. Written informed consent was obtained from each subject as per the IHEC approved procedure. Informed consent for minor subjects was obtained from guardians/parents.

Five milliliter of blood was collected in K^{2+} EDTA vacutainer tubes from the Indian study group. Genomic DNA extraction was carried out using the Promega Wizard® Genomic DNA purification kit. Visualization of isolated and quantified DNA was performed using bio-photometer and gel electrophoresis.

Genome-wide genotyping was performed using the Affymetrix Genomewide Human SNP Array 6.0 chip and Affymetrix CytoScan High-Density array. The array contained 1.8 million SNP and 2.6 million CNV markers with a median inter-marker distance of 500–600 bases. These array-based studies provided the highest physical coverage and maximum panel power for the genome.

BirdSuite algorithm (https://www.broadinstitute.org/birdsuite/birdsuite-analysis) was implemented to detect commonly known copy number polymorphisms (CNPs) based on curated literature. It detected rare and common CNVs using the hidden Markov model (HMM) algorithm from Affymetrix SNP 6.0 array data. For the HMM algorithm, the hidden state mapped a specific individual to its genomic copy number. The observed states indicated the normalized intensity measurements for each array probe. This approach identified the sample-specific variable copy number regions. Collation of sample-wise CNV calls was performed from Canary and BirdSuite algorithms using the outputs from the previous step. The selection criteria were for filtering the obtained CNV calls was postulated. This criteria suggest to include BirdSuite CNV calls with a log10 of odds score (Odds Ratio) ≥ 10 for an approximate false discovery rate (FDR) of ~ 5% for further analysis. For copy number (CN) states, all calls to be included except for those with CN state as 2 and differential CNP calls with CN states, in comparison to the population model.

Classification of copy number changes was performed using CNV Finder of Welcome Trust Sanger Institute with a varying quality score in the provided data. This method was based on two assumptions: firstly, the majority of data points were normalized around a log2 ratio of zero, and secondly, the data points localized outside of centralized log2 ratio distribution, indicative of a difference in the CN between reference and test genome.

CNP analysis was performed to obtain CN state calls in genomic regions using the Canary algorithm. Computation of single intensity summary statistics within the CNP region was completed manually using selected probes. An aggregative comparison of these intensity summaries has been used to assign individual CN state call across all samples, compared to those previously observed in training data.

Genotyping console selected quality control (QC) passed samples in CEL file format to call genotypes using the Birdseed algorithm. It detected CNVs with a threshold parameter of > 1 kb size and > 5 probes.

Genome-wide CNV study was carried out using Affymetrix Genotyping Console software as per standard protocol. The results were visualized using SVS Golden Helix Version 7. After employing Bonferroni correction for multiple testing, the corrected data output was used for CNV testing. For population-wise genotyped data, the threshold for the Bonferroni method was set between 1×10^{-7} and 7×10^{-8} for α = 0.05 on the Affymetrix 6.0 platform.

The stringency of CNP calls was met with a log10 of odds score ≥ 10 and FDR of 5%. These values corresponded to collated data output obtained from BirdSuite and Canary algorithms. All the called-SNPs with a QC call rate of > 97% were entered into the CNV analysis across subjects. Filters on call rates were used to identify call rates obtained from poor quality DNA for the overall SNPs. In the present study, contrast QC of > 2.5 with robust strength was observed across all samples. To control the possibility of spurious or artifact CNVs, the Eigenstrat approach of Price et al. was referred [26]. The principal components of the correlations among gene variants were obtained and accordingly corrected. Fifty-five individuals were extreme outliers with ≥ 1 significant Eigenstrat axes. These were excluded from the study group. Failure to meet the stipulated QC threshold resulted in the dropping of 543 CNVs in the selected individuals. Validation of CNVs was established based on ≥ 50% reciprocal overlapping with the reference set. Relative values between the comparisons of algorithms/platforms/sites were quite informative, even though the sensitivity of Jaccard statistics to the CNVs calls by each algorithm was considered. All the overlap analyses performed, handled losses and gains separately except when otherwise stated and conducted hierarchically. The algorithmic calls, called in both Canary and Birdsuite, were not considered; instead, they were collated for informative relative values between the different comparisons in terms of algorithms/platform/sites.

The reference autism gene list was prepared using two-point approaches. It was performed through an extensive well-defined PUBMED literature search matrix and SFARI database, based on inclusion-exclusion gene selection criteria. Inclusion-exclusion of gene selection was included in the criteria such as: should be an autism candidate gene expressed in brain; participate in neuronal development; interact with known autism genes; non-homozygous in controls; de novo in origin; overlap

in two or more unrelated samples; recurrent in two or more unrelated samples; and involve in the expression of brain development and participate in neuronal migration, axon growth, neuritis outgrowth, synaptic plasticity, and cell adhesion (Fig. 1). Associated genes and genes with lower significance in terms of the *p* value, pathogenicity scoring, number of studies performed, and those without validations were excluded from the final gene list. The gene list was used for the overall analysis, and the CNVs were accordingly filtered. Consistently replicated genes found across populations were selected. The shared map of autism genes under CNVs was generated for all chromosomes using the Circos software package.

Function-based gene categorization for the identified CNV autism genes was performed following GO classification: Biological process, cellular component, and molecular function using WEB-based Gene Set Analysis Toolkit (WEBGASTALT). Multiple-test adjustment was applied using the hypergeometric statistical method following the Benjamini-Hochberg procedure. The significance level and *p* value with FDR were calculated for the top seven genes. KEGG pathways were identified using classified genes based on two criteria. It included the pathway associations and quantification of genes in each pathway with p-values and its enrichment significance. In each generated pathway map, genes from the gene list were highlighted in red.

The pathways and molecular interactions were generated through the Ingenuity Pathway Analysis (www. ingenuity. com). IPA was used to identify the interaction between genes, protein-protein interactions, biological mechanisms, location, and target gene functionality. Genes and the chemical-based search were used to explore the information on protein families, protein signaling, normal cellular protein activity, and associated metabolic pathways. Localized genes and their protein products have been interconnected through edges. An edge (line) represented the relationship between two nodes. Each network edge was described using a knowledge base of pathways and the curated literature available within IPA software. The cascades of protein-protein interaction, protein binding, activation, upregulation, downregulation, and mRNA expression by targeting a mature miRNA network were observed in pathway enrichment [27].

The validation of the four recurrent CNV breakpoints was performed by amplification using polymerase chain reaction (PCR) in our laboratory and published elsewhere [28].

Results

Of the total of 1715 normal healthy subjects, 34.8% of the individuals showed significant CNVs in the autism-specific subgenome. These CNVs were ranged between 8.88 and 49.05% (Fig. 2a). The highest and lowest CNV frequencies of 49% and 8% were identified in Australia and HapMap YRI respectively (Table 1). This covered 2% autism gene-specific CNV burden across the 12 populations under study. CNVs in autism genes were seen in all the chromosomes except chromosomes 4, 14, 22, and Y (Fig. 2a). CNVs present in these chromosomes were identified by 90716 SNP and CNV combined markers with an average size and count of 261.35 kb and
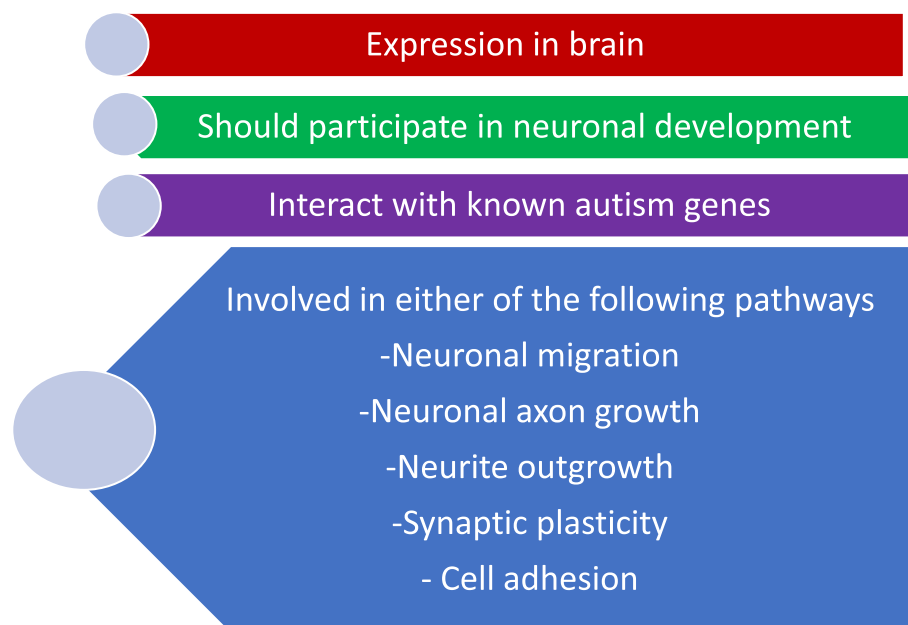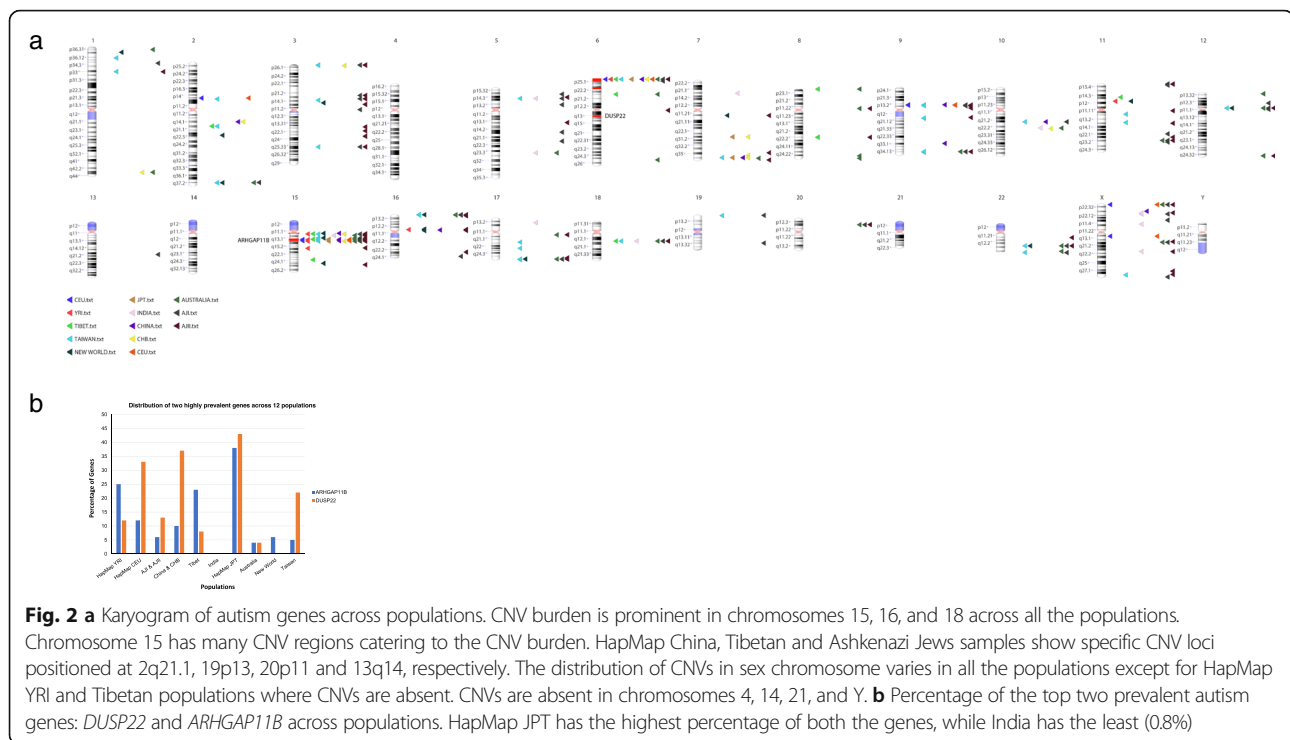

Fig. 1 Inclusion and exclusion criteria for selection of autism genes for downstream analysis with 0.05 as the *P* value

Fig. 2 a Karyogram of autism genes across populations. CNV burden is prominent in chromosomes 15, 16, and 18 across all the populations. Chromosome 15 has many CNV regions catering to the CNV burden. HapMap China, Tibetan and Ashkenazi Jews samples show specific CNV loci positioned at 2q21.1, 19p13, 20p11 and 13q14, respectively. The distribution of CNVs in sex chromosome varies in all the populations except for HapMap YRI and Tibetan populations where CNVs are absent. CNVs are absent in chromosomes 4, 14, 21, and Y. **b** Percentage of the top two prevalent autism genes: *DUSP22* and *ARHGAP11B* across populations. HapMap JPT has the highest percentage of both the genes, while India has the least (0.8%)

148.60 kb for autism CNV burden respectively (Supplementary Table 1, Supplementary Figure 1). Of the 2% CNV burden, duplication CNVs (73%) were predominant over the deletion CNVs (27%) (Table 1).

The 2% CNV autism gene landscape contained 73 singleton autism genes [54–117] (Supplementary Table 2). The notable causal autism genes mapped for these CNV regions were *ARHGAP11B, DUSP22, CHRNA7, CYFIP1, NIPA1, TUBGCP5, CACNA1H,* and *CELF4.* The frequency of these genes was highest in *ARHGAP11B,* followed by *DUSP22* and *CHRNA7* in the CNVs regions (Fig. 2b).

Two prominent findings were comprehended for two autism causal gene clusters—*CYFIP1, NIPA1, TUBGCP5* and *ARHGAP11B, DUSP22, CHRNA7.* These clusters were present in conserved CNV regions in multiple loci across many populations under study. The cluster containing *CYFIP1, NIPA1,* and *TUBGCP5* was under recurrent CNV events on 15q11.2 loci. The other cluster with *ARHGAP11B, DUSP22,* and *CHRNA7* showed significantly conserved CNV breakpoints across many populations under study. For example, *DUSP22* showed one start and two end breakpoints on the chromosome 6. The start breakpoint "257341" was present in nine populations and two end breakpoints, "381131" and "382897", were present in six populations under study. *DUSP22* and *ARHGAP11B* genes were represented in 5.65% and 2.85% of the identified CNVs, respectively. These were marked as highly recurrent genes for autism CNV burden.

All these autism genes were under the influence of CNVs with varying CN states (Supplementary Figure 2a). CN state can be studied to estimate the expression levels of proteins. Hence, baseline brain expression level in silico analysis was performed for *DUSP22* and *ARHGAP11B. ARHGAP11B* showed an expression level of one transcript (ENST00000602616) out of nine with a cut-off value of 0.4. Similarly, *DUSP22* showed a dosage level of 13 with an expression of two transcripts (ENST00000419235 and ENST00000344450) out of 16 (Table 2). *DUSP22* was under tremendous CNV burden with a frequency range of 0.09–0.16% under 70 autism genes-CNV breakpoints, both within and across populations under study. Pair-wise clustering of shared autism genes (in %) across all chromosomes and 12 populations are presented in the Circos image (Supplementary Figure 2b).

A total of 146 CNVs bearing 40 autism causal genes were limited to distinct populations with a frequency of 0.25–2.73%. There were 18 genes in 113 CNVs–AJI and AJII samples, five genes in 6 CNVs–Taiwan samples, three genes in 6 CNVs–Australia samples, three genes in 3 CNVs–Indian samples, and two genes in 2 CNVs–Tibetan samples. Similarly, CNV burden of one gene in 1 CNV–HapMap CEU, two genes in 5 CNVs–New World, four genes in 7 CNVs–China and CHB, and one gene in 3 CNVs–HapMap YRI were seen. HapMap JPT did not show any specific gene under the CNV burden. Overall, sex bias was absent for the CNV burden with CNV distribution of 51.36% in males and 48.47% in females. However, negligible male and female biases were

Table 1 Distribution of autism CNV duplication and deletion regions present across 12 populations

Sl. No	Population	Total individuals analyzed	No. of individuals with CNVs (%)	Overall autism-specific CNVs (%)				CNVs in autism genes (%)			
				Male		Female		Male		Female	
				Duplication	Deletion	Duplication	Deletion	Duplication	Deletion	Duplication	Deletion
1	HapMap YRI	90	9	42.86	57.14	100	0	50	50	100	0
2	CEU	90	9	25	75	57.14	42.86	33.33	66.67	50	50
3	Ashkenazi Jews I and II	944	44	20.83	79.17	22.5	74	68.18	31.82	76.47	23.53
4	China and CHB	199	15	43.75	56.25	50	50	45.45	54.55	77.78	22.22
5	Tibet	31	48	71.43	28.57	73.33	26.67	66.67	33.33	70	30
6	India	38	18	100	0	80	20	100	0	80	20
7	HapMap JPT	45	36	33.33	66.67	31.25	68.75	33.33	66.67	33.33	66.67
8	Australia	53	49	82.5	17.5	90.47	9.52	75	25	85.71	14.29
9	New World	41	41	65.21	34.78	-	-	57	43	-	-
10	Taiwan	184	31	92	8	82.76	17.24	86.67	13.33	87.5	12.5[*]
	Total	1715	35	73.19	26.81	73.66	21.95	67.14	32.86	76.38	23.61

observed in the majority of the populations under study with regards to the percentage of CNVs present.

For further analysis, highly prevalent seven autism gene CNVs in each population ranging from 4 to 53% were chosen. Based on gene ontology (GO) study, they were classified into major categories of biological, functional processes, and location using the WEBGASTALT tool. The encoded proteins from these genes were localized primarily in the intracellular region, followed by organelle lumen and intracellular organelle lumen, while the rest were found in cell projection and cytoskeleton regions. Further, *ARID1B*, *DBH*, *UBE3A*, *CDKL5*, *HLA-DRB1*, and *VPS13B* genes were identified through enrichment analysis of autism-gene-CNVs.

The genes identified through pathway analysis were functional in neurodevelopment, neurotransmission, and synapse formation. These genes were recognized as targets of multiple miRNAs. For instance, targets of *miR-499-3p* were IMM2PL, UBE2G1, CACNA1B, APBA1, and DLGAP2 mRNAs. Similarly, the targets for *miR-513a-5p* included CACNA1B, NIPA1, DUSP22,

KCND2, and CACNA1C mRNAs (Fig. 3). CNVs in autism genes *CACNA1C*, *CACNA1H*, *CACNA1I*, *DUSP22*, and *CHRNA7* were enriched with calcium and MAPK signaling pathways across all populations. However, *CHRNA7* and *CHRM3* were enriched for neuroactive ligand-receptor interactions and present exclusively in the Chinese population (Table 3; Fig. 4).
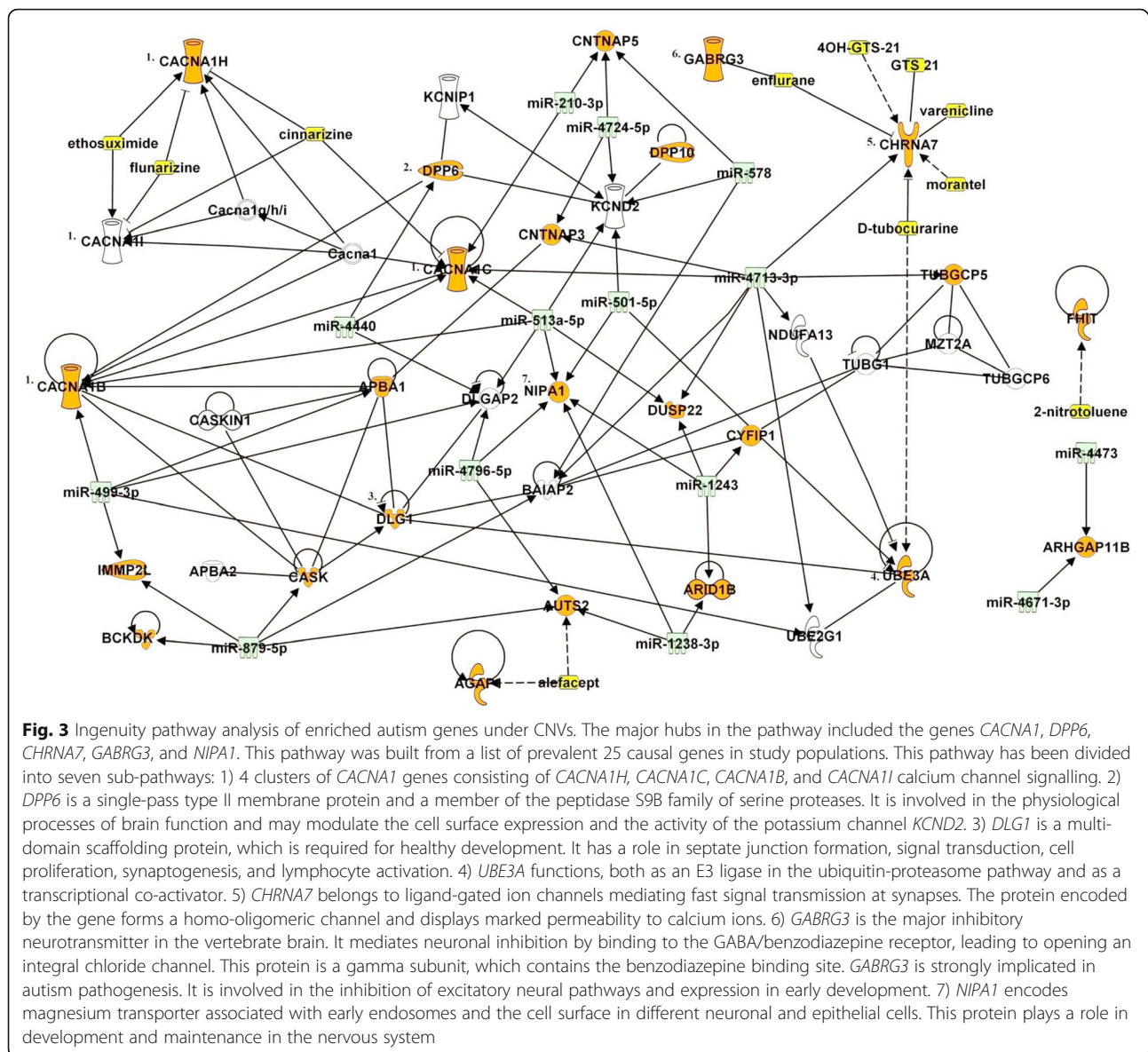
Discussion

CNVs are genomic structural variations that contribute to the disease pathogenesis through gene function disruptions. Several studies on primary CNVs have been indicative of their role in the manifestation of conditions such as asthma, nondisjunction, Parkinson's disease, diabetes, migration, and olfactory receptors [7, 29–33]. CNVs in the form of duplications and deletions manifest the gain and loss of function in a gene [34], which disrupt the protein structure and alter its transcriptional activity in the regulatory regions [35] in the autism subgenome.

Table 2 Baseline expression dosage of ARHGAP11B and DUSP22 in the brain using Human Gene Atlas

Gene with baseline expression dosage	CNV type	CN[a] state	No. of individuals	Entire/Partial	Protein dosage
ARHGAP11B	Deletion	1	9	Partial	0.2
		1	18	Entire	0.2
	Duplication	3	21		0.6
		4	1		0.8
DUSP22	Deletion	1	87	Entire	6.5
	Duplication	3	16		19.5

[a]*CN* copy number, *EBI* European Bioinformatics Institute

Fig. 3 Ingenuity pathway analysis of enriched autism genes under CNVs. The major hubs in the pathway included the genes *CACNA1*, *DPP6*, *CHRNA7*, *GABRG3*, and *NIPA1*. This pathway was built from a list of prevalent 25 causal genes in study populations. This pathway has been divided into seven sub-pathways: 1) 4 clusters of *CACNA1* genes consisting of *CACNA1H, CACNA1C, CACNA1B*, and *CACNA1I* calcium channel signalling. 2) *DPP6* is a single-pass type II membrane protein and a member of the peptidase S9B family of serine proteases. It is involved in the physiological processes of brain function and may modulate the cell surface expression and the activity of the potassium channel *KCND2*. 3) *DLG1* is a multi-domain scaffolding protein, which is required for healthy development. It has a role in septate junction formation, signal transduction, cell proliferation, synaptogenesis, and lymphocyte activation. 4) *UBE3A* functions, both as an E3 ligase in the ubiquitin-proteasome pathway and as a transcriptional co-activator. 5) *CHRNA7* belongs to ligand-gated ion channels mediating fast signal transmission at synapses. The protein encoded by the gene forms a homo-oligomeric channel and displays marked permeability to calcium ions. 6) *GABRG3* is the major inhibitory neurotransmitter in the vertebrate brain. It mediates neuronal inhibition by binding to the GABA/benzodiazepine receptor, leading to opening an integral chloride channel. This protein is a gamma subunit, which contains the benzodiazepine binding site. *GABRG3* is strongly implicated in autism pathogenesis. It is involved in the inhibition of excitatory neural pathways and expression in early development. 7) *NIPA1* encodes magnesium transporter associated with early endosomes and the cell surface in different neuronal and epithelial cells. This protein plays a role in development and maintenance in the nervous system

The present study establishes the autism-CNV atlas, prioritizing autism-specific CNV regions in healthy cohorts. It uncovers the primary-hit CNVs in the autism sub-genome which has been mainly unexplored. A similar trend is reported in the inherited CNVs with *SHAN K2* deletion, mutations with duplication in *CHRNA7*, and deletions in *CYFIP1* loci, which are indicative of putative multi-hit model for autism [16]. A similar trend is consistently identified in the present investigation.

Identified CNVs are present in autism-specific subgenome with a mean average of 34%. Investigation of CNV size has been limited to ≥ 100 kb due to maximum signal to noise ratio for CNVs below 100 kb. The majority of discovered CNVs belong to a size range of 100–500 kb. The frequency of CNV events declined beyond 500 kb. Higher CNV burden is observed in autism-specific

chromosomal regions 6, 15, 16, and 18, following a similar trend as mentioned in Girirajan et al. [23]. The chromosomes with autism genes are more susceptible to CNV accumulation. CNV distribution in terms of size, count, type, and state showed a different percentage for inter and intra populations, consistent with previous studies.

Further, autism gene-CNV duplications outnumber the deletion regions, as evident in AJI, AJII, Australia, and Taiwan. This can be because the genome can withstand duplications better than deletions. Loss of function is more damaging and hence it results in higher dosage and early disease manifestation. These findings are found in accordance with a previous study on autism in the European population and healthy cohorts [23, 32]. HapMap YRI and HapMap CEU contain an equal

Table 3 Pathway enrichment analysis of autism gene CNVs across 12 populations

Population	Pathways	Autism gene-specific CNVs		
Ashkenazi Jews I and II	Calcium signaling	CHRNA7	CACNA1I	CACNA1H
	MAPK signaling	DUSP22		
Taiwan	MAPK signaling	DUSP22	CACNA1H	
New World	Calcium signaling	CACNA1I	CACNA1H	
	MAPK signaling			
Australia	Calcium signaling	CACNA1I	CACNA1H	CACNA1C
	MAPK signaling			
China and CHB	Neuroactive ligand-receptor interaction	CHRNA7	CHRM3	

number of duplications and deletions, suggesting that these population-specific CNVs are random events. Numerous studies advocate a similar balanced contribution from deletion and duplication in these populations [36]. Hence, studies in a larger size cohort would be needed to confirm the findings.

Autism genes with a 2% CNV burden show overlapping mutations for 73 singleton genes with previously reported autism genes. This is based on relevance, research findings in various autism cohort consortiums, and SFARI gene scoring [37, 38] (Supplementary Table 2). Out of these, 14 autism genes have been mapped to SFARI gene scoring 1. These are termed as high confidence genes with clear implications for autism. These are known to have at least three de novo gene disrupting mutations reported with a rigorous threshold FDR of < 0.1. A total of 15 genes are scored 2 and referred as strong candidate genes with two de novo gene disruptive mutations. These have been implicated by genome-wide significance or replicative in multiple studies with strong evidence. Further, 25 autism genes are scored 3 with suggestive evidence. These genes contain single de novo mutations identified from significant and non-replicated association studies. These have been reported through non-association or rare-inherited case studies with no comparative statistical study in controls. Three genes are scored syndromic with the risk of autism susceptibility. An extensive literature study shows evidence for 16 genes with no scoring and confirmed as specific to autism. Hence, these genes confirmed through curated literature have been considered in the singleton gene list (Supplementary Table 2).

The selection of seven genes for downstream analysis is based on relevance to autism and recurrent CNVs present across populations. Two prominent gene clusters are identified. The first imprint gene cluster *CYFIP1*, *NIPA1*, and *TUBGCP5* is associated with changes in brain-behavior, morphology, and cognitive functions, which are key phenotypes in autism [39]. This gene cluster impacts the molecular control of synaptogenesis and neuronal connectivity in a dosage-sensitive manner [40].

Various de novo autism-specific mutations have been reported with this cluster. Further, mutations in this cluster have been marked as recurrent pathogenic CNV regions for neurodevelopmental disorders such as autism. Either side of its flanking regions contains autism genes such as *UBE3A* and *ATP10A* [41].

In the second gene cluster, ARHGAP11B and *DUSP22* are under the influence of CN states 1, 3, 4, and 1, 3, respectively. The data points for CN states 1 and 3 depict a mirror image (when halved). Populations with duplications are on the higher side and deletions on the diametrically opposite lower side. CNVs with *ARHGAP11B* are more frequent in the Tibetan population, resulting in varied protein dosage. Higher CN states (> 2) also alter the expression level of ARHGAP11B, prominent in AJI, AJII, and Taiwan. This is in conjunction with similar CNV studies performed for asthma, nondisjunction, Parkinson's disease, diabetes, and miRNA gene regulation in healthy cohorts [29–33]. Further, multiple mutations in *ARHGAP11B* include the recurrent 15q duplication and point mutations. These mutations result in early truncation and induce the proliferation of basal progenitors in the cranial neocortex [42]. *ARHGAP11B*, in such situations, triggers enhanced brain stem cell formation, which is a prerequisite for enlarged brain. This provides an advantage to *ARHGAP11B* to incur a prominent phenotype of an enlarged brain in autism [42]. *DUSP22* shows recurrent breakpoints on chromosome 6 across populations. The entire protein product is affected by a protein dosage of 19.5 for deletions and duplication variants. Similarly, the *DUSP22* gene results in the formation of excess neurons in the prefrontal brain in autism, which is the warehouse of social, language, and cognitive functions [43]. Thus, the presence of primary-hit CNVs for these genes increases susceptibility toward autism upon secondary-hit, either through point mutations or other gene mutations.

The population-specific CNVs are identified in varying frequencies across the sample cohort. The diversity and
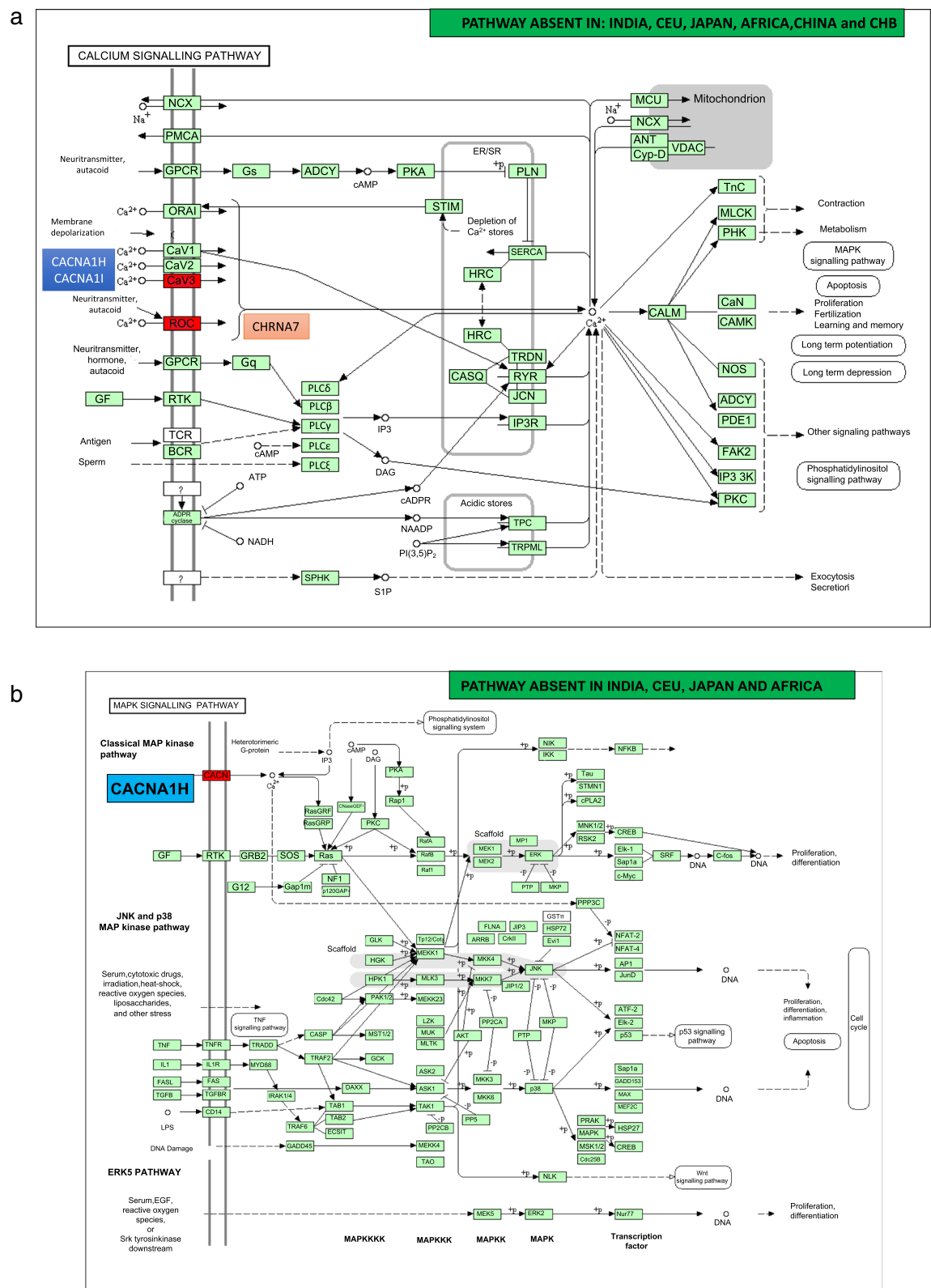
Fig. 4 (See legend on next page.)

exclusivity can be either because of variable sample size or random events. Sex bias interpretation for autism CNV regions could not be conclusive due to limited information. All these CNV genes are autosomal. Hence, it is challenging to infer sex bias-based interpretations of the study. None of the established sex bias genes for autism are identified. Therefore, sex bias is ruled out and considered as balanced across populations.

GO analysis for the highly prevalent seven genes (> 50%) pinpoints relevant autism gene functionality in each population. The majority of the identified autism genes are involved in the regulation of the cellular process, cellular response to organic substances, and regulation of cellular signaling. In biological processes, 90% of the genes are under the cellular process regulation, response to organic substances, and regulation of signaling. Under the molecular function category, genes encoding for cation binding and metal ion binding are significantly high. Neuronal stability and plasticity are regulated by actin and microtubule regulation present in cytoskeletal regions. These play a key role in brain functionality through neurite outgrowth and dendritic spine formation [44, 45]. Out of these enriched genes, *VPS13B* and *ARID1B* contribute to seizures and neurological speech impairments. These are causal for autism and result in its early onset. One recent study has identified an intragenic and multiexonic deletion in the *VPS13B* gene [46]. The subsequent gene product impairs the adaptive functionality, resulting in autism on partial inactivation [46]. *ARID1B* is a gene with high statistical significance and an FDR value of 0.01. It has strong gene-based de novo mutational evidence for autism with absence or low mutational frequency in controls [47, 48].

Seven autism CNV genes are enriched for various pathways such as calcium signaling, MAPK signaling, and neuroactive ligand-receptor interactions. The autism-specific genes for calcium signaling pathway—*CHRNA7*, *CACNA1H*, *CACNA1C*, and *CACNA1I*—are enriched across all 12 populations. The influx of Ca^{2+} from the environment or release from internal stores causes a rapid increase in cytoplasmic calcium concentration. This dysregulated modulation of $Ca2^+$ concentration results in impaired neuronal function leading to autism [49]. Products of *CHRNA7* and *CACNA1H* are neurotransmitters and voltage channels responsible for the influx of Ca^{2+}. These are involved in the regulation

of the downstream cascade of reactions in the cellular pathways [50, 51]. *CHRNA7* microduplication has been detected in a subject with autism and moderate cognitive impairment [46]. Mutations in these genes impair the protein product formed, which in turn affects various downstream signaling pathways. *CACNA1I*, *DUSP22*, *CACNA1C*, and *CACNA1H* are known to regulate the MAPK signaling pathway [52]. *CACNA1H* and *CACNA1I* contribute to CNV burden in most populations, while those for *CACNA1C* and *DUSP22* are confined to a few populations [53]. These genes are expressed in four distinct MAPK groups; extracellular signal-related kinases 1/2, Jun amino-terminal kinases 1/2/3, p38 proteins, and ERK5. These genes are involved in various cellular functions such as cell proliferation, differentiation, and migration.

The establishment of the enrichment pathway for autism-gene CNVs has identified significant genes for autism pathogenesis initiation. The minimal cut sets are computed for physical and genetic interactions. As a result, the experimental block of essential genes inevitably leads to mutants. These sets include *CASKIN1*, *KCNIP1*, and *KCND2* genes. These are closely linked to known autism causal genes and might be reliable indicators as autism candidate genes.

The primary-hit autism gene CNVs identified in 1715 individuals are cross-analyzed against 179 autism whole-exome sequence datasets with identification of overlapping regions for *CHRNA7* and *CYFIP1*. The co-occurrence of a loss of one copy of *SHANK2* and *CYFIP1* increases the risk of abnormal synaptic function in autistic subjects [16]. These autism genes CNVs contain 24 autism risk genes, resulting in autism manifestation, not found in a healthy cohort. Therefore, it can be inferred that the secondary-hit by the autism risk genes results in autism manifestation in autism cases, unlike the unaffected healthy cohorts, which escape contracting the condition.

Conclusion
Identification of recurrent CNVs in the healthy cohorts provides another dimension to assess the role of primary-hits toward the sensitization of the secondary-hits for the manifestation of autism. These primary-hits are vulnerable randomly, causing disease pathogenesis upon secondary-hits. Therefore, understanding susceptible loci in a healthy cohort would help in identifying

the soft spots to avoid increasing the probability of autism manifestation.

Limitation of the study

A detailed study in larger cohorts must be warranted to identify ethnicity-specific markers. Overlapping studies could be performed on similar datasets on other platforms like next-generation sequencing, if possible, for in silico validations.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s43042-020-00091-3.

Additional file 1: Supplementary Figure 1. Overview of a). CNV distribution in terms of count. The counts deviated significantly across population and revealed contrasting contributions towards the CNV. Chromosome 6, 15 and 16 showed higher CNV count burden, consistent with previous autism studies b). CNV distribution in terms of size. It has an average size of 2319 kbs across populations. The CNV size corresponds to the CNV count burden to chromosomes 6, 15 and 16. (c) Percentage of genes under CNV burden across populations.

Additional file 2: Supplementary Figure 2. (a) Copy number states of autism genes across populations. The data points for CN states 1 and 3 depicts a mirror image (when halved), with populations showing duplications on the higher side and lower deletions on the diametrically opposite side for all populations. (b) The pairwise clustering of shared autism genes in percent across all chromosomes for populations under study. *DUSP22* is identified to be under a tremendous CNV burden across all populations. *DUSP22* is present in varying frequency (0.09-0.16%) under several autism genes-CNV breakpoints, both within and across populations.

Additional file 3: Supplementary Table 1. Chromosome wise enrichment of autism gene CNV count (in percentage) across 12 populations

Additional file 4: Supplementary Table 2. Singleton genes identified under CNVs with overlapping CNV studies and functional relevance to autism.

Abbreviations

AJI: Ashkenazi Jews replicate 1; AJII: Ashkenazi Jews replicate 2; CEU: CEPH/ Utah Collection HapMap; CHB: Han Chinese in Beijing Japanese HapMap; CNP: Copy number polymorphism; CNVs: Copy number variations; FDR: False discovery rate; GO: Gene ontology; HapMap JPT: HapMap Japan; HapMap YRI: HapMap Yoruba; HMM: Hidden Markov model; IHEC: Institutional Human Ethics Committee; IPA: Ingenuity Pathway Analysis; Kilobases: Kb; MAPK: Mitogen-activated protein kinase; PCR: Polymerase chain reaction; QC: Quality control; SFARI: Simons Foundation Autism Research Initiative; SNP: Single-nucleotide polymorphism; WEBGESTALT: WEB-based Gene Set Analysis Toolkit

Authors' contributions

All authors have read and approved the manuscript. SA have made substantial contributions to conception and design, and analysis and interpretation of data; involved in drafting the manuscript and reviewing the manuscript critically. AMV made substantial contributions to conception and design, acquisition of data and reviewing the manuscript critically. NBR have made substantial contributions to conception and design; revised the manuscript critically for important intellectual content and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

Written consent was obtained from all participants involved in this study and the Institutional Human Ethical Committee (IHEC No.3/RI/2008–09) approved the consent procedure.

Consent for publication

Consent has been taken from all participants, the data provider, and the authors for the said publication.

Competing interests

The authors declare that they have no competing interests.

References

1. Ousley O, Cermak T (2014) Autism spectrum disorder: defining dimensions and subgroups. Curr Dev Disord Reports. https://doi.org/10.1007/s40474-013-0003-1
2. Cordeiro Q, Vallada H (2005) Genetics of autism [5]. Rev Bras Psiquiatr 27: 257. https://doi.org/10.1590/S1516-44462005000300022
3. Tye C, Runicles AK, Whitehouse AJO, Alvares GA (2018) Characterizing the interplay between autism spectrum disorder and comorbid medical conditions: An integrative review. Front Psychiatry. https://doi.org/10.3389/fpsyt.2018.00751
4. Cook EH Jr (2001) Genetics of autism. [Review] [128 refs]. Child Adolesc Psychiatr Clin North Am
5. Sener EF (2014) Association of copy number variations in autism spectrum disorders: a systematic review. Chinese J Biol. https://doi.org/10.1155/2014/713109
6. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T et al (2007) Strong association of de novo copy number mutations with autism. Science (80-). https://doi.org/10.1126/science.1138659
7. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH et al (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet. https://doi.org/10.1371/journal.pgen.1002334
8. Marshall CR, Scherer SW (2012) Detection and characterization of copy number variation in autism spectrum disorder. Methods Mol Biol. https://doi.org/10.1007/978-1-61779-507-7_5
9. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S et al (2006) Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J Med Genet. https://doi.org/10.1136/jmg.2006.043166
10. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ et al (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. https://doi.org/10.1038/ng1985
11. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J et al (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2007.12.009
12. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R et al (2008) Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. https://doi.org/10.1056/NEJMoa075974

13. Kumar RA, Karamohamed S, Sudi J, Conrad DF, Brune C, Badner JA et al (2008) Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet 17:628–638. https://doi.org/10.1093/hmg/ddm376

14. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S et al (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry. https://doi.org/10.1016/j.biopsych.2008.01.009

15. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS et al (2008) Identifying autism loci and genes by tracing recent shared ancestry. Science (80-). https://doi.org/10.1126/science.1157657

16. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G et al (2012) Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet. https://doi.org/10.1371/journal.pgen.1002521

17. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R et al (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature. https://doi.org/10.1038/nature09146

18. Takumi T, Tamada K (2018) CNV biology in neurodevelopmental disorders. Curr Opin Neurobiol. https://doi.org/10.1016/j.conb.2017.12.004

19. Yin CL, Chen HI, Li LH, Chien YL, Liao HM, Chou MC et al (2016) Genome-wide analysis of copy number variations identifies PARK2 as a candidate gene for autism spectrum disorder. Mol Autism. https://doi.org/10.1186/s13229-016-0087-7

20. Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, Patel RV et al (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. https://doi.org/10.1038/nn.4524

21. Yasukawa M, Bando S, Dölken G, Sada E, Yakushijin Y, Fujita S et al (2001) Low frequency of BCL-2/JH translocation in peripheral blood lymphocytes of healthy Japanese individuals. Blood 98:486–488. https://doi.org/10.1182/blood.V98.2.486

22. Nambiar M, Raghavan SC (2013) Chromosomal translocations among the healthy human population: Implications in oncogenesis. Cell Mol Life Sci. https://doi.org/10.1007/s00018-012-1135-x

23. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A et al (2012) Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. https://doi.org/10.1056/NEJMoa1200395

24. Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'Ang LY et al (2003) The international HapMap project. Nature. https://doi.org/10.1038/nature02168

25. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ et al (2010) Genetic evidence for high-altitude adaptation in Tibet. Science (80-). https://doi.org/10.1126/science.1189406

26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. https://doi.org/10.1038/ng1847

27. Krämer A, Green J, Pollard J, Tugendreich S (2014) Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. https://doi.org/10.1093/bioinformatics/btt703

28. Veerappa AM, Vishweswaraiah S, Lingaiah K, Murthy M, Suresh RV, Manjegowda DS et al (2015) Global spectrum of copy number variations reveals genome organizational plasticity and proposes new migration routes. PLoS One. https://doi.org/10.1371/journal.pone.0121846

29. Murthy MN, Ramachandra NB (2017) Prioritization of differentially expressed genes in Substantia nigra transcriptomes of Parkinson's disease reveals key protein interactions and pathways. Meta Gene. https://doi.org/10.1016/j.mgene.2017.07.005

30. Suresh RV, Lingaiah K, Veerappa AM, Ramachandra NB (2017) Identifying the risk of producing aneuploids using meiotic recombination genes as biomarkers: a copy number variation approach. Indian J Med Res. https://doi.org/10.4103/ijmr.IJMR_965_14

31. Prabhanjan M, Suresh RV, Murthy MN, Ramachandra NB (2016) Type 2 diabetes mellitus disease risk genes identified by genome wide copy number variation scan in normal populations. Diabetes Res Clin Pract. https://doi.org/10.1016/j.diabres.2015.12.015

32. Vishweswaraiah S, Veerappa AM, Mahesh PA, Jahromi SR, Ramachandra NB (2015) Copy number variation burden on asthma subgenome in normal cohorts identifies susceptibility markers. Allergy, Asthma Immunol Res. https://doi.org/10.4168/aair.2015.7.3.265

33. Veerappa AM, Murthy M, Vishweswaraiah S, Lingaiah K, Suresh RV, Nachappa SA et al (2014) Copy number variations burden on mirna genes reveals layers of complexities involved in the regulation of pathways and phenotypic expression. PLoS One. https://doi.org/10.1371/journal.pone.0090391

34. Haraksingh RR, Snyder MP (2013) Impacts of variation in the human genome on gene regulation. J Mol Biol. https://doi.org/10.1016/j.jmb.2013.07.015

35. Shishido E, Aleksic B, Ozaki N (2014) Copy-number variation in the pathogenesis of autism spectrum disorder. Psychiatry Clin Neurosci. https://doi.org/10.1111/pcn.12128

36. Niarchou M, SJRA C, Doherty JL, Maillard AM, Jacquemont S, Chung WK et al (2019) Psychiatric disorders in children with 16p11.2 deletion and duplication. Transl Psychiatry. https://doi.org/10.1038/s41398-018-0339-8

37. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA et al (2013) SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. https://doi.org/10.1186/2040-2392-4-36

38. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY et al (2020) Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. https://doi.org/10.1016/j.cell.2019.12.036

39. Butler MG (2017) Clinical and genetic aspects of the 15q11.2 BP1–BP2 microdeletion disorder. J Intellect Disabil Res. https://doi.org/10.1111/jir.12382

40. Van Der Zwaag B, Staal WG, Hochstenbach R, Poot M, Spierenburg HA, De Jonge MV et al (2010) A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. Am J Med Genet Part B Neuropsychiatr Genet. https://doi.org/10.1002/ajmg.b.31055

41. Rafi SK, Butler MG (2020) The 15q11.2 bp1-bp2 microdeletion (burnside–butler) syndrome: In silico analyses of the four coding genes reveal functional associations with neurodevelopmental phenotypes. Int J Mol Sci. https://doi.org/10.3390/ijms21093296

42. Heide M, Haffner C, Murayama A, Kurotaki Y, Shinohara H, Okano H et al (2020) Human-specific ARHGAP11B increases size and folding of primate neocortex in the fetal marmoset. Science. https://doi.org/10.1126/science.abb2401

43. Lin YC, Frei JA, Kilander MBC, Shen W, Blatt GJ (2016) A subset of autism-associated genes regulate the structural stability of neurons. Front Cell Neurosci. https://doi.org/10.3389/fncel.2016.00263

44. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D et al (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature. https://doi.org/10.1038/nature13908

45. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE et al (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. https://doi.org/10.1038/nature13772

46. Bacchelli E, Battaglia A, Cameli C, Lomartire S, Tancredi R, Thomson S et al (2015) Analysis of CHRNA7 rare variants in autism spectrum disorder susceptibility. Am J Med Genet Part A. https://doi.org/10.1002/ajmg.a.36847

47. Nord AS, Roeb W, Dickel DE, Walsh T, Kusenda M, O'Connor KL et al (2011) Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. Eur J Hum Genet. https://doi.org/10.1038/ejhg.2011.24

48. Halgren C, Kjaergaard S, Bak M, Hansen C, El-Schich Z, Anderson CM et al (2012) Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. Clin Genet. https://doi.org/10.1111/j.1399-0004.2011.01755.x

49. Lin CW, Chen CY, Cheng SJ, Hu HT, Hsueh YP (2014) Sarm1 deficiency impairs synaptic function and leads to behavioral deficits, which can be ameliorated by an mGluR allosteric modulator. Front Cell Neurosci. https://doi.org/10.3389/fncel.2014.00087

50. Scholl UI, Stölting G, Nelson-Williams C, Vichot AA, Choi M, Loring E et al (2015) Recurrent gain of function mutation in calcium channel CACNA1H causes early-onset hypertension with primary aldosteronism. Elife. https://doi.org/10.7554/eLife.06315

51. Weng PH, Chen JH, Chen TF, Sun Y, Wen LL, Yip PK et al (2016) CHRNA7 polymorphisms and dementia risk: interactions with apolipoprotein 4 and cigarette smoking. Sci Rep. https://doi.org/10.1038/srep27231

52. Perez-Reyes E (2003) Molecular physiology of low-voltage-activated T-type calcium channels. Physiol Rev. https://doi.org/10.1152/physrev.00018.2002

53. Alvarez-Mora MI, Calvo Escalona R, Puig Navarro O, Madrigal I, Quintela I, Amigo J et al (2016) Comprehensive molecular testing in patients with high functioning autism spectrum disorder. Mutat Res - Fundam Mol Mech Mutagen. https://doi.org/10.1016/j.mrfmmm.2015.12.006

54. Cochran L, Welham A, Oliver C, Arshad A, Moss JF (2019) Age-related behavioural change in cornelia de lange and cri du chat syndromes: a seven year follow-up study. J Autism Dev Disord. https://doi.org/10.1007/s10803-019-03966-6

55. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH et al (2015) Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.1516376112

56. Oksenberg N, Stevison L, Wall JD, Ahituv N (2013) Function and regulation of AUTS2, a gene implicated in autism and human evolution. PLoS Genet. https://doi.org/10.1371/journal.pgen.1003221

57. Kolehmainen J, Black GCM, Saarinen A, Chandler K, Clayton-Smith J, Träskelin AL et al (2003) Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. Am J Hum Genet. https://doi.org/10.1086/375454

58. Pantaleoni F, Lev D, Cirstea IC, Motta M, Lepri FR, Bottero L et al (2017) Aberrant HRAS transcript processing underlies a distinctive phenotype within the RASopathy clinical spectrum. Hum Mutat. https://doi.org/10.1002/humu.23224

59. Xu X, Li C, Gao X, Xia K, Guo H, Li Y et al (2018) Excessive UBE3A dosage impairs retinoic acid signaling and synaptic plasticity in autism spectrum disorders. Cell Res. https://doi.org/10.1038/cr.2017.132

60. Novarino G, El-Fishawy P, Kayserili H, Meguid NA, Scott EM, Schroth J et al (2012) Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. Science (80-). https://doi.org/10.1126/science.1224631

61. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M et al (2017) Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. https://doi.org/10.1038/ng.3792

62. Feyder M, Karlsson RM, Mathur P, Lyman M, Bock R, Momenan R et al (2010) Association of mouse Dlg4 (PSD-95) gene deletion and human DLG4 gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. Am J Psychiatry. https://doi.org/10.1176/appi.ajp.2010.10040484

63. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ et al (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. https://doi.org/10.1038/nature10945

64. Ekström AB, Hakenäs-Plate L, Samuelsson L, Tulinius M, Wentz E (2008) Autism spectrum conditons in myotonic dystrophy type 1: A study on 57 individuals with congenital and childhood forms. Am J Med Genet Part B Neuropsychiatr Genet. https://doi.org/10.1002/ajmg.b.30698

65. Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S et al (2014) Meta-analysis of SHANK mutations in autism spectrum disorders: a gradient of severity in cognitive impairments. PLoS Genet. https://doi.org/10.1371/journal.pgen.1004580

66. The Deciphering Developmental Disorders S, Fitzgerald TW, Gerety SS, Jones WD, van Kogelenberg M, King DA et al (2015) Large-scale discovery of novel genetic causes of developmental disorders. Nature

67. Brett M, McPherson J, Zang ZJ, Lai A, Tan ES, Ng I et al (2014) Massively parallel sequencing of patients with intellectual disability, congenital anomalies and/or autism spectrum disorders with a targeted gene panel. PLoS One. https://doi.org/10.1371/journal.pone.0093409

68. Pinggera A, Lieb A, Benedetti B, Lampert M, Monteleone S, Liedl KR et al (2015) CACNA1D de novo mutations in autism spectrum disorders activate cav1.3 l-type calcium channels. Biol Psychiatry. https://doi.org/10.1016/j.biopsych.2014.11.020

69. Van Daalen E, Kemner C, Verbeek NE, Van Der Zwaag B, Dijkhuizen T, Rump P et al (2011) Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. Neurogenetics. https://doi.org/10.1007/s10048-011-0297-2

70. Hu J, Liao J, Sathanoori M, Kochmar S, Sebastian J, Yatsenko SA et al (2015) CNTN6 copy number variations in 14 patients: A possible candidate gene for neurodevelopmental and neuropsychiatric disorders. J Neurodev Disord. https://doi.org/10.1186/s11689-015-9122-9

71. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S et al (2012) Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2011.11.025

72. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S et al (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. https://doi.org/10.1038/nature07953

73. da Silva Montenegro EM, Costa CS, Campos G, Scliar M, de Almeida TF, Zachi EC et al (2020) Meta-analyses support previous and novel autism candidate genes: outcomes of an unexplored Brazilian cohort. Autism Res. https://doi.org/10.1002/aur.2238

74. Yang S, Guo X, Dong X, Han Y, Gao L, Su Y et al (2017) GABAA receptor subunit gene polymorphisms predict symptom-based and developmental deficits in Chinese Han children and adolescents with autistic spectrum disorders. Sci Rep. https://doi.org/10.1038/s41598-017-03666-0

75. Wang L, Li J, Shuang M, Lu T, Wang Z, Zhang T et al (2018) Association study and mutation sequencing of genes on chromosome 15q11-q13 identified GABRG3 as a susceptibility gene for autism in Chinese Han population. Transl Psychiatry. https://doi.org/10.1038/s41398-018-0197-4

76. Mikhail FM, Lose EJ, Robin NH, Descartes MD, Rutledge KD, Rutledge SL et al (2011) Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. Am J Med Genet Part A. https://doi.org/10.1002/ajmg.a.34177

77. Splawski I, Yoo DS, Stotz SC, Cherry A, Clapham DE, Keating MT (2006) CACNA1H mutations in autism spectrum disorders. J Biol Chem. https://doi.org/10.1074/jbc.M603316200

78. Barone R, Fichera M, De Grandi M, Battaglia M, Lo Faro V, Mattina T et al (2017) Familial 18q12.2 deletion supports the role of RNA-binding protein CELF4 in autism spectrum disorders. Am J Med Genet Part A. https://doi.org/10.1002/ajmg.a.38205

79. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K et al (2015) Excess of rare, inherited truncating mutations in autism. Nat Genet. https://doi.org/10.1038/ng.3303

80. Halgren C, Bache I, Bak M, Myatt MW, Anderson CM, Brondum-Nielsen K et al (2012) Haploinsufficiency of CELF4 at 18q12.2 is associated with developmental and behavioral disorders, seizures, eye manifestations, and obesity. Eur J Hum Genet. https://doi.org/10.1038/ejhg.2012.92

81. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR et al (2010) A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet. https://doi.org/10.1093/hmg/ddq307

82. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE et al (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron. https://doi.org/10.1016/j.neuron.2015.09.016

83. Petersen AK, Ahmad A, Shafiq M, Brown-Kipphut B, Fong CT, Anwar IM (2013) Deletion 1q43 encompassing only CHRM3 in a patient with autistic disorder. Eur J Med Genet. https://doi.org/10.1016/j.ejmg.2012.11.003

84. Frühmesser A, Blake J, Haberlandt E, Baying B, Raeder B, Runz H et al (2013) Disruption of EXOC6B in a patient with developmental delay, epilepsy, and a de novo balanced t(2;8) translocation. Eur J Hum Genet. https://doi.org/10.1038/ejhg.2013.18

85. Pacault M, Nizon M, Pichon O, Vincent M, Le Caignec C, Isidor B (2019) A de novo 2q37.2 deletion encompassing AGAP1 and SH3BP4 in a patient with autism and intellectual disability. Eur J Med Genet. https://doi.org/10.1016/j.ejmg.2018.11.020

86. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD et al (2013) Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2012.12.016

87. Luo W, Zhang C, Jiang YH, Brouwer CR (2018) Systematic reconstruction of autism biology from massive genetic mutation profiles. Sci Adv. https://doi.org/10.1126/sciadv.1701799

88. Novara F, Beri S, Giorda R, Ortibus E, Nageshappa S, Darra F et al (2010) Refining the phenotype associated with MEF2C haploinsufficiency. Clin Genet. https://doi.org/10.1111/j.1399-0004.2010.01413.x

89. Kamath SP, Chen AI (2019) Myocyte enhancer factor 2c regulates dendritic complexity and connectivity of cerebellar purkinje cells. Mol Neurobiol. https://doi.org/10.1007/s12035-018-1363-7

90. Warren RP, Odell JD, Warren WL, Burger RA, Maciulis A, Daniels WW et al (1996) Strong association of the third hypervariable region of HLA-DR β1 with autism. J Neuroimmunol. https://doi.org/10.1016/0165-5728(96)00052-5

91. Petek E, Schwarzbraun T, Noor A, Patel M, Nakabayashi K, Choufani S et al (2007) Molecular and genomic studies of IMMP2L and mutation screening in autism and Tourette syndrome. Mol Genet Genomics. https://doi.org/10.1007/s00438-006-0173-1

92. Baldan F, Gnan C, Franzoni A, Ferino L, Allegri L, Passon N et al (2018) Genomic deletion involving the IMMP2L gene in two cases of autism spectrum disorder. Cytogenet Genome Res. https://doi.org/10.1159/000489001

93. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG et al (2012) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science (80-). https://doi.org/10.1126/science.1227764

94. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D et al (2012) A Discovery resource of rare copy number variations in individuals with autism spectrum disorder. G3 Genes, Genomes, Genet. https://doi.org/10.1534/g3.112.004689

95. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG et al (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. https://doi.org/10.1016/j.neuron.2012.12.029

96. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A et al (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485(7397):242–245

97. Hettinger JA, Liu X, Hudson ML, Lee A, Cohen IL, Michaelis RC et al (2012) DRD2 and PPP1R1B (DARPP-32) polymorphisms independently confer increased risk for autism spectrum disorders and additively predict affected status in male-only affected sib-pair families. Behav Brain Funct. https://doi.org/10.1186/1744-9081-8-19

98. Fujita E, Dai H, Tanabe Y, Zhiling Y, Yamagata T, Miyakawa T et al (2010) Autism spectrum disorder is related to endoplasmic reticulum stress induced by mutations in the synaptic cell adhesion molecule, CADM1. Cell Death Dis. https://doi.org/10.1038/cddis.2010.23

99. Babatz TD, Kumar RA, Sudi J, Dobyns WB, Christian SL (2009) Copy number and sequence variants implicate APBA2 as an autism candidate gene. Autism Res. https://doi.org/10.1002/aur.107

100. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A et al (2011) Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Hum Mol Genet. https://doi.org/10.1093/hmg/ddr243

101. Tastet J, Decalonne L, Marouillat S, Malvy J, Thépault RA, Toutain A et al (2015) Mutation screening of the ubiquitin ligase gene RNF135 in French patients with autism. Psychiatr Genet. https://doi.org/10.1097/YPG.0000000000000100

102. Lu AT-H, Dai X, Martinez-Agosto JA, Cantor RM (2012) Support for calcium channel gene defects in autism spectrum disorders. Mol Autism. https://doi.org/10.1186/2040-2392-3-18

103. Nava C, Keren B, Mignot C, Rastetter A, Chantot-Bastaraud S, Faudet A et al (2014) Prospective diagnostic analysis of copy number variants using SNP microarrays in individuals with autism spectrum disorders. Eur J Hum Genet. https://doi.org/10.1038/ejhg.2013.88

104. Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, D'Elia E et al (2010) Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2010.01.011

105. Zamboni V, Jones R, Umbach A, Ammoni A, Passafaro M, Hirsch E et al (2018) Rho GTPases in intellectual disability: from genetics to therapeutic opportunities. Int J Mol Sci. https://doi.org/10.3390/ijms19061821

106. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S et al (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. https://doi.org/10.1038/ng.835

107. Yoo J, Bakes J, Bradley C, Collingridge GL, Kaang BK (2014) Shank mutant mice as an animal model of autism. Philos Trans R Soc B Biol Sci. https://doi.org/10.1098/rstb.2013.0143

108. Correia CT, Coutinho AM, Sequeira AF, Sousa IG, Lourenço Venda L, Almeida JP et al (2010) Increased BDNF levels and NTRK2 gene association suggest a disruption of BDNF/TrkB signaling in autism. Genes, Brain Behav. https://doi.org/10.1111/j.1601-183X.2010.00627.x

109. Reiersen AM, Todorov AA (2011) Association between DRD4 genotype and autistic symptoms in DSM-IV ADHD. J Can Acad Child Adolesc Psychiatry

110. Steinmetz AB, Stern SA, Kohtz AS, Descalzi G, Alberini CM (2018) Insulin-like growth factor II targets the mTOR pathway to reverse autism-like phenotypes in mice. J Neurosci. https://doi.org/10.1523/JNEUROSCI.2010-17.2017

111. Vorstman JAS, Van Daalen E, Jalali GR, Schmidt ERE, Pasterkamp RJ, De Jonge M et al (2011) A double hit implicates DIAPH3 as an autism risk gene. Mol Psychiatry. https://doi.org/10.1038/mp.2010.26

112. Kushima I, Aleksic B, Nakatochi M, Shimamura T, Okada T, Uno Y et al (2018) Comparative analyses of copy-number variation in autism spectrum disorder and schizophrenia reveal etiological overlap and biological insights. Cell Rep. https://doi.org/10.1016/j.celrep.2018.08.022

113. Balicza P, Varga NÁ, Bolgár B, Pentelényi K, Bencsik R, Gál A et al (2019) Comprehensive analysis of rare variants of 101 autism-linked genes in a Hungarian cohort of autism spectrum disorder patients. Front Genet. https://doi.org/10.3389/fgene.2019.00434

114. An JY, Cristino AS, Zhao Q, Edson J, Williams SM, Ravine D et al (2014) Towards a molecular characterization of autism spectrum disorders: an exome sequencing and systems approach. Transl Psychiatry. https://doi.org/10.1038/tp.2014.38

115. Ping LY, Chuang YA, Hsu SH, Tsai HY, Cheng MC (2016) Screening for mutations in the TBX1 gene on chromosome 22q11.2 in Schizophrenia. Genes (Basel). https://doi.org/10.3390/genes7110102

116. Wang K, Zhang H, Ma D, Bucan M, Glessner JT (2009) Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. https://doi.org/10.1038/nature07999

117. Chung RH, Ma D, Wang K, Hedges DJ, Jaworski JM, Gilbert JR et al (2011) An X chromosome-wide association study in autism families identifies TBL1X as a novel autism spectrum disorder candidate gene in males. Mol Autism. https://doi.org/10.1186/2040-2392-2-18

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.