

RESEARCH

Open Access



# Comparative yield of molecular diagnostic algorithms for autism spectrum disorder diagnosis in India: evidence supporting whole exome sequencing as first tier test

Frenny Sheth<sup>1\*</sup>, Jhanvi Shah<sup>1†</sup>, Deepika Jain<sup>2†</sup>, Siddharth Shah<sup>3</sup>, Harshkumar Patel<sup>4</sup>, Ketan Patel<sup>5</sup>, Dhaval I Solanki<sup>6</sup>, Anand S Iyer<sup>7</sup>, Bhargavi Menghani<sup>8</sup>, Priti Mhatre<sup>9</sup>, Sanjiv Mehta<sup>3</sup>, Shruti Bajaj<sup>10</sup>, Vishal Patel<sup>11</sup>, Manoj Pandya<sup>12</sup>, Deepak Dhami<sup>13</sup>, Darshan Patel<sup>14</sup>, Jayesh Sheth<sup>1</sup> and Harsh Sheth<sup>1\*</sup>

## Abstract

**Background** Autism spectrum disorder (ASD) affects 1 in 100 children globally with a rapidly increasing prevalence. To the best of our knowledge, no data exists on the genetic architecture of ASD in India. This study aimed to identify the genetic architecture of ASD in India and to assess the use of whole exome sequencing (WES) as a first-tier test instead of chromosomal microarray (CMA) for genetic diagnosis.

**Methods** Between 2020 and 2022, 101 patient-parent trios of Indian origin diagnosed with ASD according to the Diagnostic and Statistical Manual, 5th edition, were recruited. All probands underwent a sequential genetic testing pathway consisting of karyotyping, Fragile-X testing (in male probands only), CMA and WES. Candidate variant validation and parental segregation analysis was performed using orthogonal methods.

**Results** Of 101 trios, no probands were identified with a gross chromosomal anomaly or Fragile-X. Three (2.9%) and 30 (29.7%) trios received a confirmed genetic diagnosis from CMA and WES, respectively. Amongst diagnosis from WES, SNVs were detected in 27 cases (90%) and CNVs in 3 cases (10%), including the 3 CNVs detected from CMA. Segregation analysis showed 66.6% ( $n=3$  for CNVs and  $n=17$  for SNVs) and 16.6% ( $n=5$ ) of the cases had *de novo* and recessive variants respectively, which is in concordance with the distribution of variant types and mode of inheritance observed in ASD patients of non-Hispanic white/ European ethnicity. *MECP2* gene was the most recurrently mutated gene ( $n=6$ ; 20%) in the present cohort. Majority of the affected genes identified in the study cohort are involved in synaptic formation, transcription and its regulation, ubiquitination and chromatin remodeling.

<sup>†</sup>Jhanvi Shah and Deepika Jain contributed equally to this work.

\*Correspondence:

Frenny Sheth  
frenny.sheth@frige.co.in  
Harsh Sheth  
harsh.sheth@frige.co.in

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions** Our study suggests *de novo* variants as a major cause of ASD in the Indian population, with Rett syndrome as the most commonly detected disorder. Furthermore, we provide evidence of a significant difference in the diagnostic yield between CMA (3%) and WES (30%) which supports the implementation of WES as a first-tier test for genetic diagnosis of ASD in India.

**Keywords** Autism spectrum disorder, Genetic etiology, Diagnostic yield, Chromosomal microarray, Whole exome sequencing, *De novo*, Rett syndrome, India

## Background

Autism spectrum disorder (ASD) is a heterogeneous group of neurodevelopmental disorders (NDD) with a prevalence of approximately 1 in 160 children worldwide [1] and with variable clinical presentations and outcomes [2]. According to the latest version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), it is characterized by impaired social communication along with repetitive behavior or restricted interests which can persist throughout lifetime [3, 4]. In addition to these core features, many affected individuals can be afflicted with comorbidities like intellectual disability and epilepsy. A review and meta-analysis of ASD in India reported low prevalence of only 0.0014–0.0012% in children aged 1–18 years compared to developed countries like the United States and United Kingdom with a prevalence of 1–1.5% [5]. However, a review across the South Asian population reported its prevalence rate ranging from 0.09 to 1.07% which is similar to that observed in developed countries [6].

The etiology of ASD is not fully understood, although, similar to several neurodevelopmental disorders, genetic risk and environmental exposure appears to contribute to the pathogenesis of ASD [7, 8]. Data from twin studies suggest a strong genetic role and a quantitative meta-analysis on all published twin studies in the context of ASD has estimated heritability component between 64 and 91% [9]. Therefore, genetic testing is recommended in ASD patients and as of 2013, an etiology underlying ASD could be established in around 6–15% cases [10]. Guidelines put forth a decade ago by the American College of Medical Genetics (ACMG) suggests using chromosomal microarray (CMA) as a first line test in ASD since its diagnostic yield was estimated to be between 7 and 9% [2, 10]. However, since then, studies using whole exome sequencing (WES) have evidenced sequence level contribution of *de novo* variants in the etiology of ASD and recent advancements in computational analyses of WES data suggests improvement in detection of copy number variants (CNVs) too. Indeed, two recent studies have shown that WES was able to detect nearly all clinically relevant CNVs that were detected by CMA thereby increasing its diagnostic yield by approximately 1.6% [11, 12]. In addition, a recent retrospective study using WES on clinically diagnosed 343 children with ASD from Spain suggested a diagnostic yield of ~14% with 75% of

the cases harbouring a *de novo* variant [1]. It is predicted that nearly 85% of the disease causing variants reside in the protein coding and splice site regions of the genome, which are well covered by WES [13–15]. Various studies have repeatedly shown a better yield and utility of WES over CMA in NDD and thus, WES has now been suggested as a first-tier test for patients with intellectual disability/ NDD [16, 17].

Selection and availability of a first-tier test with high diagnostic yield is desirable in low-middle income countries (LMICs) like India, since patients and families bear the cost of genetic testing. To our knowledge, no study to date has been performed in the Indian population to delineate the genetic architecture of ASD which can aid in the selection of first-tier genetic test. Here, we report the first systematic study to assess the genetic architecture and molecular diagnostic yields for karyotype, Fragile-X testing, CMA and WES in a population-based cohort of 101 patient-parent trios with ASD from India.

## Materials and methods

### Patient recruitment and sample collection

The study included consecutively recruited 101 children with a confirmed clinical diagnosis of idiopathic ASD based on the DSM-5 [3, 4]. Children with prominent syndromic features, isolated speech delay or isolated sensory processing disorders were excluded from this study. Blood samples of the patient-parent trios were collected. The parents or guardians of all probands provided a written informed consent as per the Helsinki Declaration and the study was approved by the research ethics committee at Foundation for Research in Genetics and Endocrinology, Ahmedabad (ID: FRIGE/IEC/19/2020). All the methods in the study were carried out as per the Helsinki Declaration. High molecular weight genomic DNA was extracted using desalting method [18] and was stored at -20 °C until molecular genetic testing was carried out.

### Karyotyping and Fragile-X testing

Karyotyping was performed in all cases regardless of sex, whereas Fragile-X testing was performed only in male probands. Karyotyping was carried out using GTG banding at 500 band resolution to check for gross chromosomal aberrations. Fragile-X testing was carried out by triplet repeat primed – polymerase chain reaction (TP-PCR), that involved analyzing CGG repeat expansion in

the 5' UTR of the *FMR1* gene using method as previously described [19]. Children with a normal chromosomal constitution and showing no expansion of the CGG repeats in the 5' UTR of *FMR1* gene were subsequently assessed with CMA and WES.

### Chromosomal microarray

CMA was carried out using CytoScan™ Optima array, GeneChip™ System 3000 and Affymetrix platform (Thermo Fisher Scientific, USA) as per the manufacturer's instructions. Chromosome Analysis Suite Software (ChAS) (Thermo Fisher Scientific, USA) was used to carry out the analysis of the data as per the manufacturer's recommendations which suggested a minimum resolution of 1 Mb for losses, 2 Mb for gains and 5 Mb for copy neutral loss of heterozygosity. For all candidate CNVs, variants were primarily screened for population frequency and known disease associations using publicly available databases like gnomAD database [20], DGV [21] and DECIPHER [22] and OMIM [23]. Pathogenicity of CNVs were classified in accordance with ACMG and ClinGen classification system [24]. All candidate CNVs were validated in proband and parents using SYBR Green based quantitative PCR (Q-PCR) using ABI's StepOne Real Time PCR system (Thermo Fisher Scientific, USA) (Supplementary Table 1).

### Whole exome sequencing

Genomic DNA of the proband was subjected to selective capture and sequencing of the protein coding regions that included exons and exon-intron boundaries of genes using Agilent SureSelect v6 enrichment kit (Agilent, USA). The library prepared, was subjected to paired-end sequencing with a mean coverage of >80-100x on the Illumina HiSeq or NovaSeq platform (Illumina, USA). Sequences obtained as FASTQ files were aligned to the human reference genome (GRCh37/hg19) using BWA MEM v0.7.12 [25]. SNVs and indels were called using GATK v4.12 Haplotype caller [26]. In addition to SNVs and small indels, copy number variants (CNVs) were detected from the data using the ExomeDepth v1.1.10 [27].

Variant annotation, filtration and prioritization was performed using Exomiser v12.1.0 [28]. Exomiser uses the hiPHIVE prioritization method that incorporates protein-protein interaction networks and multi-species ontologies along with ranking candidate genes based on the predicted variant pathogenicity associated with the phenotype. The phenotype information was coded in uniform human phenotype ontology (HPO) terminologies [29]. Common variants were filtered based on minor allele frequency in the 1000Genome Phase 3 [30] and gnomAD v2.1 [20] databases. The minor allele frequency cut off was set at 0.02 (2%). The cut-off was set

assuming ASD has a global prevalence of 1:100; the frequency of major and minor alleles would be 0.9 (p) and 0.1 (q), respectively, based on the Hardy-Weinberg equilibrium. As ASD is caused by dominant *de novo* variants in majority of the cases ( $pq=0.09$ ) and the prior estimates suggests genetic diagnostic yield of approximately 33%,  $pq$  would be 0.027. Only non-synonymous variants in the coding region and canonical splice site variants with a depth of >20x were used for analysis and clinical correlation. Various *in-silico* prediction tools such as PolyPhen-2 [31], SIFT [32], MutationTaster2 [33], LRT [34], CADD [35] and MetaDome [36] were used to predict pathogenicity of non-synonymous and indel variants. A CADD\_phred score of  $\geq 15$ , slightly intolerant, intolerant or highly intolerant predictions of MetaDome and at least two damaging predictions from the remaining *in-silico* tools were used for selection of candidate variants. *In-silico* predictions along with available knowledge from various sources and databases as described below was used in prioritising the variant.

Post-gross filtering, variants were prioritized based on the following: (a) known disease causing variant previously reported in databases like ClinVar [37] and HGMD [38]; (b) novel variants in known genes based on the Z-score for missense and pLoF or LOEUF score for loss of function variants available in the gnomAD database [20]; (c) variants in novel candidate genes wherein the respective gene was additionally evaluated for their function using UniProt [39] and Human Protein Atlas (proteintlas.org) [40]. Tissue expression using GTEx database (gtexportal.org), association/ interaction with known ASD genes using STRING database [41] and, plausible phenotypic outcome in murine models based on the MGI database [42] were assessed. All candidate variants were assessed using IGV [43] to evaluate their quality.

In the case of candidate CNVs, variants were primarily screened for population frequency and known disease associations using publicly available databases like gnomAD database [20], DGV [21], DECIPHER [22] and OMIM [23]. Pathogenicity of CNVs were classified in accordance with ACMG and ClinGen classification system [24].

All candidate SNVs and indels were validated in proband and parents using bi-directional Sanger sequencing using ABI's SeqStudio platform (Thermo Fisher Scientific, USA) whereas all candidate CNVs were validated using SYBR Green based quantitative PCR (Q-PCR) using ABI's StepOne Real Time PCR system (Thermo Fisher Scientific, USA) (Supplementary Table 1). This was conducted to delineate mode of inheritance and reclassify variant pathogenicity.

The classification of SNVs was carried out according to the American College of Medical Genetics – American

College of Pathologists (ACMG-AMP) guidelines [44] and ClinGen framework [24].

## Results

### Study cohort

The study cohort consisted of 101 well defined patient-parent trios diagnosed with moderate to severe ASD of unknown etiology as per the DSM-5 criteria. The average age at recruitment was  $5 \pm 3$  years and ranged from 2 to 6 months to 16 years (Table 1). The average maternal and paternal age at the time of conception was  $28 \pm 4$  years and  $30 \pm 4$  years, respectively. The cohort included 72 males (71%) and 29 females (29%), suggesting a male to female ratio of approximately 3:1. Five families had more than one child diagnosed with ASD (Supplementary Information 1). Consanguinity was noted in 8 families (7.9%), whereas non-consanguinity and endogamy in 31 (30.7%) and 62 (61.4%) families, respectively. All 101 probands with ASD also had developmental delay and intellectual disability with some of them having subtle dysmorphism (large and/ or cupped ears, long eyelashes,

telecanthus, thin upper lip) ( $n=28/101$ ; 27.7%) and epilepsy ( $n=28/101$ ; 27.7%) (Supplementary Table 2).

### Outcomes from karyotype and fragile X testing

Sequential genetic testing was performed in all 101 patients which began with karyotyping and were followed by fragile X testing (only in male probands), CMA and WES. None of the probands showed gross chromosomal aberrations or had expanded triplet repeat tracks (full-mutation alleles with  $>200$  CGG repeats) in the 5'-UTR region of the *FMR1* gene. Therefore, all probands were subsequently tested using CMA and WES.

### Outcomes from chromosomal microarray

From the 101 probands in whom CMA was performed, pathogenic CNVs were detected in 3 cases (2.9%) including two deletions and one duplication (Table 2). Proband ASD-076 had an 8 Mb deletion at the 15q11.2 locus which encompassed 20 OMIM genes and is known to cause 15q11.2 deletion syndrome (OMIM#615,656) or Angelman syndrome (OMIM#105,830). Compared to the individuals with class II deletions (BP2-BP3; ISCA-37,478), individuals with large class I deletions (BP1-BP3; ISCA-37,404) at the 15q11.2 region are observed to have a high likelihood of language impairment and autistic traits, similar to that seen in the proband in our study [45]. Patient ASD-103 was detected with a deletion of 0.19 Mb size at the 9q34.3 locus which encompassed 6 OMIM genes and is associated with Kleefstra syndrome I (OMIM#610,253). Individuals with  $>1$  Mb deletion of the 9q34 locus have a severe phenotype such as congenital anomalies including heart defects, limb anomalies, seizures and respiratory distress. In contrast individuals having  $<1$  Mb deletion are observed with a milder phenotype, which in part could explain the phenotype in the proband in the current study such as bruxism, drooling, subtle facial dysmorphism and recurrent episodes of vomiting [46, 47]. Lastly, proband ASD-050 was detected with a 0.52 Mb duplication on the 1q22 locus which consists of 8 OMIM genes. This is a rare CNV which has previously only been reported in a boy with intellectual disability and psychiatric disturbances [48]. Multiple individuals in this family were affected and the duplication variant segregated with the neurological features in all family members with this variant. All CNVs in our cohort were *de novo* in origin and were observed exclusively in male probands.

### Outcomes from whole exome sequencing

WES was carried out in 99 of 101 cases, as the cohort contained two monozygotic twin pairs and only one proband from each twin pair was processed for WES. The 99 cases also included the three cases that yielded a result by CMA to assess the sensitivity of WES to detect CNVs.

**Table 1** Demographics of 101 patient-parent trios

Variable	Whole Sample (N = 101)
<b>Gender, n (%)</b>	
Male	73 (72)
Female	28 (28)
Male-female ratio	2.6:1
<b>Age, years (SD)</b>	
Age at diagnosis of probands	5 (3)
Maternal age at conception	28 (4)
Paternal age at conception	30 (4)
<b>Type of Marriage, n (%)</b>	
Consanguineous	8 (8)
Non-consanguineous	31 (31)
Endogamous	62 (61)
<b>Phenotype, n (%)</b>	
Developmental delay	101 (100)
Speech delay	101 (100)
Intellectual disability	101 (100)
Epilepsy/ seizures	28 (27.7)
Subtle facial dysmorphism	28 (27.7)
Regression (social/ speech)	55 (54.4)
<b>Genetic testing received<sup>a</sup>, n (%)</b>	
Karyotype	101 (100)
Fragile-X ( <i>FMR1</i> triplet repeat expansion)	73 (72)
Chromosomal microarray	101 (100)
Whole exome sequencing <sup>b</sup>	99 (98)

<sup>a</sup> Genetic testing was carried out in proband only. In cases with a candidate variant, orthogonal testing approaches (Sanger sequencing and/or Q-PCR) were used to assess and validate the variant in the parents

<sup>b</sup> Whole exome sequencing was carried out in 99 of 101 cases, as the cohort contained two monozygotic twin pairs and only one proband from each twin pair was processed

**Table 2** List of cases observed with pathogenic or likely pathogenic copy number variation using CMA and/or WES.

Sr No	Case ID	Chr		Chromosomal microarray		Whole exome sequencing			CNV Type	Zygosity	Inheritance pattern	Inherited from	Variant classification	OMIM Disease (OMIM ID)
		No	Chr	Genomic coordinate	Size (Kbp)	Genomic coordinate	Variant							
1	ASD-021	15	-	-	-	?_20279345_21145729_?	chr15:g.(?_20279345_(21145729_?)del		loss	Het	AD	De novo	Variant of uncertain significance	Chromosome 15q11.2 deletion syndrome (#615,656)
2	ASD-050	1	arr[GRCh37]1q22(155446632_155971760)x3	525	?_155611966_156081998_?	chr1:g.(?_155611966_(156081998_?)dup		gain	Het	Het	AD	De novo	Likely pathogenic	Chromosome 1q22 microduplication syndrome (#615,656)
3	ASD-076	15	arr[GRCh37]15q11.2q13.2(22770421_30913574)x1	8,143	?_24955049_28272386_?	chr15:(?_24955049_(28272386_?)del		loss	Het	Het	AD	De novo	Pathogenic	Chromosome 15q11.2 deletion syndrome (#615,656)
4	ASD-103	9	arr[GRCh37]9q34.3(140338142_140533378)x1	195	?_137426248_137629023_?	chr9:g.(?_137426248_(137629023_?)del		loss	Het	Het	AD	De novo	Pathogenic	Kleefstra syndrome-1 (#610,253)

Chr = chromosome; No = number; Kbp = kilo basepair; CNV = copy number variation; Het = heterozygous; AD = autosomal dominant



On an average, approximately 3 candidate gene(s) or variant(s) were identified per proband (Supplementary Table 3).

From the 101 patients, pathogenic and/ or likely pathogenic variants were identified in 30 cases (29.7%), of which, SNVs were detected in 27 cases (90%) and CNVs in 3 cases (10%) (Table 3). Interestingly, 3 CNVs detected by CMA were also identified by WES, however, a 0.8 Mb *de novo* deletion encompassing the BP1 region of the 15q11.2 locus was detected by WES alone (Table 2). On further analysis, the lack of detection of the aforementioned CNV by CMA was due to the lack of probes covering this region on CytoScan™ Optima array.

Segregation analysis revealed that approximately 66.6% (n=3 for CNVs and n=17 for SNVs) of the cases were caused due to a *de novo* variant. *De novo* SNVs were found primarily in previously known ASD genes-*MECP2*, *SCN2A*, *KCNQ2*, *TBL1XR1*, *CNTNAP2*, *TCF4*, *CAMK2A*, *NF1*, *AUTS2*, *FOXP2* and *NLGN3*. Of 17 *de novo* variants, 6 were predicted to be loss of function (pLOF) variants (35.2%) whereas the remaining were missense variants. Remarkably, 6 of the 17 patients had a *de novo* SNV in the *MECP2* gene, which is associated with Rett syndrome (OMIM#312,750). Of these, 5 were female and 1 was a male proband. Interestingly, in a rare case of the male proband aged 2.5 years with Rett syndrome, we observed that the variant c.538 C>T (p.Arg180Ter) in the *MECP2* gene originated through a post-zygotic *de novo* event which led to somatic mosaicism in the proband (Table 3) [49].

In our cohort of patients with pathogenic/ likely pathogenic variants, 5 probands (n=5/30; 16.6%) were observed with biallelic or hemizygous variants in genes associated with NDD or metabolic disorders with a recessive mode of inheritance (Table 3). Specifically, biallelic variants were detected in (i) *ALDH4A1* gene which is associated with hyperprolinemia type II (OMIM#239,510), (ii) *NEUROG1* gene which is associated with congenital cranial dysinnervation disorder and autism spectrum disorder [50], (iii) *KDM6A* gene which is associated with Kabuki syndrome 2 (OMIM#300,867), (iv) *LMAN2L* gene which is associated with mental retardation 52 (OMIM#616,887) and, (v) *ALDH7A1* gene which is associated with pyridoxine dependent epilepsy (OMIM#266,100).

In addition, 4 probands were identified with pathogenic/ likely pathogenic heterozygous variants, which were inherited from one of their parents. In 2 cases, the variants were inherited from unaffected mother and in 1 case the variant was inherited from an unaffected father. In the 4th case, pLOF variant c.202 C>T (p.Gln68Ter) in the *RORB* gene was inherited from father who also had a clinical history of seizures (Supplementary Table 2; Supplementary Information 1). Of note, in one case

(ASD-003), paternal sample was un-available, hence the mode of inheritance couldn't be deduced. Interestingly, ASD probands with epilepsy had a higher diagnostic yield (n=15/28; 53.6%) compared to ASD probands without epilepsy (n=15/73; 20.5%) ( $\chi^2=10.6$ ,  $p=0.001$ ), however, no such association was observed for facial dysmorphism ( $\chi^2=0.67$ ,  $p=0.41$ ) and social/ speech regression phenotypes ( $\chi^2=0.53$ ,  $p=0.47$ ).

Lastly, WES identified 22 VUS variants in 21 patients (n=21/101; 20.8%; Supplementary Table 4). The variants were identified in genes that have previously been associated with or implicated in ASD etiology as per the Simons Foundation Autism Research Initiative (SFARI) Gene Database and Autism Database (AutDB). Of these, majority of the probands were detected with heterozygous variants (66.6%) which were inherited from either of the unaffected parents with equal distribution. Of note, 3 of the 21 patients following segregation analysis were detected with missense variants in the *KMT2C* gene (Kleefstra syndrome 2; OMIM#617,768) which were inherited from a healthy parent. Whilst the majority of the cases have been reported with a *de novo* variant in the *KMT2C* gene, 4 reports observed variants being inherited from a healthy parent suggesting a potential oligogenic mode of inheritance [51–54].

## Discussion

Almost a decade ago, the ACMG published guidelines recommending CMA as a first tier test for delineating the genetic cause of ASD and other NDDs [2, 10]. Since then, WES coupled with advancements in computational analyses has led to simultaneous detection of SNVs and CNVs. Studies carried out in multiple ethnic populations since 2015 have shown an increased diagnostic yield from WES compared to CMA in ASD [1, 2, 55, 56]. This outcome is supported by the observation of a high proportion of *de novo* SNVs in ASD patients which are not detectable by CMA. To our knowledge, we here report the first description of the genetic architecture of ASD and simultaneously carry out diagnostic yield comparisons of karyotype, *FMRI* triplet repeat expansion, CMA and WES in a cohort of 101 patient-parent trios of Indian origin.

Our data is in congruence with prior reports and supports the utility of WES as a primary genetic diagnostic method for ASD. In the present cohort, WES detected pathogenic/ likely pathogenic variants causative of the ASD phenotype in 29.7% of the cases in contrast with 2.9%, 0% and 0% from CMA, *FMRI* triplet repeat expansion and karyotype testing, respectively. Indeed, all three CNVs detected by CMA were also detected by WES together with a fourth CNV which was detected by WES alone. Interestingly, the low yield of CMA in the present cohort can be attributed to two potential reasons. First,

gross dysmorphism was an exclusion criteria during recruitment of cases for the study. Prior study by Tamimides et al., has shown a higher diagnostic yield of CMA in children with ASD and major congenital anomaly compared with children with minor physical anomaly [2]. Two, Affymetrix CytoScan Optima oligonucleotide array was used in the current study. The platform consists of 315,608 probes and requires at least 25 probes to call a loss or gain of approximately 100 kb in size. Prior study has shown a trend for differential diagnostic yield with CMA based on both platform resolution and phenotypic manifestation in ASD patients [2]. A higher resolution microarray (1 million probes or more) had a higher diagnostic yield in ASD patients with minor physical anomalies compared to low resolution microarray (44k platform), however, this difference was abated when the test was carried out in ASD patients with major congenital anomalies [2]. It is therefore plausible that the current platform may have missed CNVs that are beyond its detection limit, which could have been picked up with a higher resolution microarray platform. The diagnostic yield in the present cohort is concordant with those reported previously from individual cohort studies [1, 2, 55, 56]. Indeed, a recent meta-analysis in patients with NDD i.e. global developmental delay, intellectual disability and ASD showed diagnostic yield of WES to range from 31 to 53% in contrast to CMA with yield ranging 15–20% [16]. Based on these results, Srivastava et al. outlined a consensus statement and a stepwise algorithm for NDD diagnosis whereby WES is presented as the first-tier test followed by CMA and/or other orthogonal tests.

Interestingly, we observed that in 66.6% and 16.1% of the cases with a genetic diagnosis for ASD, the mode of inheritance for the variant was *de novo* and recessive, respectively. This is in congruence with prior patient-parent trio cohort studies whereby similar rates for variant's mode of inheritance was observed [1, 2, 57]. All genes identified carrying potential causative variants were subjected to STRING analysis v11.5 (Fig. 1). The network statistics consisted of 37 unique proteins resulting in 67 various protein-protein interactions (PPI) amongst themselves. In comparison, a random set of same number of proteins, would result in only 12 different interactions. With a *p*-value of  $<1.0\text{e-}16$ , a statistically significant enrichment of PPI in the present cohort indicated a biological connection amongst these proteins. Majority of these proteins are involved in synaptic formation, transcription and its regulation, ubiquitination and chromatin remodeling, as have been observed in prior studies [58]. This leads to a plausible hypothesis that the genetic architecture and etiopathogenesis of ASD is similar across ethnicities and an introduction of a uniform stepwise genetic testing algorithm would yield similar diagnostic yields.

In our cohort, three genes (*LRFN1*, *UNC13A* and *UNC79*) were identified as potential novel candidates for ASD. The variant in the *LRFN1* gene was a result of a *de novo* event. *LRFN1* interacts with *DLG4*, a known ASD gene vital in the formation of the post-synaptic complex required for signal transduction [59]. *DLG4* is classed under a high confidence category with a gene score of 1 in the SFARI database and has an Evaluation of Autism Gene Link Evidence (EAGLE) score of 2.45, which suggests limited but no contradicting evidence of its role in ASD. Due to the direct interaction between the two genes, *LRFN1* could be considered as a potential candidate for ASD, although functional validation is required and was beyond the scope of the current study. The variants in the *UNC13A* and *UNC79* genes were inherited from likely asymptomatic parents and classed as VUS. Both these genes have been listed in the AutDB and SFARI database and have been considered novel due to the absence of an associated phenotype in the OMIM database. A patient with developmental delay, dyskinetic movement disorder and autism has been previously identified with a *de novo* variant in the *UNC13A* gene [60]. Additionally, experimental evidence suggests its direct interaction with a known ASD associated gene, *STXBP1*. Only recently, *UNC79* gene has also been associated with neurodevelopmental features including autism [61].

With an increasing awareness of ASD amongst the general populous, there is a high likelihood of increase in demand for genetic testing in children with ASD. In a survey of parents having a child with ASD in USA, 80% of the parents indicated that they would pursue genetic testing to identify risk of ASD in the younger sibling [62]. However, financial concerns, not being offered genetic testing by a physician or a geneticist and lack of awareness are amongst the most common reasons for not opting for genetic diagnosis [63]. In addition, with the advent of development and deployment of new treatments such as trofinetide for Rett syndrome, there is likely to be increase in uptake for genetic testing [64]. This suggests that adoption of a uniform genetic testing algorithm coupled with educating primary care physicians and non-genetic specialists could improve rates of genetic testing and diagnosis in children with ASD.

### Limitations

The limitations of our study include a relatively small sample size, possible ascertainment bias related to patients having primarily non-syndromic form of ASD without gross congenital dysmorphism, carrying out WES and CMA in the proband only followed by segregation analysis by orthogonal approaches on prioritized variants and absence of detailed cost-effectiveness assessment. Despite this, we observe similar diagnostic yields to that observed in other cohorts [1, 2, 55]. Additionally,

**Table 3** List of cases observed with pathogenic or likely pathogenic single nucleotide variation using WES

Sr No	Case ID	Chr No	Genomic coordinate	Ref allele	Alt allele	Gene (OMIM ID)	Transcript ID	Exon	Variant	Protein change	Zygosity	Mode of inheritance	Inherited from	Variant classification	OMIM Disease (OMIM ID)
1	ASD-001	20	62,071,010	C	A	KCNQ2 (*602,235)	ENST00000359125.2 NM_172107.4	6	c.868G>T	p.Gly290Cys	Het	AD	De novo	Pathogenic	Developmental and epileptic encephalopathy 7 (#613,720);
2	ASD-003	7	114,299,416	G	A	FOXP2 (*605,317)	ENST00000408937.3 NM_148898.4	13	c.1549G>A	p.Ala517Thr	Het	AD	not Maternal <sup>a</sup>	Likely pathogenic	Speech-language disorder-1 (#602,081)
3	ASD-010	2	166,172,149	G	T	SCN2A (*182,390)	ENST00000375437.2 NM_001040142.2	11	c.1552G>T	p.Glu518Ter	Het	AD	De novo	Pathogenic	Developmental and epileptic encephalopathy 11 (#613,721)
4	ASD-011	1	19,212,049	-	GGTCTGCA	ALDH4A1 (*606,811)	ENST00000375341.3 NM_003748.4	5	c.363_370dup	p.Arg124LeufsTer9	Compound heter	AR	Maternal	Pathogenic	Hyperprolinemia, Type II (#239,510)
			19,209,638	C	G			7	c.658G>C	p.Ala220Pro			Paternal	Likely pathogenic	
5	ASD-013	17	28,530,262	T	TC	SLC6A4 (*182,138)	ENST00000261707.3 NM_001045.6	14	c.1745dup	p.Thr583AsnfsTer23	Het	AD; -	Maternal	Likely pathogenic	{Obsessive-compulsive disorder} (#164,230); {Anxiety-related personality traits} (#607,834)
6	ASD-016	X	153,363,096	CGGCG	-	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	1	c.23_27del	p.Ala8GlufsTer32	Het	XLD	De novo	Pathogenic	Rett syndrome (#312,750)



Table 3 (continued)

Sr No	Case ID	Chr	Genomic No	Ref allele	Alt allele	Gene (OMIM ID)	Transcript ID	Exon	Variant	Protein change	Zygoty	Mode of inheritance	Inherited from	Variant classification	OMIM Disease (OMIM ID)
7	ASD-017	3	184,072,060	G	A	TBL1XR1 (*608,628)	ENST00000457928.2 NM_024665.7	7	c.679G>A	p.Asp227Asn	Het	AD	De novo	Likely pathogenic	Intellectual developmental disorder, autosomal dominant 41 (#616,944); Pierpont syndrome (#602,342)
8	ASD-020	X	153,296,473	-	C	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	3	c.842dup	p.Arg282ProfsTer61	Het	XLD	De novo	Pathogenic	Rett syndrome (#312,750)
9	ASD-022	3	184,072,060	G	A	CLCN2 (*600,570)	ENST00000265593.4 NM_004366.6	15	c.1550 C>T	p.Thr517Met	Het	AD	Paternal	Likely pathogenic	[Epilepsy, idiopathic general-ized, susceptibility to, 11} (#607,628)
10	ASD-034	5	134,871,150	-		NEUROG1 (*601,726)	ENST00000314744.4 NM_006161.3	1	c.228_231dup	p.Thr78Profs*122	Homo	UNK	Maternal and Paternal	Likely pathogenic	-
11	ASD-045	19	39,805,801	A	G	LRFN1 (*612,807)	ENST00000248668.4 NM_020862.2	1	c.176T>C	p.Val59Ala	Het	UNK	De novo	Likely pathogenic	-
12	ASD-046	7	147,092,836	C	T	CNTNAP2 (*604,569)	ENST00000361727.3 NM_014141.6	10	c.1634 C>T	p.Ala545Val	Het	AD	De novo	Likely pathogenic	[Autism susceptibility 15} (#612,100)
13	ASD-047	18	52,896,081	G	A	TCF4 (*602,272)	ENST00000398339.1 NM_001243226.3	19	c.2182 C>T	p.Arg278Ter	Het	AD	De novo	Pathogenic	Pitt-Hopkins syndrome (#610,954)

Table 3 (continued)

Sr No	Case ID	Chr	Genomic No	Ref allele	Alt allele	Gene (OMIM ID)	Transcript ID	Exon	Variant	Protein change	Zygoty	Mode of inheritance	Inherited from	Variant classification	OMIM Disease (OMIM ID)
14	ASD-049	2	166,245,627	T	A	SCN2A (*182,390)	ENST00000375437.2 NM_001040142.2	27	c.5311T>A	p.Tyr1771Asn	Het	AD	De novo	Likely pathogenic	Developmental and epileptic encephalopathy-11; (#613,721); Benign familial neonatal-infantile seizures-3; (#607,745); Episodic ataxia type 9 (618,924)
15	ASD-053	5	149,629,805	G	T	CAMK2A (*114,078)	ENST00000348628.6 NM_015981.4	11	c.884 C>A	p.Ala295Asp	Het	AD	De novo	Likely pathogenic	Intellectual developmental disorder, autosomal dominant 53 (#617,798)
16	ASD-054	X	44,949,169	C	T	KDM6A (*300,128)	ENST00000377967.4 NM_021140.4	25	c.3730 C>T	p.Leu124Phe	Hemi	XL	Maternal	Likely pathogenic	Kabuki syndrome 2 (#300,867)
17	ASD-057	X	153,296,824	G	C	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	3	c.491 C>G	p.Pro164Arg	Het	XLD	De novo	Pathogenic	Rett syndrome (#312,750)
18	ASD-058	17	60,042,409	G	T	MED13 (*603,808)	ENST00000397786.2 NM_005121.3	20	c.4802 C>A	p.Ser1601Ter	Het	AD	Maternal	Pathogenic	Intellectual developmental disorder, autosomal dominant 61 (#618,009)

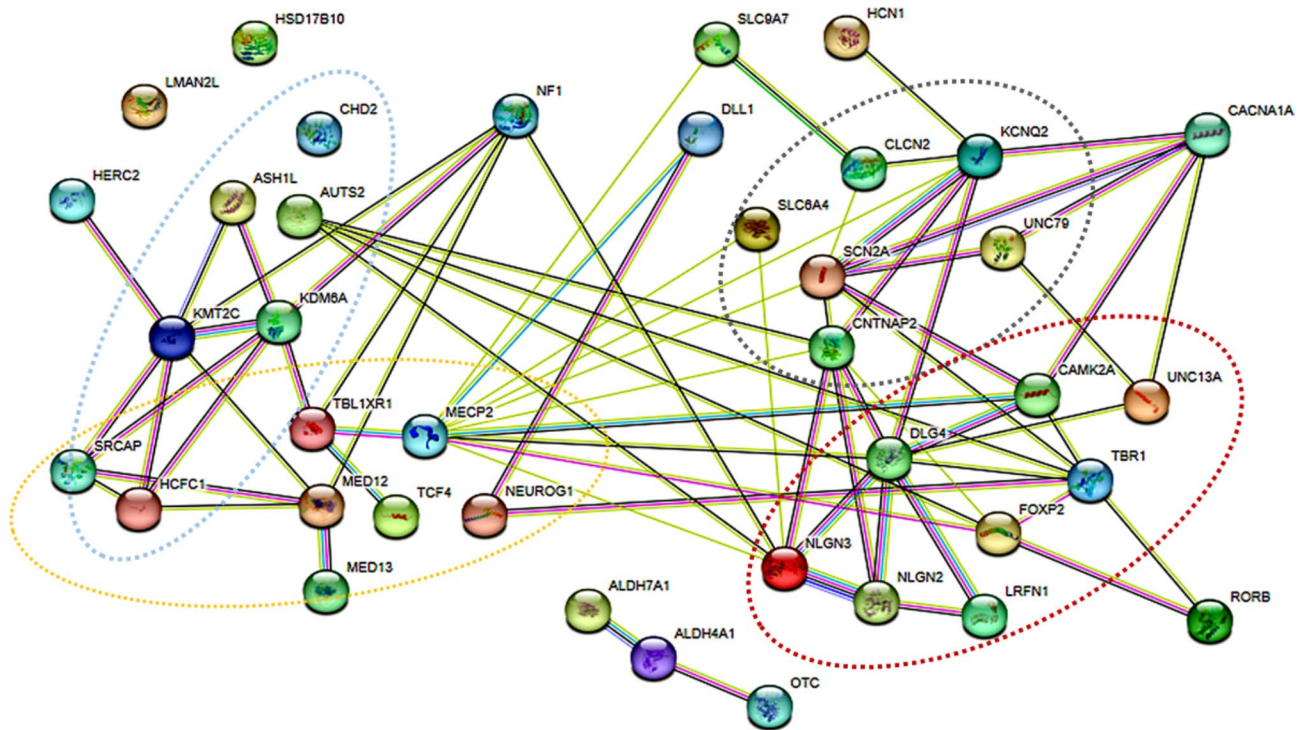
Table 3 (continued)

Sr No	Case ID	Chr	Genomic No	Ref allele	Alt allele	Gene (OMIM ID)	Transcript ID	Exon	Variant	Protein change	Zygoty	Mode of inheritance	Inherited from	Variant classification	OMIM Disease (OMIM ID)
19	ASD-064	X	153,296,882	G	A	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	3	c.433 C>T	p.Arg145Cys	Het	XLD	De novo	Pathogenic	Rett syndrome (#312,750)
20	ASD-067	X	153,296,777	G	A	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	3	c.538 C>T	p.Arg180Ter	Mosaic, somatic variant	XLD	De novo; post-zygotic event	Pathogenic	Rett syndrome (#312,750)
21	ASD-072	X	70,389,622	T	G	NLG3 (*300,336)	ENST00000358741.3 NM_181303.1	8	c.2222T>G	p.Leu741Arg	Hemi	XL	De novo	Likely pathogenic	{Autism susceptibility, X-linked 1}
22	ASD-074	9	77,249,655	C	T	RORB (*601,972)	ENST00000376896.3 NM_006914.4	3	c.202 C>T	p.Gln68Ter	Het	AD	Paternal (affected father)	Pathogenic	{Epilepsy, idiopathic generalized, susceptibility to, 15}
23	ASD-080	17	29,553,492	C	T	NFI (*613,113)	ENST00000358273.4 NM_001042492.3	18	c.2041 C>T	p.Arg681Ter	Het	AD	De novo	Pathogenic	Neurofibromatosis, type 1; (#162,200)
24	ASD-082	7	70,231,162	G	A	AUTS2 (*607,270)	ENST00000342771.4 NM_015570.4	9	c.1531G>A	p.Gly511Arg	Het	AD	De novo	Likely pathogenic	Neurofibromatosis, Noonan syndrome (#601,321)
															Intellectual developmental disorder, autosomal dominant 26 (#615,834)

Table 3 (continued)

Sr No	Case ID	Chr	Genomic No	Ref allele	Alt allele	Gene (OMIM ID)	Transcript ID	Exon	Variant	Protein change	Zygosity	Mode of inheritance	Inherited from	Variant classification	OMIM Disease (OMIM ID)
25	ASD-084	2	97,377,437	C	T	LMAN2L (*609,552)	ENST00000377079.4 NM_001142292.2	7	c.773G>A	p.Arg258His	Hom	AR	Maternal & Paternal	Likely pathogenic	? Intellectual developmental disorder, autosomal recessive 52 (#616,887)
26	ASD-085	5	125,887,798	G	A	ALDH7A1 (*107,323)	ENST00000409134.3 NM_0011182.5	14	c.1232 C>T	p.Pro411Leu	Hom	AR	Maternal & Paternal	Likely pathogenic	Epilepsy, pyridoxine-dependent (#266,100)
27	ASD-105	X	153,296,363	G	A	MECP2 (*300,005)	ENST00000453960.2 NM_001110792.2	3	c.952 C>T	p.Arg318Cys	Het	XL	De novo	Pathogenic	Autism susceptibility, X-linked 3 (#300,496)

Het: heterozygous; Hom: homozygous; Hemi: hemizygous; Comp het: Compound heterozygous; AD: autosomal dominant; AR: autosomal recessive; XL: X-linked; UNK: unknown  
\* Paternal DNA sample was unavailable and the variant was not detected in maternal sample



**Fig. 1** STRING network analysis show genes involved in synaptic junction formation (dark red), signal transduction (grey), transcription regulation (orange) and histone modification (light blue)

there are technical and interpretation limitations to the identification and prioritization of variants which were classified as VUS. Delineation of pathogenicity of these variants is often challenging because of their incomplete penetrance, variable expressivity and/or sex specific bias [65]. This however would require re-assessment of WES data every 2–3 years as per the consensus statement by Srivastava et al. using updated datasets and new computational tools [16]. Lastly, WES and CMA due to their inherent technical limitations are unable to resolve complex structural re-arrangements (e.g. inversions and translocations) which could play role in the pathogenesis of NDD [66], although, newer genomic technologies such as long-read whole genome sequencing could help to assess their role in the etiology of ASD.

## Conclusion

Data from large scale genomic and transcriptomic studies have helped to delineate the genetic architecture of ASD in European/ non-Hispanic white populations. To the best of our knowledge, this is the first study to delineate the genetic architecture of ASD in the Indian population, with *de novo* variants in genes involved in synaptic formation, transcription and its regulation, ubiquitination and chromatin remodeling as the primary cause. In congruence with data from other ethnic populations, the current study provides evidence supporting

the implementation of WES as the first-tier test in the genetic diagnosis of ASD.

## Abbreviations

ASD	autism spectrum disorder
NDD	neurodevelopmental disorder
DSM-5	Diagnostic and statistical manual – 5th edition
CMA	chromosomal microarray
WES	whole exome sequencing
SNV	single nucleotide variant
CNV	copy number variant
VUS	variants of uncertain significance
PPI	protein-protein interactions
SFARI	Simons Foundation Autism Research Initiative
AutDB	Autism Database

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12883-023-03341-0>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5

## Acknowledgements

We thank the families for their support and participation in the study.

## Author contributions

Frenny Sheth and Harsh Sheth had full access to the data in the study and take responsibility for its integrity and accuracy of data analysis. Study design was



carried out by Harsh Sheth, Frenny Sheth and Deepika Jain. Patient acquisition and recruitment was carried out by Deepika Jain, Frenny Sheth, Harshkumar Patel, Siddharth Shah, Anand S Iyer, Ketan Patel, Dhaval Solanki, Sanjiv Mehta, Priti Mhatre, Shruti Bajaj, Vishal Patel, Manoj Pandya, Deepak Dhami, Bhargavi Menghani, Darshan Patel and Jayesh Sheth. Data generation, analysis and interpretation was carried out by Jhanvi Shah and Harsh Sheth. Drafting of the manuscript was carried out by Frenny Sheth, Jhanvi Shah, Jayesh Sheth and Harsh Sheth. Statistical analysis was carried out by Jhanvi Shah and Harsh Sheth. Critical evaluation of the manuscript was carried out by all authors. Study supervision and funding acquisition was carried out by Frenny Sheth, Jayesh Sheth, Deepika Jain and Harsh Sheth.

#### Funding

This work was supported by the Gujarat State Biotechnology Mission grant GSBTM/JDR&D/608/2020/456–458 (Frenny Sheth, Deepika Jain and Harsh Sheth). Furthermore, Jhanvi Shah was supported by the CSIR-NET fellowship (09/1331(0001)/2021-EMR-I). The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

#### Data Availability

Datasets supporting the conclusions of this article are available on the EGA website (European Genome-Phenome Archive) under the title “Genetic architecture of autism spectrum disorders in India”. To access whole exome sequencing data, the study ID is EGAS00001006060 and to access chromosomal microarray data, the study ID is EGAS00001006439. Datasets can be accessed from the EGA website using the following weblink: <https://ega-archive.org/>.

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the research ethics committee at the Foundation for Research in Genetics and Endocrinology, Ahmedabad (ID: FRIGE/IEC/19/2020). A written informed consent as per the Helsinki Declaration was obtained from the parents and guardians of all probands. All the methods in the study were carried out as per the Helsinki Declaration. The entire dataset presented here does not contain any identifiable information.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>FRIGE's Institute of Human Genetics, Ahmedabad, India

<sup>2</sup>Shishu Child Development and Early Intervention Centre, Ahmedabad, India

<sup>3</sup>Royal Institute of Child Neurosciences, Ahmedabad, India

<sup>4</sup>Zydus Hospital, Ahmedabad, India

<sup>5</sup>Specialty Homeopathic Clinic, Ahmedabad, India

<sup>6</sup>Mantra Child Neurology and Epilepsy Hospital, Bhavnagar, India

<sup>7</sup>NeuroKids Clinic, Ahmedabad, India

<sup>8</sup>Children's Institute for Development and Advancement Centre, Vadodara, India

<sup>9</sup>Tender Kinds Centre for Child Development, Navi Mumbai, India

<sup>10</sup>The Purple Gene Clinic, Mumbai, India

<sup>11</sup>Little Brain Pediatric Neurocare Centre, Vadodara, India

<sup>12</sup>Kadam Maternity Home, Ahmedabad, India

<sup>13</sup>Axon Child Neurology and Epilepsy Centre, Rajkot, India

<sup>14</sup>Charotar Institute of Paramedical Sciences, Charotar University of Science and Technology, Changa, India

#### References

1. Arteche-López A, Gómez Rodríguez MJ, Sánchez Calvin MT, Quesada-Espinosa JF, Lezana Rosales JM, Palma Milla C, et al. Towards a change in the diagnostic algorithm of Autism Spectrum Disorders: evidence supporting whole Exome sequencing as a first-tier test. *Genes (Basel)*. 2021;12(4):560.
2. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular Diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in Children with Autism Spectrum Disorder. *JAMA*. 2015;314(9):895–903.
3. Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. Arlington, VA, US: American Psychiatric Publishing, Inc.; 2013. xlv, 947 p. (Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed).
4. Vahia VN. Diagnostic and statistical manual of mental disorders 5: a quick glance. *Indian J Psychiatry*. 2013;55(3):220–3.
5. Chauhan A, Sahu JK, Jaiswal N, Kumar K, Agarwal A, Kaur J, et al. Prevalence of autism spectrum disorder in Indian children: a systematic review and meta-analysis. *Neurol India*. 2019;67(1):100–4.
6. Hossain MD, Ahmed HU, Jalal Uddin MM, Chowdhury WA, Iqbal MS, Kabir RI, et al. Autism spectrum disorders (ASD) in South Asia: a systematic review. *BMC Psychiatry*. 2017;17(1):281.
7. Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, et al. The epidemiology of autism spectrum disorders. *Annu Rev Public Health*. 2007;28:235–58.
8. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The changing epidemiology of Autism Spectrum Disorders. *Annu Rev Public Health*. 2017;38:81–102.
9. Tick B, Bolton P, Happé F, Rutter M, Rijdsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry*. 2016;57(5):585–95.
10. Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Sci*. 2013;15(5):399–407.
11. Marchuk DS, Crooks K, Strande N, Kaiser-Rogers K, Milko LV, Brandt A, et al. Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *PLoS ONE*. 2018;13(12):e0209185.
12. Testard Q, Vanhoye X, Yauy K, Naud ME, Vieville G, Rousseau F, et al. Exome sequencing as a first-tier test for copy number variant detection: retrospective evaluation and prospective screening in 2418 cases. *J Med Genet*. 2022;59(12):1234–40.
13. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*. 2003;33 Suppl:228–37.
14. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*. 2012;20(5):490–7.
15. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*. 2009;106(45):19096–101.
16. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med*. 2019;21(11):2413–21.
17. Álvarez-Mora MI, Sánchez A, Rodríguez-Revenga L, Corominas J, Rabionet R, Puig S, et al. Diagnostic yield of next-generation sequencing in 87 families with neurodevelopmental disorders. *Orphanet J Rare Dis*. 2022;17:60.
18. Gautam A. Isolation of DNA from Blood Samples by Salting Method. In: Gautam A, editor. *DNA and RNA Isolation Techniques for Non-Experts [Internet]*. Cham: Springer International Publishing; 2022. p. 89–93. Available from: [https://doi.org/10.1007/978-3-030-94230-4\\_12](https://doi.org/10.1007/978-3-030-94230-4_12).
19. Rajan-Babu IS, Chong SS. Triplet-repeat primed PCR and Capillary Electrophoresis for characterizing the Fragile X Mental Retardation 1 CGG repeat hyperexpansions. *Methods Mol Biol*. 2019;1972:199–210.
20. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–43.
21. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(Database issue):D986–92.
22. Firth HV, Richards SM, Bevan AP, Clayton S, Corpes M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524–33.

Received: 18 May 2023 / Accepted: 21 July 2023

Published online: 05 August 2023

23. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33(Database issue):D514–517.
24. Thaxton C, Good ME, DiStefano MT, Luo X, Andersen EF, Thorland E, et al. Utilizing ClinGen gene-disease validity and dosage sensitivity curations to inform variant classification. *Hum Mutat.* 2022;43(8):1031–40.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
26. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, der Auwera GAV et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* 2017;201178.
27. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling - PMC [Internet]. [cited 2022 Dec 12]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3476336/>.
28. Smedley D, Jacobsen JOB, Jager M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc.* 2015;10(12):2004–15.
29. Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* 2020;49(D1):D1207–17.
30. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
31. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248–9.
32. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812–4.
33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11(4):361–2.
34. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res.* 2009;19(9):1553–61.
35. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310–5.
36. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C, MetaDome. Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat.* 2019;40(8):1030–8.
37. ClinVar: improving access to variant interpretations and supporting evidence - PMC [Internet]. [cited 2022 Dec 12]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753237/>.
38. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, et al. Human gene mutation database (HGMD®): 2003 update. *Hum Mutat.* 2003;21(6):577–81.
39. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47(D1):D506–15.
40. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv.* 2021;7(31):eab2169.
41. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13.
42. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ. Mouse Genome Database (MGD): knowledgebase for mouse–human comparative biology. *Nucleic Acids Res.* 2020;49(D1):D981–7.
43. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178–92.
44. Biesecker LG, Harrison SM. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Sci.* 2018;20(12):1687–8.
45. Dagli A, Buiting K, Williams CA. Molecular and clinical aspects of Angelman Syndrome. *Mol Syndromol.* 2012;2(3–5):100–12.
46. Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am J Hum Genet.* 2006;79(2):370–7.
47. Yatsenko SA, Cheung SW, Scott DA, Nowaczyk MJM, Tarnopolsky M, Naidu S, et al. Deletion 9q34.3 syndrome: genotype-phenotype correlations and an extended deletion in a patient with features of Opitz C trigonocephaly. *J Med Genet.* 2005;42(4):328–35.
48. Fichera M, Barone R, Grillo L, De Grandi M, Fiore V, Morana I, et al. Familial 1q22 microduplication associated with psychiatric disorders, intellectual disability and late-onset autoimmune inflammatory response. *Mol Cytogenet.* 2014;7(1):90.
49. Shah J, Patel H, Jain D, Sheth F, Sheth H. A rare case of a male child with post-zygotic de novo mosaic variant c.538 C > T in MECP2 gene: a case report of Rett syndrome. *BMC Neurol.* 2021;21:469.
50. Sheth F, Shah J, Patel K, Patel D, Jain D, Sheth J, et al. A novel case of two siblings harbouring homozygous variant in the NEUROG1 gene with autism as an additional phenotype: a case report. *BMC Neurol.* 2023;23(1):20.
51. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Mol Autism.* 2018;9:64.
52. Dhaliwal J, Qiao Y, Calli K, Martell S, Race S, Chijiwa C et al. Contribution of multiple inherited Variants to Autism Spectrum disorder (ASD) in a family with 3 affected siblings. *Genes (Basel).* 2021;12(7).
53. Tuncay IO, Parmalee NL, Khalil R, Kaur K, Kumar A, Jimale M, et al. Analysis of recent shared ancestry in a familial cohort identifies coding and noncoding autism spectrum disorder variants. *NPJ Genom Med.* 2022;7(1):13.
54. De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature.* 2014;515(7526):209–15.
55. Rossi M, El-Khechen D, Black MH, Farwell Hagman KD, Tang S, Powis Z. Outcomes of diagnostic exome sequencing in patients with diagnosed or suspected Autism Spectrum Disorders. *Pediatr Neurol.* 2017;70:34–43e2.
56. Gibitova EA, Dobrynin PV, Pomerantseva EA, Musatova EV, Kostareva A, Evsyukov I et al. A study of the genomic Variations Associated with Autistic Spectrum Disorders in a russian cohort of patients using whole-exome sequencing. *Genes (Basel).* 2022;13(5).
57. Doan RN, Lim ET, De Rubeis S, Betancur C, Cutler DJ, Chiochetti AG, et al. Recessive gene disruptions in autism spectrum disorder. *Nat Genet.* 2019;51(7):1092–8.
58. Iakouchava LM, Muotri AR, Sebat J. Getting to the Cores of Autism. *Cell.* 2019;178(6):1287–98.
59. Rodríguez-Palmero A, Boerigter MM, Gómez-Andrés D, Aldinger KA, Marcos-Alcalde I, Popp B, et al. DLG4-related synaptopathy: a new rare brain disorder. *Genet Sci.* 2021;23(5):888–99.
60. Lipstein N, Verhoeven-Duif NM, Michelassi FE, Calloway N, van Hasselt PM, Pienkowska K, et al. Synaptic UNC13A protein variant causes increased neurotransmission and dyskinetic movement disorder. *J Clin Invest.* 2017;127(3):1005–18.
61. Bayat A, Liu Z, Luo S, Fengler CD, Højte AF, Isidor B, et al. A new neurodevelopmental disorder linked to heterozygous variants in UNC79. *Genet Med.* 2023;25(9):100894.
62. Narcisa V, Disenza M, Vaccari E, Rosen-Sheidley B, Hardan AY, Couchon E. Parental interest in a genetic risk assessment test for autism spectrum disorders. *Clin Pediatr (Phila).* 2013;52(2):139–46.
63. Harrington JW, Emuren L, Restaino K, Schrier Vergano S. Parental perception and participation in genetic testing among children with Autism Spectrum Disorders. *Clin Pediatr (Phila).* 2018;57(14):1642–55.
64. Glaze DG, Neul JL, Kaufmann WE, Berry-Kravis E, Condon S, Stoms G, et al. Double-blind, randomized, placebo-controlled study of trofinetide in pediatric Rett syndrome. *Neurology.* 2019;92(16):e1912–25.
65. Wigdor EM, Weiner DJ, Grove J, Fu JM, Thompson WK, Carey CE, et al. The female protective effect against autism spectrum disorder. *Cell Genomics.* 2022;2(6):100134.
66. D’haene E, Vergult S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet Sci.* 2021;23(1):34–46.

## Publisher’s Note

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