Scientific Reports

https://doi.org/10.1038/s41598-026-44288-9

Article in Press

Autism spectrum disorder trios from consanguineous populations are enriched for rare homozygous variants, identifying 32 new candidate genes

Received: 12 June 2025

Accepted: 10 March 2026

Published online: 21 March 2026

Cite this article as: Harripaul R., Rabia A., Vasli N. *et al.* Autism spectrum disorder trios from consanguineous populations are enriched for rare homozygous variants, identifying 32 new candidate genes. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-44288-9

Ricardo Harripaul, Ansa Rabia, Nasim Vasli, Anna Mikhailov, Ashlyn Rodrigues, Stephen F. Pastore, Tahir Muhammad, Thulasi Thiruvallur Madanagopal, Aisha Nasir Hashmi, Clinton Tran, Cassandra Stan, Katherine Aw, Clement C. Zai, Maleeha Azam, Saqib Mahmood, Abolfazl Heidari, Raheel Qamar, Leon French, Shreejoy Tripathy, Zehra Agha, Muhammad Iqbal, Majid Ghadami, Susan L. Santangelo, Bita Bozorgmehr, Laila Al Ayadhi, Roksana Sasanfar, Shazia Maqbool, Arsalan Hassan, James A. Knowles, Muhammad Ayub & John B. Vincent

Autism spectrum disorder trios from consanguineous populations are enriched for rare homozygous variants, identifying 32 new candidate genes

Running title: Homozygous variants for autism and consanguinity

Ricardo Harripaul[1,2]§, Ansa Rabia[1,3,4]§, Nasim Vasli[1], Anna Mikhailov[1], Ashlyn Rodrigues[1], Stephen F. Pastore[1], Tahir Muhammad[1,2], Thulasi Thiruvallur Madanagopal[1], Aisha Nasir Hashmi[1], Clinton Tran[1], Cassandra Stan[1], Katherine Aw[1], Clement C. Zai[1,2], Maleeha Azam[5], Saqib Mahmood[3,6], Abolfazl Heidari[7], Raheel Qamar[5,8], Leon French[2,9], Shreejoy Tripathy[2,9], Zehra Agha[5,10], Muhammad Iqbal[11], Majid Ghadami[12], Susan L. Santangelo[13,14], Bita Bozorgmehr[1], Laila Al Ayadhi[15], Roksana Sasanfar[16], Shazia Maqbool[17], Arsalan Hassan[18,19,20,21], James A. Knowles[22], Muhammad Ayub[23,24]*, John B Vincent[1,2,25]*

[1]Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada;

[2]Institute of Medical Science, University of Toronto, Toronto, ON, Canada;

[3]University of Health Sciences, Lahore, Pakistan;

[4]CMH Lahore Medical College & Institute of Dentistry Lahore, Pakistan;

[5]Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan;

[6]NUR International University, Lahore, Pakistan;

[7]Sana Medical Genetics Laboratory, Qazvin, Iran; Reference Laboratory of Qazvin Medical University, Qazvin, Iran;

[8]Science and Environment Sector, ICESCO, Rabat, Kingdom of Morocco;

[9]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada;

[10]Department of Psychiatry, State University of New York at Buffalo; Jacobs School of Medicine and Biomedical Sciences, Buffalo, NY, USA;

[11]Department of Physiology, College of Medicine, King Saud University, Riyadh, Saudi Arabia;

[12]Department of Educational Sciences, Farhangian University, Tehran, Iran;

[13]Center for Clinical and Translational Research, Maine Health Institute for Research, Scarborough ME, 04074, USA;

[14]Tufts University School of Medicine, Department of Psychiatry, Boston, MA, 02110, USA;

[15]Autism Research & Treatment Center, Department of Physiology, Faculty of Medicine, National Plan for Science and Technology (NPST), King Saud University, Riyadh, Saudi Arabia;

[16]Department of Psychiatry, Harvard Medical School, Boston, MA, USA;

[17]The University of Child Health Sciences, Children's Hospital, Lahore, Pakistan;

[18]Broad Institute of MIT and Harvard, Cambridge, MA, USA;

[19]Analytic & Translational Genetics Unit, Department of Medicine, Massachussets General Hospital, Boston, MA, USA;

[20]University of Peshawar, Peshawar, Khyber Pakhtunkhwa, Pakistan;

[21]Institute of Omics and Health Research, Lahore, Pakistan;

[22]Human Genetics Institute of New Jersey and Department of Genetics, Rutgers University, Piscataway, NJ, USA;

[23]Department of Psychiatry, Queens University Kingston, Kingston, ON, Canada;

[24]Division of Psychiatry, University College London, London, UK;

[25]Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

§Equal contribution

*Correspondence should be addressed to:

Dr. John B. Vincent, Brain Molecular Science, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health (CAMH), 250 College Street, Toronto, ON, Canada, M5T 1R8; Email: john.vincent@camh.ca

Dr. Muhammad Ayub, Division of Psychiatry, University College London, 149 Tottenham Ct Rd, London W1T 7NF, United Kingdom. Email: m.ayub@ucl.ac.uk

Biallelic variants for autism....

Biallelic variants for autism....

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder (NDD) that affects about 1 in 54 children worldwide, imposing enormous economic and socioemotional burden on families and communities. Genetic studies of ASD have identified *de novo* copy number variants (CNVs) and point mutations that contribute significantly to the genetic architecture, but the majority of these studies were conducted in populations unsuited for detecting autosomal recessive (AR) inheritance. However, several ASD studies in consanguineous populations point towards AR as an under-appreciated source of ASD variants.

Methods: We used whole exome sequencing to look for rare variants for ASD in 115 proband-mother-father trios from populations with high rates of consanguinity, namely Pakistan, Iran, and Saudi Arabia. Consanguinity was assessed through microarray genotyping.

Results: We report 77 candidate single nucleotide variants and indels, with 62% homozygous, 22% autosomal dominant/*de novo*, and 16% X-linked, in 55 trios. 56% of the variants were loss of function (LoF) or putative LoF (pLoF), and 44% nonsynonymous. We found an enrichment of homozygous variants, both in 16 genes previously reported for AR ASD and/or intellectual disability (ID) and 32 previously unreported AR candidate genes (including *DAGLA*, *ENPP6*, *FAXDC2*, *ILDR2*, *KSR2*, *PKD1L1*, *SCN10A*, *SHH*, and *SLC36A1*). We also identified seven candidate homozygous exonic loss CNVs.

Biallelic variants for autism....

Conclusions: The significant enrichment for homozygous variants among individuals with high F_{roh} coefficients, compared with low F_{roh}, either in known or candidate AR genes, confirms that genetic architecture for ASD among consanguineous populations is different to non-consanguineous populations. Assessment of consanguinity may assist in the genetic diagnostic process for ASD.

Keywords: Autism spectrum disorder; Trio; Autosomal recessive; Consanguinity; Homozygous; Candidate Genes

Introduction

Autism spectrum disorder (ASD) is characterized by deficits in social communication, repetitive/restricted behaviours and interests. Apart from a small percentage of individuals who recover (1), ASD is a life-long condition and only about 20% have good outcomes, as reported in a recent meta-analysis of longitudinal studies (2). Phenotypically, individuals with ASD are heterogeneous in their level of intellectual functioning, and co-morbid psychiatric and behavioural problems. These factors play an important role in the long term end points of the disease (3). In a systematic review the prevalence of ASD was estimated to be 0.76% internationally (4). In another systematic review of international data, the median estimate was 62/10,000 (5). In this review, the ASD prevalence showed little variation by geographic region, ethnic, cultural and socioeconomic factors (5).

The reported prevalence of ASD has increased with time. In the 1960s it was estimated to affect as few as 1 in 10,000 individuals (5), prevalence studies from the 1980s suggested that as many as 72 in 10,000 individuals had ASD, rising to 1% in the 2000s (5,6). More recent studies report prevalence rates of more than 2%. (7–9). In the United States the prevalence across eleven sites in 2016 was 18.5 per 1,000 (one in 54) children (10). This increase is partly due to changes in diagnostic criteria, reporting practices, and increased awareness (11–14).

A systematic review of literature from South Asian countries reported a prevalence of ASD in the range of 0.09-1.07 % (15). To date, there has been no community study of ASD prevalence in Pakistan. Studies from child psychiatry clinics reported rates of 2.4% and 3.2%, 4.5%, 5% (16–19). Clinical services and special schools for ASD are available in Pakistan, but are typically limited to major cities (20). The literature on clinical presentation is limited. One study with a small sample size shows symptoms consistent with studies elsewhere in the world (21).

Consanguineous marriages lead to a marked increase in the frequency of rare recessive disorders (22,23). Many genetic variants will only be pathogenic in recessive form; this includes variants for ID, and almost certainly for ASD too. Populations with a high proportion of consanguinity have been important for describing autosomal recessive genes in ID in Pakistan (24,25), Iran (26,27), Syria (28), and Saudi Arabia (29), yet, to date, relatively few reports of AR genes underlying ASD in consanguineous populations have been published (30–32). Previous work has highlighted recessive inheritance as an important component of the genetic architecture of ASD. For instance, a large study of consanguineous versus non-consanguineous families in India concluded that consanguinity increases the risk for ASD with an odds ratio of 3.22 (33). Morrow and colleagues used SNP microarrays to map homozygous loci in 104 small ASD families from the

Middle East, Turkey and Pakistan, finding homozygous deletions implicating *SLC9A9*, *PCDH10*, *CNTN3* and others (30).

Research in outbred populations also supports AR inheritance as an important piece of the genetic puzzle for ASD. For example, it has been estimated that loss-of-function (LoF) recessive mutations contribute 3% of ASD genetics in two US-based case-control cohorts(34,35). These findings are not limited to population isolates or ethnic subgroups (36,37). Identification of recessive genes in outbred populations is problematic, as the analysis pipelines for WES/WGS data are non-optimal for discovering compound heterozygous mutations.

Despite the higher prevalence of recessive inheritance in consanguineous populations, data from studies of developmental disorder (DD) suggest that *de novo* variants are also prominent, albeit in a lower proportion e.g. in a UK study of 6040 families from the Deciphering Developmental Disorders (DDD) study, individuals of Pakistani ancestry had 29.8% *de novo* compared with 49.9% in the European ancestry UK population (38). In an Iranian ID study, the *de novo* rate was 27.86% (39).

The genetic involvement in ASD has been clearly established through twin, family and adoption studies (40), with heritability estimates from 50% to over 90% (41–43). The heritability of ASD is partitioned across both rare and common variants in hundreds of genes, each of which explains only a fraction of the disorder's heritability (44). ASD etiology is likely the result of a

high degree of genetic heterogeneity with many rare variants contributing to the disorder resulting in seemingly heterogeneous clinical manifestations. Describing the genetic architecture of ASD is an important step towards understanding the pathogenesis of the disorder. There has been significant progress to date, and studies of ASD have shown that *de novo* copy number variants (CNVs) and point mutations play a prominent role, however, the majority of these studies have been conducted in outbred populations (45). Here, we report on a study of 83 genetic and 10 copy number variants (seven homozygous; 3 heterozygous) from a cohort of 115 ASD trios from three countries with a high frequency of consanguineous marriages (Pakistan, Iran, and Saudi Arabia).

Materials and Methods

Trio Family Ascertainment

Institutional Research Ethics Board approval was received from the Centre for Addiction and Mental Health (CAMH) and the recruitment sites. All the subjects were informed about the study objectives, and written consent was obtained from the subjects and/or their legal guardian(s). All procedures and methods performed were carried out in accordance with relevant guidelines and regulations. A summary of the cohorts, collaborator contributions, ascertainment, and assessment tools, is in Table 1. Overall, we have collected DNA from 115 trios: 62 trios from Pakistan, 40 trios from Iran, and 13 trios from Saudi Arabia, comprised of 345 individuals. All families are

simplex, except for Saudi family SA1, for which a cousin with ASD is known, but for whom no sample is available.

Whole Exome Sequencing, Alignment and Variant Calling

Whole Exome Sequencing (WES) was performed on all samples using the Thruplex DNA-Seq (Rubicon Genomics) Library Preparation Kit with the Agilent SureSelect V5 Exome Capture kit. DNA was sheared using a Covaris ME220 Focused-Sonicator 200 bp, which was verified using a Agilent 2100 Bioanalyzer System for fragment length distribution and quantification. All trios were sequenced on Illumina HiSeq 2500 or NovaSeq sequencing systems. Details of read QC, alignment, variant calling and prioritization are provided in Supplementary Materials, and summarized in Figure 1. Fifteen trios were previously run using the Ion Proton platform/Ion Ampliseq™ Exome kit (Life Technologies)- all were re-run using the Illumina platform except trio IABB3, for which a causative frameshifting deletion was detected in the VPS13B/Cohen syndrome gene.

Microarray genotyping and analysis.

Microarrays were run using DNA from 104 of the 115 ASD probands (where sufficient DNA was available), including 13 using Affymetrix CytoScanHD, 87 using Illumina CoreExome, and four with both arrays. CoreExome data in PLINK format was used for runs-of-homozygosity (RoH) analysis, using PLINK 1.9 (http://pngu.mgh.harvard.edu/purcell/plink). RoH data was used to generate an F-coefficient of consanguinity for the probands (and affected

siblings), F_{RoH}, as described in McQuillan et al, 2008 (46), where $F_{roh} = \sum L_{roh}/L_{auto}$. Autosome size L_{auto} was estimated at 2,673,768 Kb**.** The CytoScanHD data was analysed using Chromosome Analysis Suite (ChAS), and a loss of heterozygosity (LoH) score was generated. For four samples that were run on both Illumina CoreExome and Affymetrix CytoScanHD arrays (probands from trios IABB2, IABB3, IABB4, and IABB5), we were able to generate both F_{roh} and LoH scores, which, as the relationship was almost linear (Pearson $R^2 = 0.9971$) allowed us to convert LoH scores from CytoScanHD microarrays to F_{roh} (Supplementary Materials). Comparison statistics were performed using GraphPad (https://www.graphpad.com/).

Copy Number Variant (CNV) analysis

CNV analysis of the CytoScanHD array data was performed using ChAS and PennCNV; for CoreExome data by the Illumina Genome Suite/CNVpartition and PennCNV. Evidence of homozygous loss CNVs was cross-references with WES data, using the Integrated Genome Viewer v2.3.5 (IGV: https://software.broadinstitute.org/software/igv/; (47), and CNVs corroborated in this manner were then checked by PCR. Further analysis/prioritization/validation of CNVs, including using WES data, is given in Supplementary Materials.

Cross-referencing with other datasets

In order to evaluate the strength of candidacy of the variants/genes identified here or to find supporting evidence, we cross-referenced our

findings with other datasets listed below. We identified rare variants in the candidate genes we report here (Tables 2-4) that were either: 1. *De novo*, 2. Heterozygous LoF but status either de novo or unknown; 3. Biallelic (i.e. homozygous); 4. Putative biallelic (possibly compound heterozygous, but phasing unknown); 5. X-linked in males, maternally inherited. Data are shown in Supplementary File 1.

MSSNG: More than 13,000 individuals from the Autism Genetic Research Exchange (AGRE) repository and other cohorts have been whole-genome sequenced, through Autism Speaks, and data made available through the MSSNG database (https://research.mss.ng; db7 (48). MSSNG data was accessed in May 2024.

SFARI: SSC biallelic: The Simons Simplex Collection (SSC) includes 2,600 simplex autism or related developmental disorder families. WES data is available for ~2,500 of these families, biallelic variants for our gene list were searched using the GPF browser (https://gpf.sfari.org/gpf19/datasets/SSC/browser). Putative compound heterozygous variants criteria for minor allele frequency (MAF) in SSC exome, gnomAD exome and genome frequencies, of <0.001, or, for missense variants, MCP scores >1 and CADD scores >18. **SVIP biallelic:** the autism Simons Variation in Individuals Project (SVIP) dataset was searched for putative biallelic variants through the browser gpf.sfari.org/gpf19/datasets/SVIP/browser. Putative compound heterozygous

variants checked for criteria including minor allele frequency (MAF) in SSC exome, gnomAD exome and genome frequencies, of <0.001, or, for missense variants, MCP scores >1 and CADD scores >18.

Deciphering Developmental Disorders (DDD) study: includes sequence data for ~14,000 children; this dataset was searched for variants in genes overlapping with our study through the known developmental genes list (https://decipher.sanger.ac.uk/ddd#ddgenes), as well as the list of research variants (https://decipher.sanger.ac.uk/ddd#research-variants), which are 2,723 variants of unknown significance from 4,293 children. For *de novo* variants, all are constitutive, unless specifically noted as mosaic here. CNV variants are not included here.

Autism Sequencing Consortium (ASC): genes with *de novo* variants, either in the control set or case set, from ASC exome analysis browser (https://asc.broadinstitute.org/). This dataset lists numbers of *de novo* protein-truncating variants, as well as *de novo* missense variants with MCP scores either 1-2 or ≥2. The ASC dataset includes 6,430 probands, and includes the SSC data.

ASC+SPARK *de novo* : *de novo* mutations from primary data set, N=47,061 autistic individuals (from Table S7), accessed 14 April 2025(49).

Biallelic variants for autism….

SPARK, GeneDx, DDD13K and others, *de novo* (Wang et al, 2022): *de novo* mutations, N=46,612 autistic and NDD trios (from Dataset S1), accessed 14 April 2025 (50).

DeRubeis: genes with *de novo* variants reported in the DeRubeis et al, 2014 study, with either LoF or missense variants (51). As subject IDs were not reported, overlap with other studies recorded here is not possible.

AutismKB: additional studies where variants have been reported for the genes are taken from the Autism Knowledge Base (http://db.cbi.pku.edu.cn/autismkb_v2/quick_search.php), (52).

SysNDD: all candidate genes, recessive, dominant or X-linked, were cross-referenced with the SysNDD database of curated neurodeveleopment disorder (NDD) genes (sysndd.dbmr.unibe.ch; accessed 14 April 2025).

ASD in other consanguineous populations: we compared the candidate genes/variants with those reported from similar studies of ASD in other populations enriched for consanguinity, including from Qatar (53–55), Oman (56), Iranian (57), Pakistani (58), and Middle Eastern(59) populations. We also report allele frequencies in control populations from South Asia/Pakistan and Middle East using gnomAD v4.1.0 (N~45,000 South Asian; N~3,000 Middle Eastern; accessed Mar 2025), also from Regeneron (N~16,000 Pakistani; N~3,900 Middle Eastern; rgc-research.regeneron.com/me/home; accessed Mar 2025), and from our GenSCRIP study (N=8,430 Pakistani

controls; accessed May 2025). For an Iranian control population we used the Iranian variome, “Iranome” (iranome.com; accessed Mar 2025), with N=1200 healthy controls (see Supplementary File 1).

Neuroanatomical Enrichment Analysis

The Allen Adult Human Brain Atlas (60) , Brainspan (61) , and a single-cell atlas of the mouse nervous system (62) were used to test for neuroanatomically specific expression. For each gene, expression was standardized across regions or cell types. For these compartments, each gene was then ranked from most specific to most depleted. The area under the receiver operating statistic (AUC) was used to quantify specific expression for the genes of interest within a region or cell cluster. The Mann-Whitney U test was used to test statistical significance, and the Benjamini–Hochberg procedure (63) was used to correct for multiple tests. In addition, in order to explore whether our candidate genes may contribute to electrophysiological and morphological phenotypes in neurons we cross-referenced gene expression correlates with electrophysiological phenotypes through a dataset published by Bomkamp et al, 2019 (64).

Results

We ascertained, phenotyped, collect blood and/or saliva for DNA extraction, and performed whole exome sequencing (mean depth: 27.29; average % coverage at 30X=31.01; average % coverage at 100X=7.32; see

Supplementary File 1) from 115 ASD proband (91M:24F)/mother/father trios (62 from Pakistan, 40 from Iran, and 13 from Saudi Arabia), for a total of 345 subjects (Table1). In addition, microarrays were run on 104 of the ASD probands, providing data for CNV and ROH analyses.

Whole Exome Sequencing: We discovered 77 SNVs or indels through our analysis pipeline, including 48 homozygous autosomal variants, 17 autosomal *de novo* variants, and 4 *de novo* and 8 maternally inherited X-linked variants.

Biallelic mutations (Table 2): We identified homozygous variants (7 LoF or pLoF; 9 nonsynonymous; 1 in-frame del) in 16 genes that have been previously reported to be associated with autosomal recessive non-syndromic ID, including *BTN3A2* (29), *CC2D1A* (MIM 608443; MRT3; (65)), *LINS1* (27) , *MADD* (29), *MTHFR* (28), *RSRC1* (66,67), *TECPR2* (29) and *ZNF335* (26), or for syndromic or metabolic forms of autosomal recessive ID, including *AGA* (aspartylglucosaminuria), *ASL* (arginosuccinic aciduria), *ASPA* (aspartoacylase deficiency/Canavan disease), and *HTRA2* (3-methylglutaconic aciduria). Three genes, *CC2D1A*, *DEAF1, and VPS13B* have also previously been associated with ASD or autistic features (94–96; (71)). Biallelic mutations in *DEAF1* are known to cause neurodevelopmental disorder with hypotonia, impaired expressive language, with or without seizures (MIM 617171). LoF variants were reported for new candidate genes ENPEP, *DAGLA*, *FAXDC2*, *GIMAP8*, *HRNR*, *ILDR2*, *SLC36A1*, *SCN10A*, *VPS16*,

ZNF776, and pLoF (such as canonical splice site variants) in *CNPY4*, *DENND1B*, *LRRC34*, *PKD1L1*, and *PPP1R36*. *VPS16* has previously been reported for autosomal dominant dystonia (MIM 608559), and *PKD1L1* for AR visceral heterotaxy 8 (MIM 617205). Heterozygous missense and LoF mutations in *SHH* are known to be involved in autosomal dominant holoprosencephaly 3 (HPE3; MIM 142945), and so it is somewhat surprising to find a homozygous missense variant associated with a milder phenotype and with no obvious HPE3-related dysmorphic features. The Asn69Ile variant in *SHH* substitutes an asparagine for isoleucine at a residue that is conserved across vertebrates, is predicted to be damaging, and is found in heterozygous form in only 21 out of 1.6E6 gnomAD (v4.1.0) alleles, and no homozygotes.

In one trio (trio SMPA21) we report a homozygous stop-gain variant in *DAGLA*, for which rare heterozygous variants were reported to be associated with NDDs, including autism (72). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; https://gene.sfari.org/). We also report homozygous missense variants in novel candidate genes *ANO10*, *CABP2*, *CAPSL*, *CLCA4*, *CPM*, *DNAH8*, *ERMP1*, *KSR2*, *RANBP9*, *TMEM25*, *TRIM3*, and *WDR90*, plus in *EPHB1*, for which *de novo* variants in have been reported for ASD (73), and are listed respectively as strong candidate and suggestive evidence in SFARI (scores of 2 and 3, respectively). Two of our novel

candidate AR genes, *SGSM3* and *VPS16*, have recently been confirmed in multiple NDD families in other consanguineous populations(74–77).

Although not focused on consanguineous populations, WGS data from the MSSNG study, as well as WES studies such as DDD and SFARI, support a number of our candidate AR genes (see Supplementary File 1), with homozygous rare variants in *DNAH7* (MSSNG), *DNAH8* (MSSNG), *CAPSL* (MSSNG), and *CYP2A7* (MSSNG), and putative compound heterozygous damaging rare variants in *DNAH8* (in three MSSNG individuals), *SCN10A* (DDD), and *WDR90* (DDD)(see Supplementary File 1). There was also evidence of disease-causing biallelic variants in known AR NDD genes *MTHFR* (three homozygous; DDD), *AGA* (one homozygous stop gain; DDD), *DEAF1* (two homozygous stop gain; MSSNG and DDD), *ZNF335* (one homozygous missense, DDD; one putative compound heterozygous, DDD), *ASL* (one homozygous missense; MSSNG), *CC2D1A* (one homozygous missense, MSSNG; two homozygous LoF DDD), and *VPS13B* (six homozygous, nine putative compound heterozygous).

***De novo* mutations (Table 3 & 4):** *De novo* dominant mutations have been shown to be a major causal factor in ASD etiology. We identified and validated 21 *de novo* coding variants (17 autosomal and 4 X-linked). Among these genes, variants in four of them have been identified in previous studies (*MECP2*, *MYT1L*, *SCN2A*, *ZNF292*). 13 of the 21 variants were LoF, three were putative LoF (canonical splice site), and the remainder were nonsynonymous.

The mutations include genes that have been implicated in other recent studies of ASD and/or ID, and for other trios there are variants in genes linked with other nervous system disorders, such as *DRP2* in Charcot-Marie-Tooth disease (78). Several other genes with *de novo* variants that have not been associated with ASD (or ID) previously, and may represent novel putative genes for ASD.

Our study identified a *de novo* LoF mutation in *ZNF292* (NM_015021:c.6159_6160del; p.Glu2054Lysfs*14) in two trios (IAU-65 (Iranian) and Autism-10 (Pakistani)). These two trios and 26 additional families (from multiple studies and cohorts) with mutations in *ZNF292* were reported recently (79).

Cross-referencing with other datasets, there was also support for new candidate dominant/*de novo* autosomal genes, notably the nucleolin gene *NCL* (8 LoF, 2 missense), but also including *ECM1*, *FAM53C*, *ADGRF2*, *TANGO2*, *DGKZ*, *ATP2B1* (5 individuals), *RETN*, and *PPIL2* (see Supplementary File 1). There was support for X-linked recessive (in males, maternally inherited or *de novo*) candidate gene *NRK* (three individuals), *MAGEB2* (three individuals), *MAGEC1* (12 individuals), *MSN* (five individuals), and *ZNF185* (one individual). Comparison with other studies of ASD simplex families from populations with higher frequency of consanguineous marriages, there were no candidate AR genes in common, however an LoF mutation in candidate X-linked gene *MAGEC1* was also reported in a Qatari cohort (54).

Biallelic variants for autism....

X-linked (inherited; Table 4): Two mutations were identified in well-established ASD or ID genes, namely, *AIFM1* (Cowchock syndrome; MIM 310490), and *MID2* (XLID101: MIM 300928).

Splice site variants

In addition to the detection of variants at canonical splice donor and splice acceptor sites (Tables 2-4), use of SpliceAI suggests that, for the homozygous nonsynonymous variant in the known ASD/ID gene *ASL* (trio SA7), an alternative explanation could be the generation of an alternative splice donor by the cytosine to thymine transition at this site. However, other splice algorithms such as https://www.fruitfly.org/seq_tools/splice.html do not support this, and molecular experimental procedures may be needed to corroborate or refute this prediction.

Copy Number Variants (Table 5 & Table S1)

Seven autosomal homozygous loss CNVs were identified among five trios using microarray data, that were subsequently validated through inspection of the WES data using IGV, and through molecular methods, using PCR and/or Sanger sequencing. Of these implicated genes, mutations/knock-outs of both *DHRS4* and *KLK15* have a neurobehavioral phenotype in mouse models (Supplementary File 1). *DNAH7* is highly expressed in brain, and SHPK (sedoheptulokinase) has high expression in cerebellum. Determining the ASD causative genetic variation in many of the probands with CNVs is

difficult. The proband with the *CYP2A7*, and *KLK15* CNVs (trio IABB14) also has a homozygous splice mutation in the Canavan disease gene *ASPA*. Additionally, the proband with the *SHPK* CNV (trio IABB2) also has a homozygous LoF mutation in the known ID gene *DEAF1*, and the proband with the *SIRPB1* CNV (trio IAU79) has a homozygous non-frameshift deletion in the known ID/spastic paraplegia gene *B4GALNT1*, complicating identification of the molecular causation (see Supplementary File 1). A homozygous CNV loss implicates the gene *SEMG1* in trio SMPA4, however expression of this gene is restricted to the seminal vesicle (see Supplementary File 1), and thus unlikely to be related to the ASD phenotype.

There are also three large, multi-genic *de novo* CNV losses among the trios, which may also be pathogenic (Supplementary Table S1).

Neuroanatomical Enrichment Analysis

Testing for regional and cell-type specific expression did not indicate clear anatomical targets with higher expression of the candidate genes. Enrichment in the cerebellum and cerebral cortex was observed in the BrainSpan human developmental atlas, but this was not supported by either the mouse or adult human atlases. We found no consistent neuroanatomical expression pattern for the identified genes, suggesting heterogeneity of neural circuits disrupted.

Biallelic variants for autism....

High consanguinity coefficient correlates with identification of homozygous variants

Using microarray data we calculated the consanguinity coefficient F_{RoH} using autosomal RoH for 104 of the 115 trios. This included 13 trios for which LoH scores were converted to F_{RoH} using the equation F_{RoH}=(LoH – 0.0111)/1.034 (see Methods and Supplementary Materials). Comparison of the degree of consanguinity (using F_{RoH}) between 33 trios with homozygous variants versus 20 trios with *de novo* and X-linked variants indicated a strongly significant shift (unpaired *t* test (two-tailed): $p < 0.0001$; df=51; mean F_{RoH} for homozygous variants= 0.09003, S.E.M.=0.00672, N=33; mean F_{RoH} for autosomal *de novo* and X-linked = 0.02955, S.E.M. = 0.00737, N=20). This is also true for variants identified in known ASD/ID genes (p =0.0045; df=22; mean F_{RoH} for homozygous 0.08146, N=13, S.E.M.=0.01019; mean F_{RoH} for autosomal *de novo* and X-linked = 0.03282, N=11, S.E.M.=0.01159).(Figure 2A). Comparison also showed clear differences in F_{RoH} levels between different cohorts, and country of ascertainment (Figure 2B & C).

Discussion

Many recent studies involving NGS in ASD have involved large cohorts, focusing predominantly on dominant/*de novo* inheritance. This focus is largely driven by the fact that autosomal recessive variants in novel genes

are more difficult to identify in the outbred population. Identification of candidate AR ASD genes through the study of consanguineous families should help streamline the identification of biallelic mutations through clinical genetic screening in outbred populations. Using whole exome trio analysis in consanguineous families, we have enriched for recessive variants to assess the role these variants play in ASD in populations where endogamy is common. This study of 115 trios has identified WES variants for 31 trios in genes previously associated with ASD or other NDD disorders (or since validated), resulting in a diagnostic yield of 27% (16 autosomal recessive, eight autosomal dominant and six X-linked variants in known genes). With a further three trios with large, multi-genic, loss CNVs validated experimentally as *de novo* and putatively pathogenic (Supplementary Table S1), this yield increases to 30%. There were many other variants identified that met filtering criteria which could potentially represent novel ASD genes or targets. We present 32 new candidate autosomal recessive variants/genes, including a homozygous variant in *SHH*, a gene associated with autosomal dominant holoprosencephaly. Within this cohort we identified variants in five genes that are associated with known metabolic syndromes (4.3% of the diagnosis cohort): *AGA*, *ASL*, *HTRA2*, *ASPA,* and *MTHFR*. These genes may represent better clinical management opportunities for patients, and potentially better therapies.

A number of the known ID genes identified among the 115 ASD trios have also previously been reported for ASD or ASD-like features, e.g. *CC2D1A*

(69,79,80), *VPS13B* (32), *DEAF1* (68), *ZNF335* (81), *ZNF292* (82,83), *MYT1L* (83,84), and *SCN2A* (73,85,86). Of the candidate genes identified here, MSSNG, SFARI, and other datasets provide putative support for the biallelic genes, with homozygous variants in *DNAH7*, *ANO10*, *DNAH8*, *CAPSL*, *CYP2A7*, and *SIRPB1*, and putative compound heterozygous variants in *DNAH7*, *DNAH8, SCN10A*, and *WDR90*, and for *de novo*/dominant genes *NCL*, *ADGRF2*, *DGKZ*, *ATP2B1*, *RETN*, and *PPIL2* (Supplementary File 1).

For three of the known ASD or ID genes reported here, namely *ZNF292*, *SCN2A*, and *MECP2,* mutations have been identified in two unrelated trios from our cohort. Mutations in the known ASD gene *SCN2A* (85) were identified in two trios- both *de novo*. In addition, mutations were identified in two other voltage-gated sodium channel members, including a homozygous stop gain mutation in *SCN10A*, in which heterozygous gain of function missense mutations have previously been linked to familial episodic pain syndrome 2 (FEPS2; MIM 615551)(87), as well as a *de novo* missense mutation in *SCN8A*. The *SCN8A* variant is predicted as damaging by all algorithms tested, and mutations in this gene have previously been linked to AD developmental and epileptic encephalopathy 13 (DEE13; MIM 614558). The Arg267* mutation identified in *MECP2* in trio SMPA78 has been frequently reported in cases of Rett syndrome. However more C-terminal truncating mutations such as the Thr412Asnfs*5 mutation in trio IAU54 typically have a significantly milder phenotype(88).

Biallelic variants for autism….

For some of the autosomal recessive candidate genes, there is additional support from animal models. For instance, *Vps1*6 (MGI:2136772) knockout mice have behavioral/neurological and nervous system phenotypes (http://www.informatics.jax.org/; Supplementary File 1). Additional discussion of mouse models relevant to the candidate genes is provided in Supplementary Materials.

There are several trios for which there are two or more candidate variants (Supplementary File 1). For some of these, there may be a variant that clearly delivers a more plausible etiopathological explanation. For instance, for trio SMPA3 the homozygous change in *MTHFR*; for trio IAH2, the homozygous LoF change in *TECPR2* (SPG49; MIM 615031; ID (29); hereditary sensory neuropathy with ID (89)); for trio IABB2, the homozygous LoF variant in *DEAF1* (NEDHELS; MIM 617171); for trio IABB14, the homozygous splice mutation in ASPA (Canavan disease; MIM 271900).

Since our initial methodology used relatively stringent MAF cut-off criteria for biallelic variants (<0.001), we attempted to justify this level by reiterating the analysis but with lower thresholds (<0.01). Using the more relaxed criteria resulted in just one additional candidate variant, in the gene *CLCA4* (gnomAD South Asian MAF=1.72E-3), but none in known NDD genes. Our methodology was particularly stringent regarding *de novo* nonsynonymous variants, owing to the large number of likely spurious calls, and the inclusion of only variants predicted as damaging by all algorithms

used may have excluded some *bona fide* mutations. While low read-depth is less problematic for identification of homozygous variants, it is likely to impact the identification of de novo variants in our trios. In addition, there may be genuine autosomal dominant disease-associated variants that are not *de novo*, inherited from an unaffected parent and thus excluded, but are non-penetrant in that parent- we included one such variant in *CHD8*, as it is a known ASD gene (MIM 615032), but there may be others that are as yet unknown. Other variants, particularly intronic or intergenic, would also likely be missed by our approach. Also, a proportion of exons, and particularly in GC-rich regions, are either missed or have low coverage in WES. Follow-up with whole genome sequencing could be considered as a possible next step. However, parsing and assessing rare non-coding variants is particularly challenging. Further, a more gene-level statistical modelling approach would also potentially benefit the study.

It was recently estimated in the DDD cohort that recessive variants make up 3.6% of diagnoses, whereas *de novo* variants contribute 48.6% of variants (38). The same study also compared homozygous variants in a narrowly defined group of Pakistani Ancestry in the British Isles (PABI), which is most like our cohort (38). The PABI cohort had 356 (333 undiagnosed) probands, 110 of whom (30.9%) had homozygous coding candidate variants, approximately half of which are in known DD genes and half in novel candidates (38). For comparison, our study reports very similar yields for

recessive variants, with 14% of probands with homozygous variants in known ASD/ID genes (16/115), or 30% including those with homozygous variants in novel candidate genes (35/115). We report 24 *de novo* variants (17 autosomal, four X-linked, three CNVs) in 17 of 115 individuals (15%), which is somewhat lower than those reported for the PABI cohort (29.8%). This difference is most likely due to differences in the methodologies or stringency used for calling *de novo* variants in the respective studies. In a study of 100 small ASD families from Qatar, while the total yield was 27%, only 7% were autosomal recessive variants (53).

Comparison of consanguinity (F_{RoH}) between trios with homozygous variants (either known or candidates) *versus* trios with *de novo* and X-linked variants indicated a strongly significant shift. Thus, unsurprisingly, the chances of finding homozygous rare variants are strongly associated with higher consanguinity, and should be a consideration for diagnostic genetic analysis.

Overall, the findings demonstrate the importance of autosomal recessive mutations in ASD in countries, or in cohorts, with high rates of consanguinity. In countries where consanguinity is rare, when screening ASD individuals' genomes attention should still be paid to the relatedness of the parents, and, for those with a high F-coefficient, attention should be focused on homozygous variants. Additionally, the candidate genes reported here should be examined for both autozygous *and* allozygous mutations in

outbred populations, just as would be done for more established autosomal recessive gene disorders.

Declaration of Interests Statement

L.F. owns shares in Quince Therapeutics and has received consulting fees from PeopleBio Co., GC Therapeutics Inc., Cortexyme Inc., and Keystone Bio. M.A. is co-founder and director of Institute of Omics and Health Research, Lahore, a research organization. The organization was not involved in the research on which this manuscript is based. All other authors declare no competing interests.

Acknowledgements

The authors thank the families for their participation in this study. We acknowledge the assistance of Natalie Freeman with PLINK microarray analysis. Computations were performed on the CAMH Specialized Computing Cluster. The SCC is funded by The Canada Foundation for Innovation, Research Hospital Fund. Additional acknowledgements are included in Supplementary Materials.

Funding

This work was supported by grants from the Canadian Institutes of Health Research to JBV (#MOP-102758 and #PJT-156402), also a Brain & Behavior

Research Foundation (BBRF/NARSAD) Young Investigator Award to NV. RH was supported by a Peterborough K.M. Hunter Charitable Foundation Graduate Scholarship. AR and AH were supported by International Research Fellowship Program scholarships from the Pakistani Higher Education Commission. CS was supported by a Summer University Research Program award from the Institute of Medical Science, University of Toronto.

Data and code availability

New variants reported here are available through ClinVar SUB14665852, and SCV accession numbers are included in Supplementary File 1. Microarray genotype data and whole exome sequence data will be made available from the corresponding author upon request.

References

1. Helt M, Kelley E, Kinsbourne M, Pandey J, Boorstein H, Herbert M, Fein D (2008): Can children with autism recover? If so, how? *Neuropsychol Rev* 18: 339–366.
2. Mason D, Capp SJ, Stewart GR, Kempton MJ, Glaser K, Howlin P, Happé F (2021): A Meta-analysis of Outcome Studies of Autistic Adults: Quantifying Effect Size, Quality, and Meta-regression. *J Autism Dev Disord* 51: 3165–3179.

3. Lai M-C, Kassee C, Besney R, Bonato S, Hull L, Mandy W, *et al.* (2019): Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry* 6: 819–829.

4. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG (2015): The epidemiology and global burden of autism spectrum disorders. *Psychol Med* 45: 601–613.

5. Elsabbagh M, Divan G, Koh Y-J, Kim YS, Kauchali S, Marcín C, *et al.* (2012): Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Res* 5: 160–179.

6. Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, Charman T (2006): Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet Lond Engl* 368: 210–215.

7. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC) (2014): Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morb Mortal Wkly Rep Surveill Summ Wash DC 2002* 63: 1–21.

8. Kim YS, Leventhal BL, Koh Y-J, Fombonne E, Laska E, Lim E-C, *et al.* (2011): Prevalence of autism spectrum disorders in a total population sample. *Am J Psychiatry* 168: 904–912.

9. Lundström S, Reichenberg A, Anckarsäter H, Lichtenstein P, Gillberg C (2015): Autism phenotype versus registered diagnosis in Swedish children: prevalence trends over 10 years in general population samples. *BMJ* 350: h1961.

10. Maenner MJ (2020): Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill Summ* 69. https://doi.org/10.15585/mmwr.ss6904a1

11. Fombonne E (2018): Editorial: The rising prevalence of autism. *J Child Psychol Psychiatry* 59: 717–720.

12. Hansen SN, Schendel DE, Parner ET (2015): Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. *JAMA Pediatr* 169: 56–62.

13. Lord C, Brugha T, Charman T, Cusack J, Dumas G, Frazier T, *et al.* (2020): Autism spectrum disorder. *Nat Rev Dis Primer* 6: 5.

14. Rødgaard E-M, Jensen K, Vergnes J-N, Soulières I, Mottron L (2019): Temporal Changes in Effect Sizes of Studies Comparing Individuals With and Without Autism: A Meta-analysis. *JAMA Psychiatry* 76: 1124–1132.

15. Hossain MD, Ahmed HU, Jalal Uddin MM, Chowdhury WA, Iqbal MS, Kabir RI, *et al.* (2017): Autism Spectrum disorders (ASD) in South Asia: a systematic review. *BMC Psychiatry* 17: 281.

16. Imran N, Bhatti MR, Anwar A, Najmi F, Haider II (2012): Children's Mental Health: Pattern of referral, distribution of disorders and service use in child psychiatry outpatient setting. *Pak J Med Sci* 28: 22–26.

17. Syed EU, Hussein SA, Yousafzai AW (2007): Developing Services with Limited Resources: Establishing a CAMHS in Pakistan. *Child Adolesc Ment Health* 12: 121–124.

18. Tareen A, Mirza I, Minhas A, Minhas F, Rahman A (2009): Developing a child and adolescent mental health service in a low-income country: a global partnership model. *Psychiatr Bull* 33: 181–183.

26. Sarwat A, Ali SMI, Ejaz MS. Mental health comorbidity in children: a hospital based study in child psychiatry clinic. Pak J Med Sci. 2009;25(6):982–5.

20. Minhas A, Vajaratkar V, Divan G, Hamdani SU, Leadbitter K, Taylor C, *et al.* (2015): Parents' perspectives on care of children with autistic spectrum disorder in South Asia – Views from Pakistan and India. *Int Rev Psychiatry* 27: 247–256.

21. Rauf NK, Anis-ul-Haq null, Aslam N, Anjum U (2014): Characteristic symptoms and adaptive behaviors of children with autism. *J Coll Physicians Surg--Pak JCPSP* 24: 658–662.

22. Bundey S, Alam H (1993): A five-year prospective study of the health of children in different ethnic groups, with particular reference to the effect of inbreeding. *Eur J Hum Genet EJHG* 1: 206–219.

23. Modell B, Darr A (2002): Science and society: genetic counselling and customary consanguineous marriage. *Nat Rev Genet* 3: 225–229.

24. Harripaul R, Vasli N, Mikhailov A, Rafiq MA, Mittal K, Windpassinger C, *et al.* (2018): Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. *Mol Psychiatry* 23: 973–984.

25. Riazuddin S, Hussain M, Razzaq A, Iqbal Z, Shahzad M, Polla DL, *et al.* (2017): Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. *Mol Psychiatry* 22: 1604–1614.

26. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, *et al.* (2018): Genetics of intellectual disability in consanguineous families. *Mol Psychiatry*. https://doi.org/10.1038/s41380-017-0012-2

27. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, *et al.* (2011): Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 478: 57–63.

28. Reuter MS, Tawamie H, Buchert R, Hosny Gebril O, Froukh T, Thiel C, *et al.* (2017): Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders. *JAMA Psychiatry* 74: 293–299.

29. Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, *et al.* (2017): Expanding the genetic heterogeneity of intellectual disability. *Hum Genet* 136: 1419–1429.

30. Morrow EM, Yoo S-Y, Flavell SW, Kim T-K, Lin Y, Hill RS, *et al.* (2008): Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321: 218–223.

31. Al-Mubarak B, Abouelhoda M, Omar A, AlDhalaan H, Aldosari M, Nester M, *et al.* (2017): Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from Saudi families. *Sci Rep* 7: 5679.

32. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, *et al.* (2013): Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77: 259–273.

33. Mamidala MP, Kalikiri MK, Praveen Kumar PTV, Rajesh N, Vallamkonda OR, Rajesh V (2015): Consanguinity in India and its association with autism spectrum disorder. *Autism Res Off J Int Soc Autism Res* 8: 224–228.

34. 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* 77: 235–242.

35. Doan RN, Lim ET, De Rubeis S, Betancur C, Cutler DJ, Chiocchetti AG, *et al.* (2019): Recessive gene disruptions in autism spectrum disorder. *Nat Genet* 51: 1092–1098.

36. Howrigan DP, Simonson MA, Davies G, Harris SE, Tenesa A, Starr JM, *et al.* (2016): Genome-wide autozygosity is associated with lower general cognitive ability. *Mol Psychiatry* 21: 837–843.

37. Kaiser VB, Svinti V, Prendergast JG, Chau Y-Y, Campbell A, Patarcic I, *et al.* (2015): Homozygous loss-of-function variants in European cosmopolitan and isolate populations. *Hum Mol Genet* 24: 5464–5474.

38. Martin HC, Jones WD, McIntyre R, Sanchez-Andrade G, Sanderson M, Stephenson JD, *et al.* (2018): Quantifying the contribution of recessive coding variation to developmental disorders. *Science* 362: 1161–1164.

39. Kahrizi K, Hu H, Hosseini M, Kalscheuer VM, Fattahi Z, Beheshtian M, *et al.* (2019): Effect of inbreeding on intellectual disability revisited by trio sequencing. *Clin Genet* 95: 151–159.

40. Le Couteur A, Bailey A, Goode S, Pickles A, Robertson S, Gottesman I, Rutter M (1996): A broader phenotype of autism: the clinical spectrum in twins. *J Child Psychol Psychiatry* 37: 785–801.

41. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, *et al.* (2011): Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* 68: 1095–1102.

42. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A (2014): The familial risk of autism. *JAMA* 311: 1770–1777.

43. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, *et al.* (2015): Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry* 72: 415–423.

44. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger JI, Hallmayer JF (2017): Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet* 18: 362–376.

45. Ramaswami G, Geschwind DH (2018): Genetics of autism spectrum disorder. *Handb Clin Neurol* 147: 321–329.

46. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, *et al.* (2008): Runs of homozygosity in European populations. *Am J Hum Genet* 83: 359–372.

47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011): Integrative genomics viewer. *Nat Biotechnol* 29: 24–26.

48. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, *et al.* (2016): Genome-wide characteristics ofde novomutations in autism. *NPJ Genomic Med* 1: 160271–1602710.

49. Koko M, Satterstrom FK, Autism Sequencing Consortium, APEX consortium, Warrier V, Martin H (2025): Contribution of autosomal

rare and de novo variants to sex differences in autism. *Am J Hum Genet* 112: 599–614.

50. Wang T, Kim CN, Bakken TE, Gillentine MA, Henning B, Mao Y, *et al.* (2022): Integrated gene analyses of de novo variants from 46,612 trios with autism and developmental disorders. *Proc Natl Acad Sci U S A* 119: e2203491119.
51. 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* 515: 209–215.
52. Yang C, Li J, Wu Q, Yang X, Huang AY, Zhang J, *et al.* (2018): AutismKB 2.0: a knowledgebase for the genetic evidence of autism spectrum disorder. *Database J Biol Databases Curation* 2018. https://doi.org/10.1093/database/bay106
53. Abdi M, Aliyev E, Trost B, Kohailan M, Aamer W, Syed N, *et al.* (2023): Genomic architecture of autism spectrum disorder in Qatar: The BARAKA-Qatar Study. *Genome Med* 15: 81.
54. Ben-Mahmoud A, Gupta V, Abdelaleem A, Thompson R, Aden A, Mbarek H, *et al.* (2024): Genome Sequencing Identifies 13 Novel Candidate Risk Genes for Autism Spectrum Disorder in a Qatari Cohort. *Int J Mol Sci* 25: 11551.
55. Gupta V, Ben-Mahmoud A, Ku B, Velayutham D, Jan Z, Yousef Aden A, *et al.* (2023): Identification of two novel autism genes, TRPC4 and

SCFD2, in Qatar simplex families through exome sequencing. *Front Psychiatry* 14: 1251884.

56. Gupta V, Ben-Mahmoud A, Idris AB, Hottenga J-J, Habbab W, Alsayegh A, *et al.* (2024): Genetic Variant Analyses Identify Novel Candidate Autism Risk Genes from a Highly Consanguineous Cohort of 104 Families from Oman. *Int J Mol Sci* 25: 13700.
57. Ghasemi M-R, Sadeghi H, Hashemi-Gorji F, Mirfakhraie R, Gupta V, Ben-Mahmoud A, *et al.* (2024): Exome sequencing reveals neurodevelopmental genes in simplex consanguineous Iranian families with syndromic autism. *BMC Med Genomics* 17: 196.
58. Khan H, Harripaul R, Mikhailov A, Herzi S, Bowers S, Ayub M, *et al.* (2024): Biallelic variants identified in 36 Pakistani families and trios with autism spectrum disorder. *Sci Rep* 14: 9230.
59. Al-Sarraj Y, Taha RZ, Al-Dous E, Ahram D, Abbasi S, Abuazab E, *et al.* (2024): The genetic landscape of autism spectrum disorder in the Middle Eastern population. *Front Genet* 15: 1363849.
60. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, *et al.* (2012): An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489: 391–399.
61. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, *et al.* (2018): Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362. https://doi.org/10.1126/science.aat7615

62. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, *et al.* (2018): Molecular Architecture of the Mouse Nervous System. *Cell* 174: 999-1014.e22.

63. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* 57: 289–300.

64. Bomkamp C, Tripathy SJ, Bengtsson Gonzales C, Hjerling-Leffler J, Craig AM, Pavlidis P (2019): Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes. *PLoS Comput Biol* 15: e1007113.

65. Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, Walsh CA, *et al.* (2006): The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation. *J Med Genet* 43: 203–210.

66. Perez Y, Menascu S, Cohen I, Kadir R, Basha O, Shorer Z, *et al.* (2018): RSRC1 mutation affects intellect and behaviour through aberrant splicing and transcription, downregulating IGFBP3. *Brain J Neurol* 141: 961–970.

67. Maddirevula S, AlZahrani F, Anazi S, Almureikhi M, Ben-Omran T, Abdel-Salam GMH, *et al.* (2018): GWAS signals revisited using human knockouts. *Genet Med Off J Am Coll Med Genet* 20: 64–68.

68. Nabais Sá MJ, Jensik PJ, McGee SR, Parker MJ, Lahiri N, McNeil EP, *et al.* (2019): De novo and biallelic DEAF1 variants cause a phenotypic spectrum. *Genet Med Off J Am Coll Med Genet* 21: 2059–2069.

69. Manzini MC, Xiong L, Shaheen R, Tambunan DE, Di Costanzo S, Mitisalis V, *et al.* (2014): CC2D1A regulates human intellectual and social function as well as NF-κB signaling homeostasis. *Cell Rep* 8: 647–655.

70. Douzgou S, Petersen MB (2011): Clinical variability of genetic isolates of Cohen syndrome. *Clin Genet* 79: 501–506.

71. Rajab A, Schuelke M, Gill E, Zwirner A, Seifert F, Morales Gonzalez S, Knierim E (2015): Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. *J Med Genet* 52: 607–611.

72. Smith DR, Stanley CM, Foss T, Boles RG, McKernan K (2017): Rare genetic variants in the endocannabinoid system genes CNR1 and DAGLA are associated with neurological phenotypes in humans. *PloS One* 12: e0187926.

73. 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* 485: 237–241.

74. Birnbaum R, Ezer S, Lotan NS, Eilat A, Sternlicht K, Benyamini L, *et al.* (2024): Intellectual disability syndrome associated with a homozygous

founder variant in SGSM3 in Ashkenazi Jews. *J Med Genet* 61: 289–293.

75. Turkyilmaz A, Saglam KA, Yilmaz M, Cebi AH (2025): A Strong Candidate Gene for Nonsyndromic Intellectual Disability Phenotype: SGSM3. *Clin Genet* 107: 196–200.

76. Sofou K, Meier K, Sanderson LE, Kaminski D, Montoliu-Gaya L, Samuelsson E, *et al.* (2021): Bi-allelic VPS16 variants limit HOPS/CORVET levels and cause a mucopolysaccharidosis-like disease. *EMBO Mol Med* 13: e13376.

77. Yıldız Y, Koşukcu C, Aygün D, Akçaboy M, Öztek Çelebi FZ, Taşcı Yıldız Y, *et al.* (2021): Homozygous missense VPS16 variant is associated with a novel disease, resembling mucopolysaccharidosis-plus syndrome in two siblings. *Clin Genet* 100: 308–317.

78. Berciano J, García A, Gallardo E, Peeters K, Pelayo-Negro AL, Álvarez-Paradelo S, *et al.* (2017): Intermediate Charcot-Marie-Tooth disease: an electrophysiological reappraisal and systematic review. *J Neurol* 264: 1655–1677.

79. Mirzaa GM, Chong JX, Piton A, Popp B, Foss K, Guo H, *et al.* (2020): De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. *Genet Med* 22: 538–546.

80. Sener EF, Onal MG, Dal F, Nalbantoglu U, Ozkul Y, Canatan H, Oztop DB (2020): Novel alterations of CC2D1A as a candidate gene in a

Turkish sample of patients with autism spectrum disorder. *Int J Neurosci* 1–8.

81. Wu J, Yu P, Jin X, Xu X, Li J, Li Z, *et al.* (2018): Genomic landscapes of Chinese sporadic autism spectrum disorders revealed by whole-genome sequencing. *J Genet Genomics Yi Chuan Xue Bao* 45: 527–538.

82. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, *et al.* (2018): Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Mol Autism* 9: 64.

83. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, *et al.* (2016): De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nat Commun* 7: 13316.

84. Coursimault J, Guerrot A-M, Morrow MM, Schramm C, Zamora FM, Shanmugham A, *et al.* (2021): MYT1L-associated neurodevelopmental disorder: description of 40 new cases and literature review of clinical and molecular aspects. *Hum Genet*. https://doi.org/10.1007/s00439-021-02383-z

85. Weiss LA, Escayg A, Kearney JA, Trudeau M, MacDonald BT, Mori M, *et al.* (2003): Sodium channels SCN1A, SCN2A and SCN3A in familial autism. *Mol Psychiatry* 8: 186–194.

86. Jiang Y, Yuen RKC, Jin X, Wang M, Chen N, Wu X, *et al.* (2013): Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet* 93: 249–263.

87. Faber CG, Lauria G, Merkies ISJ, Cheng X, Han C, Ahn H-S, *et al.* (2012): Gain-of-function Nav1.8 mutations in painful neuropathy. *Proc Natl Acad Sci U S A* 109: 19444–19449.

88. Bebbington A, Percy A, Christodoulou J, Ravine D, Ho G, Jacoby P, *et al.* (2010): Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *J Med Genet* 47: 242–248.

89. Neuser S, Brechmann B, Heimer G, Brösse I, Schubert S, O'Grady L, *et al.* (2021): Clinical, neuroimaging, and molecular spectrum of TECPR2-associated hereditary sensory and autonomic neuropathy with intellectual disability. *Hum Mutat* 42: 762–776.

90. Schopler E, Reichler RJ, DeVellis RF, Daly K (1980): Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J Autism Dev Disord* 10: 91–103.

91. Lord C, Rutter M, Le Couteur A (1994): Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24: 659–685.

92. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989): Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* 19: 185–212.

93. American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision*, 4th edition. Washington, DC: American Psychiatric Association.

94. Ehlers S, Gillberg C, Wing L (1999): A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *J Autism Dev Disord* 29: 129–141.

95. Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, Place M (2004): The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 43: 548–558.

96. American Psychiatric Association, American Psychiatric Association, DSM-5 Task Force (2013): *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* Retrieved December 14, 2016, from http://dsm.psychiatryonline.org/book.aspx?bookid=556

97. World Health Organization (2010): The ICD-10 Classification of Mental and Behavioural Disorders. World Health Organization. Retrieved from https://www.who.int/classifications/icd/en/bluebook.pdf

98. Packer JS, Maxwell EK, O'Dushlaine C, Lopez AE, Dewey FE, Chernomorsky R, *et al.* (2016): CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. *Bioinformatics* 32: 133–135.

Table 1: Cohort Description and clinical assessment.

Cohort ID (Recruited By)	Institute	Nationality (Age range)	Number of Trios (Number of DNAs)	Proband Male: Female	Instruments and diagnostic system	Assessors
SMPA (Ansa Rabia, Saqib Mahmood, Shazia Maqbool)	University of Health Sciences and Children Hospital, Lahore	Pakistan	54 (162)	42:12	Childhood Autism Rating Scale (CARS)(90) DSM IV	Physicians with expertise in neurodevelopmental disorders
IAU (Dr. Sasanfar)	Children's Health and Evaluation Project/ Special Education Organization	Iran (4-11)	27 (81)	23:4	ADI-R (91) and ADOS (92) (DSM-IV-TR) (93); SCQ (50)	Certified Physicians
IABB (Dr. Bita Bozorgmehr)	Shahid Beheshti Medical University, Teheran	Iran	10 (30)	9:2	Autism Spectrum Screening Questionnaire (94) DSM IV	Bita Bozorgmehr
SA (Laila Al Ayadhi)	Autism Treatment Centre, King Saud University, Riyadh	Saudi Arabia (2-12)	13 (39)	12:1	ADI-R (91), Autism Diagnostic Observation Schedule (ADOS) (92) and 3DI (Developmental, dimensional diagnostic interview) scales (95) Childhood Autism Rating Scale (CARS)(90); DSM IV	Laila Al Ayadhi

Autism (Raheel Qamar, Maleeha Azam and Zehra Agha)	COMSATS University, Islamabad, and National Rural Support Program and the Ghazi Barotha Taraqiata iDara	Pakistan	4 (12)	4:4	Childhood Autism Rating Scale (CARS)(90) DSM-5 (96) and ICD-10 criteria (97)	Maimoona Siddiqui at Shifa International Hospitals Limited, Islamabad, *et al*.
RQPA (Raheel Qamar, Maleeha Azam and Zehra Agha)			4 (12)			Maimoona Siddiqui at Shifa International Hospitals Limited, Islamabad, *et al*.
IAH (Abolfazl Heidari)	Qazvin Medical University, Qazvin	Iran	3 (9)	2:1	DSM IV	Hossein Mojdehipanah
Total			115 (345)	91:24		

Table 2: Biallelic/homozygous variants: Genes previously reported as pathogenic for ASD/ID or other disorder are indicated, with annotation using (in order of priority): OMIM: https://omim.org/; SFARI ASD genes https://gene.sfari.org/; DDD gene2phenotype genes: https://www.ebi.ac.uk/gene2phenotype/; Geisinger developmental brain disorder gene database: https://dbd.geisingeradmi.org/; NP*denovo*: http://www.wzgenomics.cn/NPdenovo/; Gene4denovo: http://www.genemed.tech/gene4denovo/search. Genes previously unreported as definitively disease-causing/pathogenic for ASD or ID are in bold type. Control population frequencies shown are from gnomAD (all, and for South Asian sub-population). Other control population datasets, such as the Regeneron Pakistani and Middle Eastern subsets, and Iranome, were also checked (see Supplementary File 1). No homozygotes were present in the controls. All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in Supplementary Materials. Allele frequencies are from gnomAD v4.1.0 (https://gnomad.broadinstitute.org/; accessed June 2025). N.B. *SETD2* and *SHH* are indicated here as novel AR, but are known for autosomal dominant NDD phenotypes.

Variant	Family	Chr.	Coordinates (hg38)	Ref.	Alt.	Gene	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomAD S. Asian Freq. (or ME=Middle Eastern)
1	SMPA3	1	11795347	C	A	MTHFR (28)(MIM 236250; SFARI score 2)	nonsynonymous	NM_001330358.2: exon6: c. 905G>T: p. Gly302Val	6.20E-06	1.10E-04
2	IABB16	1	86563694	G	A	**CLCA4**	nonsynonymous	NM_012128.3:c.482G>A:p.Arg161Gln	1.93E-05	2.24E-05
3	SMPA4	1	152221199_152221208	GACGTTTC TG	-	**HRNR**	frameshift del	NM_001009931.3: exon3: c.423_432del:p.Arg142Glufs*15	2.48E-05	4.17E-04
4	SMPA76	1	166922618_166922619	CT	-	**ILDR2**	frameshift del	NM_199351.3: exon8: c.1187_1188del: p. Glu396Glyfs*39	6.20E-07	1.10E-05

Biallelic variants for autism....

5	SMPA40	1	197652234	C	A	**DENND1B**	Splicing	NM_001195215.1: exon7: c.447+1G>T	0	0
6	Autism8	2	74532659	G	T	HTRA2 (MIM 617248)	Nonsynonymous	NM_013247.3: exon7: c.1156G>T: p.Asp386Tyr	3.72E-06	3.30E-05
7	SMPA19	3	38760694	C	A	**SCN10A**	Stopgain	NM_001293306.2: exon7: c.937G>T; p.Gly313*	6.82E-06	1.21E-04
8	SMPA76	3	43551578	C	A	**ANO10** (spinocerebellar ataxia 10, MIM 613728)	Nonsynonymous	NM_001346463.2:exon11:c.1679G>T:p.Cys560Phe	4.38E-06	3.10E-05
9	SA15	3	47122939	G	C	**SETD2** (LLS (AD) MIM 616831; SFARI score 1; Geisinger Tier 1)	Nonsynonymous	NM_001349370.3: exon2:c.1565C>G: p.Ser522Cys	3.72E-06	0 (1.66E-06 ME)
10	SMPA27	3	135192815	T	C	**EPHB1** (SFARI score 3**)**	Nonsynonymous	NM_004441.3: exon11: c.2122T>C: p.Phe708Leu	6.20E-07	1.10E-05
11	Autism3	3	158122213	C	T	RSRC1 (MRT70; MIM 618402; SFARI score S)	Stopgain	NM_001271834.2: exon2:c.109C>T: p.Arg37*	1.25E-06	0
12	SMPA8	3	169806847	C	T	**LRRC34**	Splicing	NM_001172779.1: exon5: c.528+1G>A	1.90E-06	3.41E-05
13	SMPA40	4	110559675	C	T	**ENPEP**	Stop gain	NM_001977.4: exon19: c.2671C>T; p.Arg891*	8.48E-05	1.19E-03
14	RQPA12	4	177439668	G	A	AGA (MIM 208400)	Nonsynonymous	NM_000027.3: exon3: c.302C>T: p.Ala101Val	4.35E-06	4.39E-05
15	SMPA35	4	184112778_184112779	CAT	C	**ENPP6**	frameshift del	NM_153343.3: exon6: c.886_887del: p.Met296Aspfs*28	0	0
16	IABB11	5	35910482	C	T	**CAPSL**	Nonsynonymous	NM_144647.3:c.200G>A:p.Gly67Arg	3.60E-05	0
17	Autism8	5	151467266_151467269	-	CTGG	**SLC36A1**	frameshift ins	NM_001308150.2: exon6:c.487_490dup: p.Asp164Glyfs*2	5.59E-06	9.92E-05

Biallelic variants for autism....

18	IAU56	5	154820404	C	-	**FAXDC2**	frameshift del	NM_032385.3: exon9: c.915delG: p.Thr306Profs*100	3.35E-05	4.50E-04
19	SMPA48	6	13634491	C	G	**RANBP9**	Nonsynonymous	NM_005493.3: exon11: c.1735G>C: p.Gly579Arg	0	0
20	IABB8	6	38883898	A	T	**DNAH8**	Nonsynonymous	NM_001206927.2: exon56: c.8159A>T: p.Asp2720Val	3.15E-06	5.77E-05
21	IAU7	7	47905327	T	C	**PKD1L1** (AR laterality defects; MIM 616205)	Splicing	NM_138295.3: exon11: c.1523-2A>G	0	0
22	SA7	7	66086640	C	T	ASL (MIM 207900)#	Nonsynonymous	NM_001024943.2: exon6:c.502C>T: p.Arg168Cys	6.20E-06	0
23	IABB7	7	100122386	G	C	**CNPY4**	Splicing	NM_152755.1: exon2: c.245+1G>C	4.15E-05	0
24	IAU79	7	150477735_150477736	TCC	T	**GIMAP8**	frameshift del	NM_175571.3: exon5: c.1952_1953del: p.Gln652Serfs*6	0	0
25	SMPA11	7	155811917	T	A	**SHH** (HPE3 (AD) MIM 142945)	Nonsynonymous	NM_000193.4: exon1: c.206A>T: p.Asn69Ile	1.30E-05	2.20E-4
26	IABB3	8	99143151	AG	A	VPS13B (COH1 MIM 216550; SFARI score 1; Geisinger AR)	frameshift del	NM_015243.3: exon13: c.1829delG: p.Ser610Thrfs*3	0	0
27	SA4	9	5805613	T	C	**ERMP1**	Nonsynonymous	NM_024896.3: exon9: c.1721A>G: p.His574Arg	1.102E-05	0 (1.694E-04 ME)
28	IABB2	11	679703	G	GA	DEAF1 (MIM 617171; SFARI score 1; Geisinger AR)	frameshift ins	NM_001293634.1: exon6: c.845dupT: p.Ala283Argfs*20	0	0
29	SMPA11	11	6456659	C	A	**TRIM3**	Nonsynonymous	NM_033278.3:exon6:c.1067G>T:p.Arg356Leu	0	0
30	SMPA20	11	47284198	A	G	MADD (26,29)	Nonsynonymous	NM_001135943.2: exon11: c. 1883A>G: p.Tyr628Cys	6.82E-06	1.21E-04

Biallelic variants for autism....

31	SMPA21	11	61728163	C	G	**DAGLA** (SFARI score 2)	Stopgain	NM_006133.3: exon7:c.647C>G: p.Ser216*	0	0
32	SMPA80	11	67522557	C	T	**CABP2** (Deafness, MIM 607314)	Nonsynonymous	NM_016366.3: exon2: c.202G>A: p.Ala68Thr	1.87E-05	2.02E-04
33	SMPA8	11	118533148	A	G	**TMEM25**	Nonsynonymous	NM_001144036.2: exon3: c. 302A>G: p.Asn101Ser	1.81E-05	2.75E-04
34	IAU79	12	57627731_57627751	GGGAAGCCGACGAGCTCGTGGT	G	B4GALNT1 (Spastic paraplegia 26; MIM 609195 ID reported in some cases)	nonframeshift del	NM_001276468.2: exon9: c.1094_1114del: p.His365_Phe372del	0	0
35	IABB10	12	68869360	C	G	**CPM**	Nonsynonymous	NM_001005502.2: exon6: c.752G>C: p.Gly251Ala	1.86E-06	1.10E-05
36	IABB11	12	117761234	G	A	**KSR2**	Nonsynonymous	NM_173598.6:exon4:c.763C>T:p.Arg255Trp	2.92E-05	2.53E-05
37	SMPA8	14	64565356	G	C	**PPP1R36**	Splicing	NM_172365: exon5: c.270-1G>C	3.73E-06	6.63E-06
38	IAH2	14	102497094_102497107	TGTGGGGACCGCCTG	T	TECPR2 Anazi et al, 2017 PMID 27431290); Neuropathy, hereditary sensory and autonomic, type IX, with developmental delay (HSAN9): MIM 615031	frameshift del	NM_014844.3: exon18: c.3904_3917del: p.Val1302Glyfs*60	0	0
39	IABB4	15	100573758	T	-	LINS1 (MRT27; MIM 614340; Geisinger AR)	frameshift del	NM_001040616.3: exon5: c.1116delA: p.Glu372Aspfs*9	0	0
40	SMPA76	16	666004	C	T	**WDR90**	Nonsynonymous	NM_145294.4: exon36: c.4489C>T: p.Arg1497Trp	4.37E-06	2.20E-05
41	IABB14	17	3489343	G	T	ASPA (Canavan disease; MIM 271900)	Splicing	NM_001128085: exon5: c.634+1G>T	4.40E-06	2.21E-05
42	SA2	19	13913639	G	T	CC2D1A (MRT3; MIM 608443; SFARI score 2; Geisinger AR)	Splicing	NM_017721: exon6: c.748+1G>T	6.25E-07	0 (1.704E-04 ME)

Biallelic variants for autism....

43	IABB16	19	49818402	G	A	MED25 (BVFYS; MIM 616449)	Nonsynonymous	NM_030973.3: exon1: c.61G>A: p. Val21Met	1.24E-06	0
44	SA11	19	57754064	C	-	**ZNF776**	Frameshift del	NM_173632.4:c.934delC: p.Arg312Glufs*80	6.82E-06	0 (1.155E-03 ME)
45	IABB16	20	2866555	C	A	**VPS16** (Dystonia (AD) MIM 619291; NDD (AR))(76,77)	Nonsynonymous	NM_080413.3: exon20: c.2069C>A: p. Ala690Glu	1.67E-05	2.20E-05
46	IABB16	20	16380092	A	G	KIF16B (PMID 29736960)	Nonsynonymous	NM_001199865.2: exon19: c. T1910T>C: p. Val637Ala	0	0
47	IABB5	20	45950360	C	T	ZNF335 Hu et al, 2019; MCPH10; MIM 615095; Geisinger Tier 4)	Nonsynonymous	NM_022095.3: exon22: c.3346G>A: p. Gly1116Arg	3.77E-06	1.14E-05
48	IABB11	22	40407248	G	C	**SGSM3** (SFARI score 2; Geisinger Tier 4)(74,75)	Nonsynonymous	NM_001350039.2:c.1288G>C:p.Glu430Gln	6.26E-06	8.78E-05

SpliceAI predicts the generation of a splice donor site.

Biallelic variants for autism....

Table 3: *De novo* variants. Genes previously reported as pathogenic for ASD/ID or other disorder are indicated, with annotation using (in order of priority): OMIM: https://omim.org/; SFARI ASD genes https://gene.sfari.org/; DDD gene2phenotype genes: https://www.ebi.ac.uk/gene2phenotype/; Geisinger developmental brain disorder gene database: https://dbd.geisingeradmi.org/; NP*denovo*: http://www.wzgenomics.cn/NPdenovo/; Gene4denovo: http://www.genemed.tech/gene4denovo/search. Genes previously unreported for ASD or ID are in bold type. Control population frequencies shown are from gnomAD (all, and for South Asian sub-population). Allele frequencies are from gnomAD v4.1.0 (https://gnomad.broadinstitute.org/; accessed June 2025). All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in Supplementary Materials.

Variant	Family	Chr	Coordinates (hg38)	Ref.	Alt.	Gene (previously reported)	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomADS. Asian Freq.
1	RQPA20	1	28592734	G	A	**RAB42**	nonsynonymous	NM_001193532.3: exon1: c. 223G>A: p. Glu75Lys	3.26E-06	0
2	RQPA20	1	111488692	A	-	**TMIGD3**	frameshift del	NM_001302680.2: exon2: c.286delT: p. Trp96Glyfs*36	0	0
3	RQPA20	1	150511490	G	T	**ECM1** (DDDG2P: lipoid proteinosis, AR)	Stopgain	NM_001202858.2: exon7: c.823G>T: p. Glu275*	3.10E-06	4.39E-05
4	IAU29	2	1943094	C	-	MYT1L (MRD39: MIM 616521; SFARI score 1; MYT1L)	frameshift del	NM_001303052.2: exon9: c.394delG: p. Glu132Argfs*42	0	0
5	SMPA53	2	165310448	C	T	SCN2A (DEE11: MIM 6137219; SFARI score 1; Geisinger Tier 1)	Stopgain	NM_001040143.2: exon6:c.823C>T: p. Arg275*	1.56E-05	6.20E-0

Biallelic variants for autism....

6	SMPA38	2	165344646	C	A	SCN2A (DEE11: MIM 6137219; SFARI score 1; Geisinger Tier 1)	Nonsynonymous	NM_001040143.2:exon15:c.2654C>A:p.Thr885Asn	0	0
7	IAU66	2	231455467	C	-	**NCL**	frameshift del	NM_005381: exon13: c.1991delG: p.Gly664Glufs*70	0	0
8	SMPA75	5	138344944	-	T	**FAM53C**	frameshift ins	NM_001135647.2: exon4: c.256dupT: p.Ser86Phefs*14	0	0
9	SMPA75	6	47679135	G	A	**ADGRF2**	Splicing	NM_153839.7: exon5: c.267+1G>A	1.86E-06	3.29E-05
10	Autism10	6	87259789-87259790	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1; Geisinger Tier 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p.Glu2054Lysfs*14	1.25E-06	0
11	IAU65	6	87259789-87259790	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1; Geisinger Tier 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p.Glu2054Lysfs*14	1.25E-06	0
12	RQPA20	11	46368000	A	C	**DGKZ**	Splicing	NM_201533: exon4: c.379-2A>C	1.24E-06	2.20E-05
13	Autism9	12	51794454	G	A	SCN8A (DEE13: MIM 614558; SFARI score 1)	Nonsynonymous	NM_014191.4:c.4608G>A: p.Met1536Ile	0	0
14	SMPA73	12	89604231	G	C	ATP2B1 (MRD66: MIM 619910; SFARI score 3S)	Stopgain	NM_001001323.2: exon15:c.2558C>G: p.Ser853*	0	0
15	SMPA75	19	7669338-7669339	CT	-	**RETN**	frameshift del	NM_001193374.1: exon2: c.8_9del: p.Cys5Serfs*13	2.48E-05	4.28E-05

16	RQPA20	22	20061537	C	G	**TANGO2 (Geisinger AR)**	Stop gain	NM_152906.7:exon7:c.459C>G: p.Tyr153*	0	0
17	RQPA20	22	21696745	A	G	**PPIL2**	Splicing	NM_148175: exon21: c.1570-2A>G	6.49E-07	1.10E-05

Table 4: X-linked variants. Genes previously reported as pathogenic for ASD/ID or other disorder are indicated, with annotation using (in order of priority): OMIM: https://omim.org/; SFARI ASD genes https://gene.sfari.org/; DDD gene2phenotype genes: https://www.ebi.ac.uk/gene2phenotype/; Geisinger developmental brain disorder gene database: https://dbd.geisingeradmi.org/; NP*denovo*: http://www.wzgenomics.cn/NPdenovo/; Gene4denovo: http://www.genemed.tech/gene4denovo/search. Genes previously unreported for ASD or ID are in bold type. Control population frequencies shown are from gnomAD (all, and for South Asian sub-population). Allele frequencies are from gnomAD v4.1.0 (https://gnomad.broadinstitute.org/; accessed June 2025). All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in Supplementary Materials.

Variant	Family	Mode of Inheritance	Coordinates (hg38)	Ref	Alt	Gene (previous reports)	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomADS. Asian Freq.
1	SMPA3	Recessive	30218663	C	-	**MAGEB2**	frameshift del	NM_002364.4:c.83delC:p.Pro28Leufs*84	0	0
2	IAU66	Recessive	48814541	A	G	**HDAC6**	Nonsynonymous	NM_006044.4:c.908A>G: p.Tyr303Cys	0	0
3	IAU45	Recessive	65727856	C	A	**MSN**	Nonsynonymous	NM_002444.3: exon3: c.139C>A: p. Leu47Met	0	0
4	SMPA15	Recessive	105949596	C	T	**NRK**	Stopgain	NM_198465.4: exon27: c.4375C>T: p. Gln1459*	0	0
5	IAU47	Recessive	107926868	A	G	MID2 (MRX101; MIM 300928)	nonsynonymous	NM_012216.4:c.2003A>G:p.Tyr668Cys	1.65E-06	0
6	IAU24	Recessive	130156472	C	T	AIFM1 (Cowchock	nonsynonymous	NM_001130847.4: exon2: c.238G>A: p. Ala80Thr (ClinVar	3.30E-06	0

						syndr.; MIM 310490)		variant ID 643584; variant of uncertain significance)		
7	IAU51	Recessive	141906875-141906876	-	A	**MAGEC1**	frameshift ins	NM_005462.3:c.1471_1472insA; p.Leu491Tyrfs*10	7.87E-05	0
8	IABB8	Recessive	152938072	A	G	**ZNF185**	Splicing	NM_001178110: exon13: c.942-2A>G	4.26E-06	0
9	SMPA71	De Novo	63706329	G	A	ARHGEF9 (DEE8; MIM 300607; SFARI score 1S; Geisinger Tier1)	nonsynonymous	NM_015185.3: exon3:c.310C>T: :p.Arg104Trp (ClinVar variation ID 501568; conflicting interpretations of pathogenicity)	0	0
10	SMPA84	De Novo	101252612	A	C	DRP2 (CMTX1 MIM 302800; PMID 26227883)	nonsynonymous	NM_001171184.2: exon15: c. 1639A>C: p. Ser547Arg	0	0
11	IAU54	De Novo	154030630	-	T	MECP2 (RTT; MIM 312750; SFARI score 1; Geisinger Tier1)	frameshift ins	NM_001110792.2:exon3:c.1234dup:p.Thr412Asnfs*5	0	0
12	SMPA78	De Novo	154031065	G	A	MECP2 (RTT; MIM 312750; SFARI score 1’ Geisinger Tier1)	Stopgain	NM_001110792.2:exon3:c.799C>T:p. Arg267* (ClinVar: known pathogenic variant)	0	0

Biallelic variants for autism....

[&]MutationTaster predicts splice donor created

Biallelic variants for autism....

Table 5: CNV summary: validated homozygous loss CNVs called by CNVpartition or ChAS. Homozygous losses called from microarray data by PennCNV and from WES data by CLAMMS (98)) did not validate using IGV, and hence are not included. Comparison with control population CNVs used the Database of Genomic Variants (DGV) Gold, accessed through the Decipher genomic browser (www.deciphergenomics.org), also through gnomAD. Coordinates given using hg38.

Family	Mode of Inheritance	(hg38)	Cytoband	Size (Kb)	CN	CNV confidence	Genes	FREQ IN DGV GOLD STANDARD, gnomAD, or rare?	IGV	PCR
SMPA4	Recessive	14:23985715-23995116	14q11.2	9.402	0	345.9659	DHRS4 exon 8; DHRS4L2 exons 1-4	Partial overlap with gssvL33876 FREQ 0.0029	Y	Y
SMPA4	Recessive	20:45208144-45208323	20q13.12	0.18	0	228.5597	SEMG1 exon 2, 180bp, 60 amino acids	Partial overlap with gssvL72784 FREQ 0.0038	Y	Y
SMPA8	Recessive	2:195934670-195936638	2q32.3	1.969	0	139.6294	DNAH7 exon 20-21	NONE	Y	Y
IAU79	Recessive	20:1578613-1578684	20p13	0.072	0	171.6751	SIRPB1 exon 2	Partial OVERLAP WITH COMMON CNV (0.0894)	Y	Y
IABB2	Recessive	17:3610562-3610760	17p13.2	0.198	0	318.3258	SHPK exon 7	Partial OVERLAP WITH 80Kb gssvL47583 (0.00274)	Y	Y
IABB14	Recessive	19:40850414-40880128	19q13.2	29.714	0	579.0146	CYP2A7 6 exons	gssvL58578 0.002 Plus gssvL58575 0.0017	Y	Y
IABB14	Recessive	19:50827695-50827807	19q13.33	0.112	0	579.0146	KLK15 exon 2	Overlap with gssvL59230 0.02	Y	Y

Biallelic variants for autism....

Figure Legends

Figure 1: Variant prioritization methods where variants that pass GATK best practices would be filtered for MAF less than or equal to 10^{-4} and then categorized into recessive or homozygous variants. Variants are then prioritized based on assessment of potential damage to the protein, and then filtered based on scoring algorithms to predict pathogenicity.

Figure 2: Violin plots comparing consanguinity coefficient F_{roh}, for A: different categories of variant identified: All biallelic variants (N=27 observations), variants in known AR ID and/or ASD genes (N=10 observations), variants in novel candidate genes (N=17 observations), all X-linked (XL) variants plus *de novo* autosomal dominant (AD) variants (N=18 observations), variants in known XL and AD genes (N=11 observations), and variants in novel candidate XL and AD genes (N=7 observations). Unpaired *t* test (2-tailed) for comparison of means showed comparison for i) all homozygous versus all *de novo* autosomal plus X-linked to be extremely significant ($p<0.0001$, $t=5.8309$, d.f.=51), ii) known AR versus AD ID and/or ASD genes to be strongly significant ($p=0.0045$, $t=3.1642$, d.f.=22), iii) novel candidate AR versus AD genes to be strongly significant ($p<0.0001$, $t=5.4795$, d.f.=26). No significant difference in mean F_{roh} was seen for comparisons between known and novel candidate AR genes, or between

known and novel AD plus XL genes; **B:** different cohorts, categorized by last name of principal investigator for each collection (Agha, N=3, Ansar, N=52, Heidari, N=2, Al Ayadhi, N=13, Bozorgmehr, N=10, Sasanfar, N=24 observations). While F_{roh} distribution was similar for most cohorts, the Iranian cohort from Sasanfar had a significantly lower mean than most other cohorts; **C:** grouped by country of origin (Pakistan, N=55, Iran, N= 36, Saudi Arabia=13 observations). Comparison of mean F_{roh} was non-significant for Pakistan versus Iran and versus Saudi Arabia, but significant for Iran versus Saudi Arabia, but non-significant after correction for multiple testing ($p=0.0206$, $t=2.396$, $d.f.=36$). Plots were prepared using R software using ggplot2, and show mean, standard deviation, and outliers marked by “X”.