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

Running title: Biallelic variants for autism and consanguinity

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

[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 (affiliated with National University of Medical Sciences, Rawalpindi), Pakistan;

[5]Genome Diagnostics, Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, ON, Canada;

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

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

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

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

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

[11]Center for Psychiatric Research, Maine Medical Center Research Institute, 509 Forest Avenue, Portland, ME, 04101, USA;

[12]Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA, 02110, USA;

[13]Department of Psychiatry, Maine Medical Center, 22 Bramhall Street, Portland, ME, 04102, USA;

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

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

[16]Children's Hospital & The Institute of Child Health, Lahore, Pakistan;

[17]Department of Cell Biology, SUNY Downstate Health Sciences University, New York City, NY, USA;

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

[19]Department of Academic Psychiatry, University College London, London, UK;

[20]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, Department of Psychiatry, Queen's University, 191 Portsmouth Avenue, Kingston, ON, K7M 8A6, Canada. Email: ma84@queensu.ca

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects about 1 in 55 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 of ASD, but the majority of these studies were conducted in outbred populations, which are not ideal for detecting autosomal recessive (AR) inheritance. However, several studies of ASD in consanguineous populations point towards AR as an under-appreciated source of ASD variants. Here, we used trio 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. We report 87 candidate sequence variants, with 57% biallelic, 21% autosomal dominant/*de novo*, and the rest X-linked. 52% of the variants were loss of function (LoF) or putative LoF (splice site, stop loss), and 47% nonsynonymous. Our analysis indicates an enrichment of biallelic variants. These include variants in genes previously reported for AR ASD and/or intellectual disability (ID) (*AGA*, *ASL*, *ASPA*, *BTN3A2, CC2D1A*, *DEAF1*, *HTRA2*, *KIF16B*, *LINS1*, *MADD*, *MED25*, *MTHFR*, *RSRC1*, *TECPR2*, *VPS13B*, *ZNF335*), and 32 previously unreported candidates, including 15 LoF or splice variants, in genes such as *DAGLA*, *EFCAB8*, *ENPP6*, *FAXDC2*, *ILDR2*, *PKD1L1*, *SCN10A*, and *SLC36A1*. We also identified several candidate biallelic exonic loss CNVs. The enrichment shown here for biallelic variants confirms that the genetic architecture for ASD among consanguineous populations is different to non-

consanguineous populations. However, as has been shown in many recessive disorders, the genes reported here may also be relevant to outbred populations.

Introduction

Autism spectrum disorder (ASD) is characterized by deficits in social communication, also repetitive and 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 outcome as reported in a recent meta-analysis of longitudinal studies (2). ASD is quite heterogeneous in terms of level of intellectual functioning, co-morbid psychiatric and behavioural problems, and these factors play an important role in the long term outcome of the diagnosis (3). In a systematic review its prevalence 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 prevalence of ASD 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 10000 individuals (5), prevalence studies from the 1980s suggested that as many as 72 in 10000 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 2012 was 14.5 per 1,000 (one in 69) children aged 8 years (10) and in 2016 it increased to 18.5 per 1,000 (one in 54) children (11). This increase at least partly is because of

changes in diagnostic criteria, reporting practices, increased awareness (12–15). For example, a Swedish study showed that although ASD diagnoses became more common over time among individuals born between 1992 and 2002, the underlying level of autistic traits did not (9). Studies that examined the differences between individuals with and without ASD on symptom and cognitive measures indicated a decrease in the differences between the two groups (15,16). However, in a recent study from Denmark only a third of ASD diagnoses could be explained because of broadened diagnostic criteria (13). The search for alternative explanations for the increase in the prevalence includes the possibility of a bigger contribution of the environmental factors that might have become more common with time and have caused a genuine increase in the numbers of affected individuals (17). Many of the reported environmental exposures associated with ASD, such as air pollution exposure, paternal age, and maternal psychotropic medication use during pregnancy tend to occur during the prenatal period (18,19). This has been further examined in a longitudinal twin study where there was an increase in the ASD traits while the contribution of environmental factors had not changed (20).

Describing the genetic architecture of ASD is the first step towards understanding the pathogenesis of the disorder. There has been significant progress to date, and in genetic studies of ASD it has been 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 (21)

A systematic review of published literature from South Asian countries reported a prevalence of ASD in the range of 0.09-1.07 % (22). To date, there has been no community study

of prevalence from Pakistan. Studies from child psychiatry clinics reported rates of 2.4% and 3.2%, 4.5%, 5% (23–26). Clinical services for ASD are available but limited to big cities. There are special schools in every big city for children with ASD (27). The literature on clinical presentation is limited. One study with a small sample size shows symptoms consistent with studies elsewhere in the world (28).

Several projects have highlighted recessive inheritance as an important component of the genetic architecture of ASD. A large study of consanguineous versus non-consanguineous families in India concluded that consanguinity increases risk for ASD with an odds ratio of 3.22 (29). 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 support AR inheritance as an important piece of the genetic puzzle for ASD. For example, previous work has estimated that loss-of-function (LoF) recessive mutations contribute 3% of ASD genetics in two US-based case-control cohorts (31). These findings are not limited to population isolates or ethnic subgroups (32,33). Consanguineous marriages lead to a marked increase in the frequency of rare recessive disorders (34,35). Many genetic variants will only be pathogenic in recessive form; this includes variants for ID, and almost certainly for ASD too. 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. Populations with a high proportion of consanguinity have been important for describing autosomal recessive genes in ID in Pakistan (36,37), Iran (38,39), Syria (40), and Saudi Arabia (41), yet, to date, relatively

few reports of AR genes underlying ASD in consanguineous populations have been published (30,42,43). 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 (44). In an Iranian ID study, the *de novo* rate was 27.86% (45).

Here, we report on a study of genetic and genomic variants in a cohort of 115 ASD trios from three countries with a high frequency of consanguineous marriages (Pakistan, Iran, and Saudi Arabia). Of the candidate sequence variants we report (not including CNVs), the proportion that are *de novo* is ~24%, similar to the rates reported among Pakistani ancestry DD individuals in the UK and in ID individuals in Iran. We report 57% of the candidate sequence variants identified in the study are AR. This is somewhat higher than the 31% AR reported for Pakistani ancestry subjects in the UK DDD study (3.6% for those of European ancestry). Of the 28 sequence variants identified in genes previously linked to neurodevelopmental disorders, 25% were *de novo*, and 55% AR. This compares well with rates in a study of Iranian ID families (28% de novo, 69% AR) (45).

Materials and Methods

Trio Family Ascertainment

Institutional Research Ethics Board approval was received for this study through the Centre for Addiction and Mental Health (CAMH) and other institutional recruiting sites. Twenty-seven trios were ascertained by Dr. Sasanfar by the Children's Health and Evaluation project (CHEP) sponsored by the Special Education Organization (SEO), as described previously (46,47). Children in the cohort were between ages 4 and 11, and received a clinical diagnosis of ASD by a child psychiatrist. All parents from Iranian cohorts filled out the Social Communication Questionnaire (SCQ) (48), using the *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition-text revision (DSM-IV-TR) (49). The diagnosis of ASD was confirmed in all subjects using ADI-R (50) and ADOS (51) by a certified physician. Another 13 trios were ascertained through clinical services in Iran: Dr. Bita Bozorgmehr clinically assessed 10 trios and Dr. Abolfazl Heidari assessed three trios. Fifty-four Pakistani trios were recruited from the University of Health Sciences and Children Hospital, Lahore, Pakistan. Pakistani ASD trio probands were assessed using the Childhood Autism Rating Scale (CARS)(52). The clinical assessment has been performed by Drs. Ansa Rabia Saqib Mahmood, and Shazia Maqbool. A further eight Pakistani trios were recruited through COMSATS University, Islamabad, by Drs Raheel Qamar and Zehra Agha, with ascertainment through the National Rural Support Program and the Ghazi Barotha Taraqiata iDara. Diagnosis of ASD was performed using combined DSM-5 (53) and ICD-10 criteria (54), after assessment using CARS (52). The 13 Saudi Arabian trios were provided by Dr. Laila Al Ayadhi through the Autism Treatment Centre, King Saud University, Riyadh, Saudi Arabia, and included subjects aged 2-12 years old who had a diagnosis of ASD confirmed using the ADI-R (50), Autism Diagnostic Observation Schedule (ADOS) (51) and 3DI (Developmental, dimensional diagnostic interview) scales (55). Disease

severity was also measured using the Childhood Autism Rating Scale (CARS)(52). Saudi subjects were excluded if they had dysmorphic features, or diagnosis of fragile X or other serious neurological (e.g., seizures), psychiatric (e.g., bipolar disorder) or known medical conditions.

Overall, we have sequenced 115 trios: 62 trios from Pakistan, 40 trios from Iran, and 13 trios from Saudi Arabia. In total, we have sequenced 345 samples from the different institutions, with collaborator contributions listed in Table 1.

Whole Exome Sequencing, Alignment and Variant Calling

Whole Exome Sequencing (WES) was performed using the Thruplex DNA-Seq (Rubicon Genomics) Library Preparation Kit with the Agilent SureSelect V5 Exome Capture kit. Samples were sheared using the Covaris ME220 Focused-Sonicator to ensure that the DNA was sheared to 200 bp. Samples were then analyzed using the Agilent 2100 Bioanalyzer System for fragment length distribution and quantification. Library preparation and adapter ligation was performed using the Thruplex DNA-seq kit. All trios, representing the proband, mother and father, were sequenced on the Illumina HiSeq 2500 or the Illumina NovaSeq sequencing system.

After sequencing, FastQC was used to assess the quality of sequences and determine the adapter contamination and read quality. Unligated adapters and low quality sequences were removed using FastX Toolkit (Updated Aug 4th 2017: https://github.com/agordon/fastx_toolkit) (56) and Trimmomatic (v 0.33: Updated Mar. 10th, 2015: https://github.com/timflutre/trimmomatic) (57) . Cleaned fastq files were then aligned to the hg19 reference genome using BWA 0.7.17 (58) and variants were called using the GATK best practices for variant calling and quality control (59). Briefly, after alignment, duplicate reads

were removed using Samtools version 1.4 (60) and the GATK v 3.8 (61) was used for base recalibration, indel realignment and then variant calling using the Global Haplotyper (61). This created raw Variant Call Format (VCF) files. The variants called per sample were then checked for their transition-transversion ratio which was approximately 2.5.

Annotation and Variant Prioritization

VCF files were annotated using ANNOVAR (62), integrating allele frequencies, from gnomAD (63), The Greater Middle Eastern Variome Server (64) and ExAC (65) , as well as functional pathogenicity scores from Polyphen2 (66) , SIFT (67) and MutationTaster (68) , MutationAssessor (69), FATHMM (70) , and CADD (71). Variants were assessed, firstly, if they passed the quality metrics as set by the Convolutional Neural Network scoring algorithm by GATK (61). Briefly, variants were compared against known variants from dbSNP and a convolutional neural network was trained to filter out poor variant calls based on several metrics (61) and minor allele frequency (MAF) in gnomAD (https://gnomad.broadinstitute.org/) of equal to or lower than 1×10^{-3} for autosomal recessive, or 1×10^{-5} for autosomal dominant/*de novo*, with a preference for variants that had no homozygotes in the gnomAD non-Neuro Cohort. For autosomal recessive variants, as the MAF used was on the conservative side, reanalysis was also performed using less stringent values (1×10^{-3}, 1×10^{-2}), to see if any variants in known ASD/ID genes were missed under the stringent analysis. Variants were then split into either homozygote or heterozygote categories and then prioritized by functional impact of mutation type. For example, LoF variants would carry more

weight than missense variants leading to assessing stopgain, frameshift, splicing, and missense variants, in that order. The LoF mutations were given highest priority as it was assumed that LoF variants will disrupt protein function more severely and, depending on the position within the transcript, cause nonsense mediated mRNA decay (NMD). The variants were then assessed for their functional pathogenicity score, where variants that met at least two functional prediction scores between Polyphen2 (66), and MutationTaster (68) and CADD (71) were further investigated. Special attention was given to splicing with additional functional annotation scores being used for splicing events with the addition of FATHMM (70) and dbscSNV (72). Due to the large number of missense variants for autosomal dominant/*de novo* inheritance, stricter criteria were used where a variant would only be considered if it was predicted as pathogenic by all software used including Polyphen2 (66) , MutationTaster (68), MutationAssessor (69), FATHMM (70), CADD(71) and M-CAP(73). The overall prioritization method is outlined in Figure 1. Variants were also analyzed to determine if they are in genes associated with any previous studies reporting association with ASD or ID. Variants in genes in pathways such as neuron growth and guidance were also considered as a higher priority. Finally, variants were Sanger-sequenced and segregation checked with both parents for validation.

Splice site annotation

Standard annotation software has been noted to be suboptimal for splicing variants and cryptic splice sites (74). To improve detection of splicing variants, particularly in neurodevelopmental disorders, SpliceAI was used to splicing variants or cryptic splice sites (75). All VCF files were run

through the standard SpliceAI algorithm as per documentation and filtered for variants with a SpliceAI score greater than 0.5. Once variants were extracted, they were annotated with ANNOVAR (76) to add additional meta data for biological interpretation. Variants were then binned into heterozygous, X-linked and homozygous and filtered by gnoMAD (77) allele frequency (MAF < 0.01). Variants that survived filtering were then investigated for appropriate inheritance patterns (i.e. *de novo* for heterozygous variants and both parents contributing one allele for homozygous variants).

Copy Number Variant (CNV) analysis

Microarrays were run using DNA from 103 ASD probands from the 115 trios (where sufficient DNA was available), including 19 using Affymetrix CytoScanHD, and 84 using Illumina CoreExome. CNV analysis for the CytoScanHD arrays was performed using ChAS and PennCNV; for CoreExome using Illumina Genome Suite/CNVpartition and PennCNV. Evidence of homozygous loss CNVs was cross-references with WES data, using the Integrated Genome Viewer (IGV: https://software.broadinstitute.org/software/igv/; (78), and CNVs corroborated in this manner were then checked by PCR.

Also, for WES data we performed CNV analysis using CLAMMS (79), XHMM (80) and CoNIFER (81). The analysis of exome sequencing data for CNVs has drawbacks due to the capture technology used. A majority of these methods employ a read based exome calling strategy after normalizing for the depth of sequencing across different regions. CLAMMS was used because it offered the ability to create subsets of the samples and construct more specific references through principal component analysis (PCA) of sequencing metrics. All softwares used were run

according to standard procedures described on their respective documentation. Filtering was done for quality metrics outlined in the documentation. Once CNVs were called they were also flited against the gnomAD structural variation call set (82) and the Database of Genomic Variants (DGV: http://dgv.tcag.ca/dgv/app/home). CNVs that overlapped more than 50% with gnomAD or DGV SVs were filtered out. Once the final call set after filtration was created, CNVs were then binned into homozygous, heterozygous or X-linked variants and further validated with qPCR to check whether they followed expected inheritance patterns for their respective zygosity. For instance, homozygous CNVs in the proband would have to have both parents that were heterozygous, and heterozygous proband variants would have to be *de novo*. X-linked variants were also further validated in males, as they lack a second copy of the X chromosome and are thus more likely to be damaging. CNV validations were also performed for family members, to confirm the inheritance pattern of candidate variants.

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 attempted to cross-reference with other available datasets. We attempted where possible to identify rare variants in the 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 Tables S1 to S3.

Biallelic variants for autism….

MSSNG: More than 10,000 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; (83). MSSNG data was accessed in April 2020.

SFARI: SSC biallelic: The Simons Simplex Collection (SSC) includes 2600 simplex autism or related developmental disorder families. Whole exome sequence 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 GPF browser (https://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.

Biallelic variants for autism….

Autism Sequencing Consortium: genes with *de novo* variants, either in the control set or case set, from the Autism Sequencing Consortium (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.

DeRubeis: genes with *de novo* variants reported in the DeRubeis et al, 2014 study, with either LoF or missense variants (84). 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), (85).

Neuroanatomical Enrichment Analysis

The Allen Adult Human Brain Atlas (86) , Brainspan (87) , and a single-cell atlas of the mouse nervous system(88) 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 (89)was used to correct for multiple tests.

Results

Our study used whole exome sequence data for a total of 115 proband/mother/father trios where the proband had a diagnosis of ASD. These included 62 from Pakistan, 40 from Iran, and 13 from Saudi Arabia. Of the probands, 91 were male, 24 were female (Table 1).

Whole exome sequencing: Of the 115 trios, 18 were found to have an autosomal biallelic mutation in a gene previously reported for ASD, ID or other neurodevelopmental disorder (see Table 2). Of these, eight were loss-of-function (LoF), or putative LoF (e.g. canonical splice site) mutations. *De novo* autosomal or X-linked mutations in known or previously reported ASD, ID, or other neurodevelopmental disorder were identified in a further seven trios (see Tables 3 and 4). A known *de novo* Rett syndrome-associated mutation (NM_004992: c.763C>T; p.Arg255*) was identified in one trio.

Biallelic mutations: We identified mutations in several genes that have been previously reported to be associated with autosomal recessive non-syndromic ID, such as *TECPR2* (41) , *MTHFR* (40) , *MADD* (41), *BTN3A2* (41), *LINS1* (39) , *CC2D1A* (MIM 608443; MRT3; (90)), *RSRC1* (91,92) and *ZNF335* (38), and a number known for syndromic or metabolic forms of autosomal recessive ID, including *VPS13B* (Cohen syndrome), *AGA* (aspartylglucosaminuria), *ASL* (arginosuccinic aciduria), *HTRA2* (3-methylglutaconic aciduria) *ASPA* (aspartoacylase deficiency/Canavan disease), and *MED25* (Basel-Vanagaite-Smirin-Yosef syndrome). *CC2D1A*, *VPS13B*, and *DEAF1* have also previously been associated with ASD or autistic features (93–95). Biallelic mutations in *DEAF1* are known to cause neurodevelopmental disorder with hypotonia, impaired expressive language, with or without seizures (MIM 617171). LoF mutations were reported for new candidate genes *HRNR*, *ILDR2*, *SCN10A*, *DDIT4L*, *SLC36A1*, *FAXDC2*, *GIMAP8*,

DAGLA, *B4GALNT1* , *VPS16*, and *EFCAB8*, and potentially LoF (such as canonical splice site mutations) in *DENND1B*, *LRRC34*, *PKD1L1*, *CNPY4*, and *PPP1R36*. *VPS16* has previously been reported for autosomal dominant dystonia (MIM 608559), *B4GALNT1* for AR spastic paraplegia (MIM 609195), and *PKD1L1* for 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 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 6 out of 125,748 gnomAD exomes.

We report a homozygous missense variant in *SGSM3*. Previously *de novo* variants in *SGSM3* have been reported for ASD (96), and SFARI lists this an ASD gene with a score of 3 (https://gene.sfari.org/).

***De novo* mutations:** *De novo* dominant mutations have been shown to be a major causal factor in ASD etiology. In this set of 87 identified WES variants from 115 trios, 21 variants were identified as *de novo* variants (18 autosomal and 3 X-linked). Among the *de novo* variants identified, variants in three of the genes have been identified in previous studies (*MECP2*, *SCN2A*, *MYT1L*, *ZNF292*). 12 were LoF variants in the genes *MECP2*, *TMIGD3*, *ECM1*, *SLAMF7*, *MYT1L*, *SCN2A*, *NCL*, *FAM53C*, *ZNF292* (2 trios), *ATP2B1*, and *RETN* (Tables 3 and 4). Four were putative LoF (splice site or stop loss) variants, including *ADGRF2*, *DGKZ*, *CLEC4E*, and *PPIL2*, and the remainder were nonsynonymous. There were also several other genes that were listed as

de novo variants that have not been associated with ASD (or ID) previously. These variants have met the criteria for filtering as described in the Materials and Method, and may represent novel putative variants for ASD.

ZNF292 (Mirzaa et al, 2020(46)**):** 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 (97). Out of 28 families reported in the study, the mutations were *de novo* in all but one, in which the mutation was inherited in dominant fashion. The clinical features associated with *ZNF292* mutations spanned a range of severities and diagnoses, including ASD and ID.

X-linked: A number of mutations were identified in well-established ASD or ID genes, for instance *MECP2* (Rett syndrome: MIM 312750), *AIFM1* (Cowchock syndrome; MIM 310490), *MID2* (XLID101: MIM 300928). For other trios there are mutations in genes that have been implicated in other recent studies of ASD and/or ID. For instance, we report a hemizygous nonsynonymous mutation in *CLTRN*- hemizygous mutations in *CLTRN* have recently been reported in several males with neutral aminoaciduria accompanied with autistic features, anxiety, depression, compulsions and motor tics (98)- a clinical picture reminiscent of the autosomal recessive Hartnup disorder (MIM 234500). For other trios there are variants in genes linked with other nervous system disorders, such as *DRP2* in Charcot-Marie-Tooth disease (99).

Biallelic variants for autism….

Splice site variants

Eight of the biallelic autosomal mutations (16%), four of the *de novo* autosomal variants (22%), and two of the X-linked recessive variants (11%) were at canonical splice positions. In addition to detection of variants at canonical splice donor and splice acceptor sites, use of SpliceAI suggests that, for the biallelic nonsynonymous variant in the known ASD/ID gene *ASL* in 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.

Overlap with studies in non-consanguineous populations: Although not focused on consanguineous populations, WGS data from the MSSNG study, as well as WES studies such as DDD and SFARI, shows possible support for several of the candidate AR genes, with possible compound heterozygous damaging rare variants in *DNAH8* (in three MSSNG individuals), *CLCA4*, *ANO10*, *SCN10A*, and *WDR90* (see Supplementary Table S1). There was also evidence of disease-causing biallelic variants in known neurodevelopmental genes *MTHFR* (one homozygous), *AGA* (one homozygous), *DEAF1* (one homozygous), *ZNF335* (one homozygous, one putative compound heterozygous), and *VPS13B* (six homozygous, nine putative compound heterozygous). There was also support for new candidate dominant/*de novo* autosomal genes, including *ECM1*, *SLAMF7*, *NCL*, *DGKZ*, *ATP2B1* (5 individuals), *CBFA2T3*, *RETN*, and *PPIL2* (see Supplementary Table S2). There was strong support for X-linked recessive (in males, maternally inherited or *de novo*) candidate gene *CNGA2* (16 individuals), *GUCY2F* (13 individuals), *SAGE1*

(eight individuals), as well as support for *EGFL6* (six individuals), *NRK* (six individuals), *PIR* (five individuals), *MSN* (five individuals), *ACTRT1* (five individuals), *TRPC5* (four individuals), and *ZNF185* (two individuals).

Copy Number Variants

A number of autosomal biallelic loss CNVs were identified using microarray data, which were then validated through observation of WES reads using IGV, and through molecular methods, using PCR and/or Sanger sequencing. Of these implicated genes, *DHRS4* and *KLK15* both have supportive evidence of neurobehavioral phenotype in mouse models (Supplementary Table S5). *DNAH7* shows strong transcription in brain tissues, and both *SHPK* and *WDR73* have high protein levels in cerebellum. Of the biallelic CNVs identified, *WDR73* has previously been associated with the autosomal recessive Galloway-Mowat syndrome 1 (MIM 251300), which includes microcephaly, delayed psychomotor development and cerebellar atrophy. A biallelic *WDR73* mutation has also been reported for ID (37). In other ASD or ID genome or exome datasets, a homozygous stop-gain and two homozygous missense mutations in *SIRPB1* were found in affected individuals in the MSSNG data (1-0157-003: Chr20: 1585484C>A; NM_001135844.2:c.655G>T:p.Glu219*; AU3915301: Chr20:1584614T>G; NM_001135844.2:c.926A>C:p.Tyr309Ser; 3-0410-000: Chr20:1585517T>A; NM_001135844.2:c.622A>T:p.Ile208Phe). None of these variants are present in the gnomAD v3.1.2 control database. For *CYP2A7*, the MSSNG data affected individuals included a homozygous frameshift insertion (1-1093-003: Chr19: 41383105dupA; NM_000764.2: c.1151dupT: p.Leu385Profs*3) and two with homozygous missense mutations (both

AU4302301 and AU008505: Chr19: 41382543C>T; NM_000764.2:c.1192G>A: p.Val398Met), neither variant present in gnomAD. While there is evidence of brain expression for *SIRPB1*, expression of *CYP2A7* appears to be restricted to liver and testes (Supplementary Table S9). However, the *WRD73*, *CYP2A7*, and *KLK15* CNVs were all identified in a proband with a biallelic splice mutation in the Canavan gene *ASPA*, the *SHPK* CNV is in an individual with a biallelic LoF mutation in known ID gene *DEAF1*, and the *SIRPB1* CNV is in an individual with a biallelic non-frameshift deletion in known gene *B4GALNT1*, and are thus less likely to be the main causative variants in these individuals (see Supplementary Table S4). A biallelic CNV loss implicates the gene *SEMG1*, however expression of this gene is restricted to the seminal vesicle (see Supplementary Table S9), 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 (Table 6).

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 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.

Discussion

Biallelic variants for autism….

Many recent studies involving NGS in ASD have involved large cohorts, focusing predominantly on dominant/*de novo* inheritance. This is largely driven by the fact that autosomal recessive variants in novel genes are more difficult to identify in the outbred population, and informatics pipelines are not sufficiently tuned to identify compound heterozygote variants. Compound heterozygous variants would need to have variants on opposite alleles, and in order to identify them would require informatics pipelines that focus on multiple damaging heterozygous variants in the same gene in a proband, and comparing with genotypes from the maternal and paternal WES or WGS data. This requires extra care when analyzing variants. The use of genomewide association studies (GWAS) to identify variants is also inadequate for finding recessive variants because the sample size required would be significantly larger than current Psychiatric Genomics Consortium studies to have sufficient power to find biallelic variants. Also, currently most ASD genetics research emanates from North American and European countries, where larger families, which are optimal for recessive variant mapping are not as readily available.

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 28 trios in genes previously associated with ASD or other neurodevelopmental disorders, resulting in a diagnostic yield of 24%. In our ASD cohort, 18 recessive, five *de novo*, and five X-linked known genes. With a further three trios with large, multi-genic, loss CNVs validated experimentally as *de novo* and thus likely pathogenic (Table 6), this yield increases to 27%. There were many other variants

identified that met filtering criteria which could potentially represent novel ASD genes or targets. Cumulatively, 18 candidate autosomal *de novo* variants were identified in the 115 trios and 50 candidate autosomal recessive variants identified in this cohort. 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.

We identified a biallelic nonsynonymous mutation in calcyphosine-like protein gene, *CAPSL*. Biallelic mutations have previously been reported for ID in both of the calcyphosine–related genes, *CAPS* (36) and *CAPS2* (41). Other than being calcium-binding EF-hand domain proteins, remarkably little is currently known about the role of calcyphosine. *EFCAB8*, in which a bialleic nonsense mutation was identified in trio SMPA9, is also a calcium-binding EF-hand domain protein. Given the possible involvement of all three calcyphosine related genes in neurodevelopmental disorders, further investigation into this biological pathway is clearly warranted.

A number of the ID genes identified here among ASD trios have also previously been reported for ASD or ASD-like features, e.g. *CC2D1A* (94,100), *VPS13B* (43), *DEAF1* (93), *ZNF335* (101), *ZNF292* (97,102,103), *MYT1L* (103,104), *SCN2A* (96,105,106), and *PLXNA3* (107,108). Of the candidate genes identified here, MSSNG, SFARI, and other datasets provide putative support (although it is not possible to determine whether two reported mutations, i.e. putative compound heterozygotes, in the same gene are on the same alleles or not, unless occurring within the same sequence read) for the biallelic genes *DNAH8*, *SCN10A*, *CLCA4*, *ANO10*, *WDR90*

(Supplementary Table S1), and for *de novo*/dominant genes *NCL*, *SLAMF7*, *ADGRF2*, *DGKZ*, *ATP2B1*, *CBFA2T3*, *RETN*, and *PPIL2* (Supplementary Table S2).

There are two genes identified here for which multiple trios from our cohort with mutations were identified: *ZNF292* and *SCN2A*. *De novo* mutations in *ZNF292* in were found in two families, including two frameshift deletions (Autism-10 and IAU65). These two trios, one from Pakistan, the other from Iran, and the frameshift mutation they share, were reported previously in Mirzaa et al, 2020 (97). *ZNF292* encodes a zinc finger protein, a transcription factor that binds to the promoter of the growth hormone gene, *GH* (109). Mutations in the known ASD gene *SCN2A* (105) were identified in two trios- one *de novo* and the other paternally inherited. In addition, a biallelic LoF mutation was identified in another voltage-gated sodium channel member (*SCN10A*), in which heterozygous gain of function missense mutations have previously linked to familial episodic pain syndrome 2 (FEPS2; MIM 615551)(110).

For some of the autosomal recessive candidate genes, there is additional support from animal models. *Rasal2* (MGI:2443881) and *Vps1*6 (MGI:2136772) knockout mice have behavioral/neurological and nervous system phenotypes (http://www.informatics.jax.org/; Supplementary Table S5). Note that VPS16 is a vesicle mediated trafficking protein similar to *VPS13B*. Homozygous knockout of *Ephb1* results in impaired contextual and cued conditioning, as well as abnormal freezing behaviour, and homozygous knockout of *Slc36a1* results in embryonic growth retardation, decreased freezing behaviour, and preweaning lethality (www.mousephenotypes.org; Supplementary Table S5). Biallelic knockout of *Dagla* results in decreased brain size, hypoactivity, abnormal behaviour, and decreased thigmotaxis

(Supplementary Table S5). In general, there is a high rate of neurodevelopmental phenotypes in knockout mice available for the genes with biallelic variants (~70%), and in biallelic or hemizygous knockouts for the X-linked genes (~58%). However, it is a remarkably contrasting story for the *de novo*/dominant variants, for which there is little or no support by way of neurodevelopmental or behavioural phenotypes in heterozygous mouse models (0%). Moreover, where there are biallelic knockout mouse models for genes from the *de novo*/dominant set, the phenotypes described are mainly unrelated to CNS and behaviour (<10%; Supplementary Table S5). This would suggest that the autosomal biallelic (and X-linked) mutations among our cohort are much more likely to be etiopathologically relevant to ASD.

There are a number of trios for which there are two or more candidate variants (Supplementary Materials Table S4). For some of these, there may be a variant that clearly delivers a more plausible etiopathological explanation. For instance, trio SMPA3 carries a biallelic nonsynonymous change in the gene *MTHFR*, mutations in which have been reported to cause recessive homocystinuria and a known cause of neurodevelopmental and psychiatric disorders. For the same trio, there is a biallelic nonsynonymous change in *RASAL2*, and a maternally inherited hemizygous LoF mutation in *MAGEB2*, as well as a biallelic loss CNV encompassing *PCDHA9* and *10*. The likelihood of disease due to mutation in *MTHFR* thus diminishes the candidacy of alternative *RASAL2*, *MAGEB2* and *PCDHA9/10*. Similarly, for trio IAH2, a biallelic LoF in *TECPR2* – a gene which has been reported for autosomal recessive spastic paraplegia 49 (SPG49; MIM 615031) as well as for autosomal recessive ID (41) and hereditary sensory neuropathy with ID (111) – also lessens the case for the PCDHA9/10 biallelic

loss CNV. For trio IABB2, the biallelic LoF mutation in *DEAF1* – a gene associated with an AR neurodevelopmental disorder with hypotonia, impaired expressive language, with or without seizures (NEDHELS; MIM 617171) – seems a highly probable candidate, and thus diminishes the possible involvement of biallelic loss CNVs at *UBE2U* and *SHPK*. The initial report for *DEAF1* was, in fact, in a consanguineous Omani family with autism, ID and epilepsy (112). A biallelic splice mutation in the Canavan disease (a severe neurodevelopmental disorder; MIM 271900) gene, *ASPA*, is the likely cause in trio IABB14, making candidate CNVs/genes *WDR73*, *KLK15*, and *CYP2A7* more likely to be benign.

For other trios with two or more variants reported, the decision process was not as clear-cut or unequivocal. Seven *de novo* autosomal mutations reported for trio RQPA20 were all validated by Sanger sequencing, however, none of these candidate genes are particularly convincing in terms of the biology, support from other human disease studies, or support from mouse models. However, expression of the genes *DGKZ* and *CBFA2T3* are associated with numerous electrophysiological phenotypes (113)(Supplementary Table S6), and show high transcription and protein expression in brain regions, which could warrant further investigation into the effects of the variants identified.

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 two additional candidate variants, in the genes *CLCA4* (gnomAD South Asian MAF=1.72E-3), and *RASAL2* (gnomAD South Asian MAF=6E-3), but none in known neurodevelopmental genes. Our methodology was

particularly stringent in regard to *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. 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.

It was recently estimated in the DDD cohort that recessive variants make up 3.6% of diagnoses, whereas *de novo* variants would contribute 48.6% of variants (44). 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 (44). The PABI cohort had 356 (333 undiagnosed) probands, 110 of whom (30.9%) had biallelic coding candidate variants, approximately half of which are in known DD genes and half in novel candidates (44). For comparison, our study reports very similar yields for recessive variants, with 16% of probands with biallelic variants in known NDD genes (18/115), or 34% including those with biallelic variants in novel candidate genes (39/115). We report 24 *de novo* variants (18 autosomal, three X-linked, three CNVs) in 15 of 115 individuals (13%), 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. Overall, the findings demonstrate the importance of autosomal recessive mutations in ASD in countries, or in cohorts, with high rates of consanguinity, as well as identifying genes that should be

examined for autozygous *and* allozygous mutations in outbred populations, just as would be done for more established autosomal recessive gene disorders.

Acknowledgements

The authors thank the families for their participation in this study. This work was supported by grants from the Canadian Institutes of Health Research to JBV (#MOP-102758 and #PJT-156402), also a National Alliance for Research on Schizophrenia and Depression (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. Computations were performed on the CAMH Specialized Computing Cluster. The SCC is funded by The Canada Foundation for Innovation, Research Hospital Fund.

The authors wish to acknowledge the resources of MSSNG (research.mss.ng), Autism Speaks and The Centre for Applied Genomics at The Hospital for Sick Children, Toronto, Canada. We also thank the participating families for their time and contributions to this database, as well as the generosity of the donors who supported this program.

DDD: The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute [grant number

WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of Wellcome or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

Conflict of Interest Statement

The authors declare no conflicts of interest in relation to the current study.

References

1. Helt M, Kelley E, Kinsbourne M, Pandey J, Boorstein H, Herbert M, et al. Can children with autism recover? If so, how? Neuropsychol Rev. 2008 Dec;18(4):339–66.

2. Mason D, Capp SJ, Stewart GR, Kempton MJ, Glaser K, Howlin P, et al. A Meta-analysis of Outcome Studies of Autistic Adults: Quantifying Effect Size, Quality, and Meta-regression. J Autism Dev Disord. 2021 Sep;51(9):3165–79.

3. Lai M-C, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. Lancet Psychiatry. 2019 Oct;6(10):819–29.

4. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychol Med. 2015 Feb;45(3):601–13.

5. Elsabbagh M, Divan G, Koh Y-J, Kim YS, Kauchali S, Marcín C, et al. Global Prevalence of Autism and Other Pervasive Developmental Disorders. Autism Res. 2012 Jun;5(3):160–79.

6. Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. 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. 2006 Jul 15;368(9531):210–5.

7. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). 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. 2014 Mar 28;63(2):1–21.

8. Kim YS, Leventhal BL, Koh Y-J, Fombonne E, Laska E, Lim E-C, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry. 2011 Sep;168(9):904–12.

9. Lundström S, Reichenberg A, Anckarsäter H, Lichtenstein P, Gillberg C. Autism phenotype versus registered diagnosis in Swedish children: prevalence trends over 10 years in general population samples. BMJ. 2015 Apr 28;350:h1961.

10. Christensen DL, Braun KVN, Baio J, Bilder D, Charles J, Constantino JN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Morb Mortal Wkly Rep Surveill Summ Wash DC 2002. 2018 Nov 16;65(13):1–23.

11. Maenner MJ. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ [Internet]. 2020 [cited 2021 Jun 5];69. Available from: https://www.cdc.gov/mmwr/volumes/69/ss/ss6904a1.htm

12. Fombonne E. Editorial: The rising prevalence of autism. J Child Psychol Psychiatry. 2018;59(7):717–20.

13. Hansen SN, Schendel DE, Parner ET. Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. JAMA Pediatr. 2015 Jan;169(1):56–62.

14. Lord C, Brugha T, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Nat Rev Dis Primer. 2020 Jan 16;6(1):5.

15. Rødgaard E-M, Jensen K, Vergnes J-N, Soulières I, Mottron L. Temporal Changes in Effect Sizes of Studies Comparing Individuals With and Without Autism: A Meta-analysis. JAMA Psychiatry. 2019 Nov 1;76(11):1124–32.

16. Arvidsson O, Gillberg C, Lichtenstein P, Lundström S. Secular changes in the symptom level of clinically diagnosed autism. J Child Psychol Psychiatry. 2018 Jul;59(7):744–51.

17. Hertz-Picciotto I, Schmidt RJ, Krakowiak P. Understanding environmental contributions to autism: Causal concepts and the state of science. Autism Res Off J Int Soc Autism Res. 2018 Apr;11(4):554–86.

18. Bölte S, Girdler S, Marschik PB. The contribution of environmental exposure to the etiology of autism spectrum disorder. Cell Mol Life Sci CMLS. 2019 Apr;76(7):1275–97.

19. Modabbernia A, Velthorst E, Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Mol Autism. 2017;8:13.

20. Taylor MJ, Rosenqvist MA, Larsson H, Gillberg C, D'Onofrio BM, Lichtenstein P, et al. Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. JAMA Psychiatry. 2020 Sep 1;77(9):936–43.

21. Ramaswami G, Geschwind DH. Genetics of autism spectrum disorder. Handb Clin Neurol. 2018;147:321–9.

22. 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 Aug 1;17(1):281.

23. Imran N, Bhatti MR, Anwar A, Najmi F, Haider II. Children's Mental Health: Pattern of referral, distribution of disorders and service use in child psychiatry outpatient setting. Pak J Med Sci. 2012;28(1):22–6.

24. Syed EU, Hussein SA, Yousafzai AW. Developing Services with Limited Resources: Establishing a CAMHS in Pakistan. Child Adolesc Ment Health. 2007;12(3):121–4.

25. Tareen A, Mirza I, Minhas A, Minhas F, Rahman A. Developing a child and adolescent mental health service in a low-income country: a global partnership model. Psychiatr Bull. 2009 May;33(5):181–3.

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.

27. Minhas A, Vajaratkar V, Divan G, Hamdani SU, Leadbitter K, Taylor C, et al. Parents' perspectives on care of children with autistic spectrum disorder in South Asia – Views from Pakistan and India. Int Rev Psychiatry. 2015 May 4;27(3):247–56.

28. Rauf NK, Anis-ul-Haq null, Aslam N, Anjum U. Characteristic symptoms and adaptive behaviors of children with autism. J Coll Physicians Surg--Pak JCPSP. 2014 Sep;24(9):658–62.

29. Mamidala MP, Kalikiri MK, Praveen Kumar PTV, Rajesh N, Vallamkonda OR, Rajesh V. Consanguinity in India and its association with autism spectrum disorder. Autism Res Off J Int Soc Autism Res. 2015 Apr;8(2):224–8.

30. Morrow EM, Yoo S-Y, Flavell SW, Kim T-K, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. Science. 2008 Jul 11;321(5886):218–23.

31. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. 2013 Jan 23;77(2):235–42.

32. Howrigan DP, Simonson MA, Davies G, Harris SE, Tenesa A, Starr JM, et al. Genome-wide autozygosity is associated with lower general cognitive ability. Mol Psychiatry. 2016;21(6):837–43.

33. Kaiser VB, Svinti V, Prendergast JG, Chau Y-Y, Campbell A, Patarcic I, et al. Homozygous loss-of-function variants in European cosmopolitan and isolate populations. Hum Mol Genet. 2015 Oct 1;24(19):5464–74.

34. Bundey S, Alam H. 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. 1993;1(3):206–19.

35. Modell B, Darr A. Science and society: genetic counselling and customary consanguineous marriage. Nat Rev Genet. 2002 Mar;3(3):225–9.

36. Harripaul R, Vasli N, Mikhailov A, Rafiq MA, Mittal K, Windpassinger C, et al. Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Mol Psychiatry. 2018;23(4):973–84.

37. Riazuddin S, Hussain M, Razzaq A, Iqbal Z, Shahzad M, Polla DL, et al. Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. Mol Psychiatry. 2017 Nov;22(11):1604–14.

38. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, et al. Genetics of intellectual disability in consanguineous families. Mol Psychiatry. 2018 Jan 4;

39. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature. 2011 Sep 21;478(7367):57–63.

40. Reuter MS, Tawamie H, Buchert R, Hosny Gebril O, Froukh T, Thiel C, et al. Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders. JAMA Psychiatry. 2017 Mar 1;74(3):293–9.

41. Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, et al. Expanding the genetic heterogeneity of intellectual disability. Hum Genet. 2017 Nov;136(11–12):1419–29.

42. Al-Mubarak B, Abouelhoda M, Omar A, AlDhalaan H, Aldosari M, Nester M, et al. Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from Saudi families. Sci Rep. 2017 Jul 18;7(1):5679.

43. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al. Using whole-exome sequencing to identify inherited causes of autism. Neuron. 2013 Jan 23;77(2):259–73.

44. Martin HC, Jones WD, McIntyre R, Sanchez-Andrade G, Sanderson M, Stephenson JD, et al. Quantifying the contribution of recessive coding variation to developmental disorders. Science. 2018 07;362(6419):1161–4.

45. Kahrizi K, Hu H, Hosseini M, Kalscheuer VM, Fattahi Z, Beheshtian M, et al. Effect of inbreeding on intellectual disability revisited by trio sequencing. Clin Genet. 2019;95(1):151–9.

46. Mirzaa GM, Chong JX, Piton A, Popp B, Foss K, Guo H, et al. De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. Genet Med. 2020 Mar;22(3):538–46.

47. Sasanfar R, Haddad SA, Tolouei A, Ghadami M, Yu D, Santangelo SL. Paternal age increases the risk for autism in an Iranian population sample. Mol Autism. 2010 Feb 22;1(1):2.

48. Rutter, Michael, Bailey A, Lord C, Cianchetti C, Fancello GS. SCQ: Social Communication Questionnaire: Manuale. OS: Giunti; 2007.

49. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision. 4th edition. Washington, DC: American Psychiatric Association; 2000. 943 p.

50. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994 Oct;24(5):659–85.

51. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord. 1989 Jun;19(2):185–212.

52. Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). J Autism Dev Disord. 1980 Mar;10(1):91–103.

53. American Psychiatric Association, American Psychiatric Association, DSM-5 Task Force. Diagnostic and statistical manual of mental disorders: DSM-5. [Internet]. 2013 [cited 2016 Dec 14]. Available from: http://dsm.psychiatryonline.org/book.aspx?bookid=556

54. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders [Internet]. World Health Organization; 2010. Available from: https://www.who.int/classifications/icd/en/bluebook.pdf

55. Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2004 May;43(5):548–58.

56. FASTX-Toolkit [Internet]. [cited 2021 Dec 18]. Available from: http://hannonlab.cshl.edu/fastx_toolkit/

57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl. 2014 Aug 1;30(15):2114–20.

58. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. 2013 May 26 [cited 2021 Dec 18]; Available from: http://arxiv.org/abs/1303.3997

59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491–8.

60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009 Aug 15;25(16):2078–9.

61. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43:11.10.1-11.10.33.

62. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015 Oct;10(10):1556–66.

63. Wang Q, Pierce-Hoffman E, Cummings BB, Alföldi J, Francioli LC, Gauthier LD, et al. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. Nat Commun. 2020 May 27;11(1):2539.

64. Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat Genet. 2016;48(9):1071–6.

65. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 18;536(7616):285–91.

66. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013 Jan;Chapter 7:Unit7.20.

67. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul 1;31(13):3812–4.

68. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010 Aug;7(8):575–6.

69. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011 Sep 1;39(17):e118.

70. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinforma Oxf Engl. 2015 May 15;31(10):1536–43.

71. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019 Jan 8;47(D1):D886–94.

72. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014 Dec 16;42(22):13534–44.

73. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016 Dec;48(12):1581–6.

74. Song L, Sabunciyan S, Florea L. CLASS2: accurate and efficient splice variant annotation from RNA-seq reads. Nucleic Acids Res. 2016 Jun 2;44(10):e98.

75. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019 Jan 24;176(3):535-548.e24.

76. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.

77. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May;581(7809):434–43.

78. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

79. Packer JS, Maxwell EK, O'Dushlaine C, Lopez AE, Dewey FE, Chernomorsky R, et al. CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. Bioinformatics. 2016 Jan 1;32(1):133–5.

80. Fromer M, Purcell SM. Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data. Curr Protoc Hum Genet. 2014 Apr 24;81:7.23.1-21.

81. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012 Aug;22(8):1525–32.

82. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. Nature. 2020 May;581(7809):444–51.

83. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, et al. Genome-wide characteristics ofde novomutations in autism. NPJ Genomic Med. 2016 Aug 3;1:160271–1602710.

84. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Nov 13;515(7526):209–15.

85. Yang C, Li J, Wu Q, Yang X, Huang AY, Zhang J, et al. AutismKB 2.0: a knowledgebase for the genetic evidence of autism spectrum disorder. Database J Biol Databases Curation. 2018 Jan 1;2018.

86. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012 Sep 20;489(7416):391–9.

87. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science. 2018 14;362(6420).

88. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018 Aug 9;174(4):999-1014.e22.

89. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995 Jan;57(1):289–300.

90. Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, Walsh CA, et al. The CC2D1A, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation. J Med Genet. 2006 Mar;43(3):203–10.

91. Perez Y, Menascu S, Cohen I, Kadir R, Basha O, Shorer Z, et al. RSRC1 mutation affects intellect and behaviour through aberrant splicing and transcription, downregulating IGFBP3. Brain J Neurol. 2018 Apr 1;141(4):961–70.

92. Maddirevula S, AlZahrani F, Anazi S, Almureikhi M, Ben-Omran T, Abdel-Salam GMH, et al. GWAS signals revisited using human knockouts. Genet Med Off J Am Coll Med Genet. 2018 Jan;20(1):64–8.

93. Nabais Sá MJ, Jensik PJ, McGee SR, Parker MJ, Lahiri N, McNeil EP, et al. De novo and biallelic DEAF1 variants cause a phenotypic spectrum. Genet Med Off J Am Coll Med Genet. 2019 Sep;21(9):2059–69.

94. Manzini MC, Xiong L, Shaheen R, Tambunan DE, Di Costanzo S, Mitisalis V, et al. CC2D1A regulates human intellectual and social function as well as NF-κB signaling homeostasis. Cell Rep. 2014 Aug 7;8(3):647–55.

95. Douzgou S, Petersen MB. Clinical variability of genetic isolates of Cohen syndrome. Clin Genet. 2011 Jun;79(6):501–6.

96. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012 Apr 4;485(7397):237–41.

97. Mirzaa GM, Chong JX, Piton A, Popp B, Foss K, Guo H, et al. De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. Genet Med Off J Am Coll Med Genet. 2019 Nov 14;

98. Pillai NR, Yubero D, Shayota BJ, Oyarzábal A, Ghosh R, Sun Q, et al. Loss of CLTRN function produces a neuropsychiatric disorder and a biochemical phenotype that mimics Hartnup disease. Am J Med Genet A. 2019 Dec;179(12):2459–68.

99. Berciano J, García A, Gallardo E, Peeters K, Pelayo-Negro AL, Álvarez-Paradelo S, et al. Intermediate Charcot-Marie-Tooth disease: an electrophysiological reappraisal and systematic review. J Neurol. 2017 Aug;264(8):1655–77.

100. Sener EF, Onal MG, Dal F, Nalbantoglu U, Ozkul Y, Canatan H, et al. Novel alterations of CC2D1A as a candidate gene in a Turkish sample of patients with autism spectrum disorder. Int J Neurosci. 2020 Dec 28;1–8.

101. Wu J, Yu P, Jin X, Xu X, Li J, Li Z, et al. Genomic landscapes of Chinese sporadic autism spectrum disorders revealed by whole-genome sequencing. J Genet Genomics Yi Chuan Xue Bao. 2018 Oct 20;45(10):527–38.

102. 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.

103. Wang T, Guo H, Xiong B, Stessman HAF, Wu H, Coe BP, et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun. 2016 Nov 8;7:13316.

104. Coursimault J, Guerrot A-M, Morrow MM, Schramm C, Zamora FM, Shanmugham A, et al. MYT1L-associated neurodevelopmental disorder: description of 40 new cases and literature review of clinical and molecular aspects. Hum Genet. 2021 Nov 8;

105. Weiss LA, Escayg A, Kearney JA, Trudeau M, MacDonald BT, Mori M, et al. Sodium channels SCN1A, SCN2A and SCN3A in familial autism. Mol Psychiatry. 2003 Feb;8(2):186–94.

106. Jiang Y, Yuen RKC, Jin X, Wang M, Chen N, Wu X, et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J Hum Genet. 2013 Aug 8;93(2):249–63.

107. Steele JL, Morrow MM, Sarnat HB, Alkhunaizi E, Brandt T, Chitayat DA, et al. Semaphorin-Plexin Signaling: From Axonal Guidance to a New X-Linked Intellectual Disability Syndrome. Pediatr Neurol. 2021 Oct 18;126:65–73.

108. Wang T, Zhang Y, Liu L, Wang Y, Chen H, Fan T, et al. Targeted sequencing and integrative analysis of 3,195 Chinese patients with neurodevelopmental disorders prioritized 26 novel candidate genes. J Genet Genomics Yi Chuan Xue Bao. 2021 Apr 20;48(4):312–23.

109. Lipkin SM, Näär AM, Kalla KA, Sack RA, Rosenfeld MG. Identification of a novel zinc finger protein binding a conserved element critical for Pit-1-dependent growth hormone gene expression. Genes Dev. 1993 Sep;7(9):1674–87.

110. Faber CG, Lauria G, Merkies ISJ, Cheng X, Han C, Ahn H-S, et al. Gain-of-function Nav1.8 mutations in painful neuropathy. Proc Natl Acad Sci U S A. 2012 Nov 20;109(47):19444–9.

111. Neuser S, Brechmann B, Heimer G, Brösse I, Schubert S, O'Grady L, et al. Clinical, neuroimaging, and molecular spectrum of TECPR2-associated hereditary sensory and autonomic neuropathy with intellectual disability. Hum Mutat. 2021 Jun;42(6):762–76.

112. Rajab A, Schuelke M, Gill E, Zwirner A, Seifert F, Morales Gonzalez S, et al. Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. J Med Genet. 2015 Sep;52(9):607–11.

113. Bomkamp C, Tripathy SJ, Bengtsson Gonzales C, Hjerling-Leffler J, Craig AM, Pavlidis P. Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes. PLoS Comput Biol. 2019 Jun;15(6):e1007113.

114. Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, et al. Expanding the genetic heterogeneity of intellectual disability. Hum Genet. 2017 Nov;136(11–12):1419–29.

115. Kaiser VB, Svinti V, Prendergast JG, Chau Y-Y, Campbell A, Patarcic I, et al. Homozygous loss-of-function variants in European cosmopolitan and isolate populations. Hum Mol Genet. 2015 Oct 1;24(19):5464–74.

Table 1: Cohort Description.

Cohort ID (Recruited By)	Institute	Nationality	Number of Trios (Number of DNAs)	Proband Male: Female
SMPA (Ansa Rabia, Saqib Mahmood, Shazia Maqbool)	University of Health Sciences and Children Hospital, Lahore	Pakistan	54 (162)	42:12
IAU (Dr. Sasanfar)	Special Education Organization	Iran	27 (81)	23:4
IABB (Dr. Bita Bozorgmehr)	Shahid Beheshti Medical University	Iran	10 (30)	9:2
SADD (Laila Al Ayadhi)	King Saud University	Saudi Arabia	13 (39)	12:1
Autism (Raheel Qamar, Maleeha Azam and Zehra Agha)	COMSATS University, Islamabad	Pakistan	4 (12)	4:4
RQPA (Raheel Qamar, Maleeha Azam and Zehra Agha)			4 (12	
IAH (Abolfazl Heidari)	Qazvin Medical University, Qazvin	Iran	3 (9)	2:1
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 for ASD or ID are in bold type. Control population frequencies show are from gnomAD (all, and for South Asian sub-population). All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing (except for IAU80/BTN3A2, for which unique PCR primers could not be designed). Validations are shown in Supplementary Materials.

Variant	Family	Chr.	Coordinates (hg19)	Ref.	Alt.	Gene	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomAD S. Asian Freq.
1	SMPA3	1	11855404	C	A	MTHFR (40)(MIM 236250; SFARI score 3)	nonsynonymous	NM_001330358.2: exon6: c. 905G>T: p. Gly302Val	4.07E-06	3.25E-05
2	IABB16	1	87029377	G	A	**CLCA4**	nonsynonymous	NM_012128.3:c.482G>A:p.Arg161Gln	4.11E-04	1.72E-03
3	SMPA4	1	152193673-152193682	GACGTTTCTG	-	**HRNR**	frameshift del	NM_001009931.3: exon3: c.423_432del: p.Arg142Glufs*15	4.467E-05	4.0E-4
4	SMPA76	1	166891853-166891854	CT	-	**ILDR2**	frameshift del	NM_199351.3: exon8: c.1187_1188del: p. Glu396Glyfs*39)	0	4.07E-06
5	SMPA3	1	178425880	C	A	**RASAL2**	Nonsynonymous	NM_004841.3: exon11: c.1813C>A: p. Leu605Ile	8.0E-4	6E-3
6	SMPA40	1	197621364	C	A	**DENND1B**	Splicing	NM_001195215.1: exon7: c.447+1G>T	0	0
7	Autism8	2	74759786	G	T	HTRA2 (MIM 617248)	Nonsynonymous	NM_013247.3: exon7: c.1156G>T: p. Asp386Tyr	8.12E-06	6.50E-05
8	SMPA19	3	38802185	C	A	**SCN10A**	Stopgain	NM_001293306.2: exon7: c.937G>T; p. Gly313*	0	0
9	SMPA76	3	43593070	C	A	**ANO10** (spinocerebellar ataxia 10, MIM 613728)	Nonsynonymous	NM_001346463.2:exon11:c.1679G>T:p.Cys560Phe	0	0
10	SA15	3	47164429	G	C	SETD2 (LLS (AD) MIM 616831; SFARI score 1)	Nonsynonymous	NM_001349370.3: exon2:c.1565C>G: p. Ser522Cys	0	0
11	SMPA27	3	134911657	T	C	**EPHB1**	Nonsynonymous	NM_004441.3: exon11: c.2122T>C: p. Phe708Lys	4.07E-06	3.25E-05

12	Autism3	3	157840002	C	T	RSRC1 (MRT70; MIM 618402; SFARI score S)	Stopgain	NM_001271834.2: exon2:c.109C>T: p. Arg37*	0	0
13	SMPA8	3	169524635	C	T	**LRRC34**	Splicing	NM_001172779.1: exon5: c.528+1G>A	0	0
14	IAU66	4	101108876	TT	C	**DDIT4L**	frameshift del	NM_145244.3: exon3: c.539_540delinsG: p. Lys180Serfs*5	0	0
15	RQPA12	4	178360822	G	A	AGA (MIM 208400)	Nonsynonymous	NM_000027.3: exon3: c.302C>T: p. Ala101Val	0	4.06E-06
16	SMPA35	4	185033930	CAT	C	**ENPP6**	frameshift del	NM_153343.3: exon6: c.886_887del: p.Met296Aspfs*28	0	0
17	IABB11	5	35910583	C	T	**CAPSL**	Nonsynonymous	NM_144647.3:c.200G>A:p.Gly67Glu	0	0
18	Autism8	5	150846823	-	CTGG	**SLC36A1**	frameshift ins	NM_001308150.2: exon6:c.483_484insCTGG: p. Asp164Glyfs*2	8.01E-06	6.6E-05
19	IAU56	5	154199963	C	-	**FAXDC2**	frameshift del	NM_032385.3: exon9: c.915delG: p.Thr306Profs*98	5.35E-05	4E-4
20	SMPA48	6	13634723	C	G	**RANBP9**	Nonsynonymous	NM_005493.3: exon11: c.1735G>C: p. Gly579Arg	0	2.44E-05
21	IAU80	6	26368279	G	A	BTN3A2 (114)	Splicing	NM_001197247.2: exon2: c.-1+1G>A (5'UTR)	0	0
22	IABB8	6	38851674	A	T	**DNAH8**	Nonsynonymous	NM_001206927.2: exon56: c.8159A>T: p. Asp2720Val	0	0
23	Autism9	6	170058095	C	A	**WDR27**	Nonsynonymous	NM_001350625.2: exon13: c.1246G>T: p. Gly416Trp	0	0
24	IAU7	7	47944924	T	C	**PKD1L1** (AR laterality defects; MIM 616205)	Splicing	NM_138295.3: exon11: c.1523-2A>G	0	0
25	SA7	7	65551627	C	T	ASL (MIM 207900)#	Nonsynonymous	NM_001024943.2: exon6:c.502C>T: p. Arg168Cys	0	0
26	IABB7	7	99720009	G	C	**CNPY4**	Splicing	NM_152755.1: exon2: c.245+1G>C	0	0
27	IAU79	7	150174821	TCC	T	**GIMAP8**	frameshift del	NM_175571.3: exon5: c.1952_1953del: p. Gln652Serfs*6	0	0
28	SMPA11	7	155604611	T	A	SHH (HPE3 (AD) MIM 142945)	Nonsynonymous	NM_000193.4: exon1: c.206A>T: p. Asn69Ile	2.39E-06	1.96E-4
29	IABB3	8	100155379	AG	A	VPS13B (COH1 MIM 216550; SFARI score 1)	frameshift del	NM_015243.3: exon13: c.1829delG: p. Ser610Thrfs*3	0	0
30	SA4	9	5805613	T	C	**ERMP1**	Nonsynonymous	NM_024896.3: exon9: c.1721A>G: p.His574Arg	0	0

31	IABB2	11	679701	G	GA	DEAF1 (MIM 617171; SFARI score 1))	frameshift ins	NM_001293634.1: exon6: c.845dupT: p. Ala283Argfs*20	0	0
32	SMPA11	11	6477889	C	A	**TRIM3**	Nonsynonymous	NM_033278.3:exon6:c.1067G>T:p.Arg356Leu	0	0
33	SMPA20	11	47305749	A	G	MADD (38,114)	Nonsynonymous	NM_001135943.2: exon11: c. 1883A>G: p. Tyr628Cys	0	0
34	SMPA21	11	61495635	C	G	**DAGLA** (SCA20; MIM 608687; SFARI score 3)	Stopgain	NM_006133.3: exon7:c.647C>G: p. Ser216*	0	0
35	SMPA80	11	67290028	C	T	**CABP2** (Deafness, MIM 60314)	Nonsynonymous	NM_016366.3: exon2: c.202G>A: p. Ala68Thr	0	6.60E-05
36	SMPA8	11	118403863	A	G	**TMEM25**	Nonsynonymous	NM_001144036.2: exon3: c. 302A>G: p. Asn101Ser	0	0
37	IAU79	12	58021505	GGGAAGCCGAC GAGCTCGTGGT	G	**B4GALNT1** (Spastic paraplegia 26; MIM 609195)	nonframeshift del	NM_001276468.2: exon9: c.1094_1114 del: p.His365_Phe372del	0	0
38	IABB10	12	69263140	C	G	**CPM**	Nonsynonymous	NM_001005502.2: exon6: c.752G>C: p. Gly251Ala	0	0
39	SMPA8	14	65032074	G	C	**PPP1R36**	Splicing	NM_172365: exon5: c.270-1G>C	0	0
40	IAH2	14	102963429	TGTGGGGACCGC CTG	T	TECPR2 Anazi et al, 2017 PMID 27431290); SPG49 MIM 615031	frameshift del	NM_014844.3: exon18: c.3904_3917del: p. Val1302Glyfs*60	0	0
41	IABB4	15	101113962	T	-	LINS1 (MRT27; MIM 614340)	frameshift del	NM_001040616.3: exon5: c.1116delA: p. Glu372Aspfs*9	0	0
42	SMPA76	16	716004	C	T	**WDR90**	Nonsynonymous	NM_145294.4: exon36: c.4489C>T: p. Arg1497Trp	3.24E-05	0
43	IABB14	17	3392637	G	T	ASPA (Canavan disease; MIM 271900)	Splicing	NM_001128085: exon5: c.634+1G>T	0	0
44	SA2	19	14024452	G	T	CC2D1A (MRT3; MIM 608443; SFARI score 2)	Splicing	NM_017721: exon6: c.748+1G>T	0	/0
45	IABB16	19	50321659	G	A	MED25 (BVFYS; MIM 616449)	Nonsynonymous	NM_030973.3: exon1: c.61G>A: p. Val21Met	0	0
46	IABB16	20	2847201	C	A	**VPS16** (Dystonia (AD) MIM 619291)	Nonsynonymous	NM_080413.3: exon20: c.2069C>A: p. Ala690Glu	0	0
47	IABB16	20	16360737	A	G	KIF16B (PMID 29736960)	Nonsynonymous	NM_001199865.2: exon19: c. T1910T>C: p. Val637Ala	0	0

48	SMPA9	20	31531544	C	T	**EFCAB8**	Stopgain	NM_001143967.2: exon23:c.2893C>T: p. Gln965*	0	0
49	IABB5	20	44578999	C	T	ZNF335 Hu et al, 2019; MCPH10; MIM 615095)	Nonsynonymous	NM_022095.3: exon22: c.3346G>A: p. Gly1116Arg	0	0
50	IABB11	22	40803252	G	C	**SGSM3** (SFARI score 3)	Nonsynonymous	NM_001350039.2:c.1288G>C:p.Glu430Gln	1.99E-05	1.63E-4

SpliceAI predicts the generation of a splice donor site.

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 show are from gnomAD (all, and for South Asian sub-population). All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing (except for SMPA39/CLEC4E, for which DNA was exhausted). Validations are shown in Supplementary Materials.

Variant	Family	Chr	Coordinates (hg19)	Ref.	Alt.	Gene (previously reported)	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomAD S. Asian Freq.
1	RQPA20	1	28919246	G	A	**RAB42**	nonsynonymous	NM_001193532.3: exon1: c. 223G>A: p. Glu75Lys	0	0
2	RQPA20	1	112031311	A	-	**TMIGD3**	frameshift del	NM_001302680.2: exon2: c.286delT: p. Trp96Glyfs*36	0	0
3	RQPA20	1	150483966	G	T	**ECM1** (DDDG2P: lipoid proteinosis, AR)	Stopgain	NM_001202858.2: exon7: c.823G>T: p. Glu275*	0	8.12E-06
4	RQPA20	1	160722912	C	A	**SLAMF7**	Stopgain	NM_001282589.2: exon5:c.560C>A: p. Ser187*	0	0
5	IAU29	2	1946865	C	-	MYT1L (MRD39: MIM 616521)	frameshift del	NM_001303052.2: exon9: c.394delG: p. Glu132Argfs*42	0	0
6	SMPA53	2	166166958	C	T	SCN2A (DEE11: MIM 6137219)	Stopgain	NM_001040143.2: exon6:c.823C>T: p. Arg275*	0	0
7	SMPA38	2	166201156	C	A	SCN2A (DEE11: MIM 6137219)	Nonsynonymous	NM_001040143.2:exon15:c.2654C>A:p.Thr885Asn	0	0
8	IAU66	2	232320177	C	-	**NCL**	frameshift del	NM_005381: exon13: c.1991delG: p. Gly664Glufs*70	0	0
9	SMPA75	5	137680632	-	T	**FAM53C**	frameshift ins	NM_001135647.2: exon4: c.256dupT: p. Ser86Phefs*14	0	0

10	SMPA75	6	47646871	G	A	**ADGRF2**	Splicing	NM_153839: exon5: c.267+1G>A	4.07E-06	3.25E-05
11	Autism10	6	87969506-87969507	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p. Glu2054Lysfs*14	0	0
12	IAU65	6	87969506-87969507	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p. Glu2054Lysfs*14	0	0
13	RQPA20	11	46389550	A	C	**DGKZ**	Splicing	NM_201533: exon4: c.379-2A>C	0	0
14	SMPA39	12	8687235	-	A	**CLEC4E**	Stoploss	NM_014358.3: exon6: c.658dupT: p.*220Leuext*36	1.22E-05	9.81E-05
15	SMPA73	12	89998008	G	C	**ATP2B1**	Stopgain	NM_001001323.2: exon15:c.2558C>G: p. Ser853*	0	0
16	RQPA20	16	89006655	G	A	**LOC100129697/CBFA2T3**	Nonsynonymous	NM_001290330.2: exon1: c.53G>A: p. Arg18Lys	0	0
17	SMPA75	19	7734220-7734221	CT	-	**RETN**	frameshift del	NM_001193374.1: exon2: c.8_9del: p. Cys5Serfs*13	4.47E-05	0.0003
18	RQPA20	22	22051034	A	G	**PPIL2**	Splicing	NM_148175: exon21: c.1570-2A>G	0	0

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 show are from gnomAD (all, and for South Asian sub-population). 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 (hg19)	Ref	Alt	Gene (previous reports)	Variant Type	cDNA/Amino Acid Change	gnomAD All Freq.	gnomAD S. Asian Freq.
1	SMPA43	Recessive	13621531	GC	G	**EGFL6**	frameshift del	NM_001167890.1: exon5: c.497delC: p. Pro167Glnfs*7	0	0
2	SMPA27	Recessive	15474108	G	T	**PIR**	Nonsynonymous	NM_003662.4:c.343C>A:p.Gln115Lys	5.60E-05	0
3	SMPA3	Recessive	30236779	C	-	**MAGEB2**	frameshift del	NM_002364.4:c.82delC:p.Pro28Leufs*84	0	0
4	IAU69	Recessive	55779942	T	C	**RRAGB**	Nonsynonymous	NM_006064.4: exon7: c.730T>C: p. Phe244Leu	0	0
5	IAU45	Recessive	64947718	C	A	**MSN**	Nonsynonymous	NM_002444.3: exon3: c.139C>A: p. Leu47Met	0	0
6	SMPA15	Recessive	105193588	C	T	**NRK**	Stopgain	NM_198465.4: exon27: c.4375C>T: p. Gln1459*	0	0
7	IAU47	Recessive	107170098	A	G	MID2 (MRX101; MIM 300928)	nonsynonymous	NM_012216.4:c.2003A>G:p.Tyr668Cys	0	0

8	SMPA15	Recessive	108625363	C	T	**GUCY2F**	nonsynonymous	NM_001522.3: exon17: c.3134G>A: p. Gly1045Glu	6.25E-05	0.0006
9	Autism6	Recessive	111019623	A	G	**TRPC5**	nonsynonymous	NM_012471.3: exon11: c.2840T>C: p. Phe947Ser	4.59E-05	0.0004
10	IAU78	Recessive	127185449	C	T	**ACTRT1**	nonsynonymous	NM_138289.4: exon1: c.737G>A: p. Gly246Glu	0	0
11	IAU24	Recessive	129290446	C	T	AIFM1 (Cowchock syndr.; MIM 310490)	nonsynonymous	NM_001130847.4: exon2: c. 238G>A: p. Ala80Thr	0	0
12	SMPA35	Recessive	134987552	G	A	**SAGE1**	Splicing	NM_018666: exon5: c.454+1G>A	0	0
13	IAU51	Recessive	140994661	-	A	**MAGEC1**	frameshift ins	NM_005462.3:c.1471_1472insA; p.Leu491Tyrfs*10	4.37E-05	0
14	SMPA70	Recessive	150912120	G	A	**CNGA2**	nonsynonymous	NM_005140.3: exon7: c.1145G>A: p. Arg382Gln	0	0
15	IABB8	Recessive	152106616	A	G	**ZNF185**	Splicing	NM_001178110: exon13: c.942-2A>G	0	0
16	SMPA41	Recessive	153689047	C	T	PLXNA3 (ASD: PMID: 33994118; SFARI score 3)	nonsynonymous	NM_017514.3: exon2: c.524C>T: p. Ser175Leu	4.62E-05	7.37E-05
17	SA1	De Novo	15677205	T	A	**CLTRN (TMEM27)**	nonsynonymous[&]	NM_020665.3:exon3:c.137A>T:p.Glu46Val	0	0
18	SMPA84	De Novo	100507601	A	C	DRP2 (CMTX1 MIM 302800; PMID 26227883)	nonsynonymous	NM_001171184.2: exon15: c. 1639A>C: p. Ser547Arg	0	0

19	SMPA78	De Novo	153296516	G	A	MECP2 (RTT; MIM 312750; SFARI score 1)	Stopgain	NM_001110792.2:exon3:c.799C>T:p. Arg267*	0	0

[&]MutationTaster predicts splice donor created

Table 5: CNV summary: validated biallelic loss CNVs called by CNVpartition or ChAS. Biallelic losses called from microarray data by PennCNV and from WES data by CLAMMS (79)) 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 hg19.

Family	Mode of Inheritance	(hg19)	Cytoband	Size (Kb)	CN	CNV confidence	Genes	FREQ IN DGV GOLD STANDARD, gnomAD, or rare?	IGV	PCR
SMPA4	Recessive	14:24,454,924-24,464,325	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: 43836785-43836964	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	2:196,799,394-196,801,362	2q32.3	1.969	0	139.6294	DNAH7 exon 20-21	NONE	Y	Y
SMPA3	Recessive	5:140,224,038-140,242,319	5q31.3	18.282	0	1417.866	PCDHA9, PCDHA10 (SFARI score 3)	Also Hom in mother	Y	Y
IAH2	Recessive	5:140,224,038-140,242,319	5q31.3	18.282	0	1417.866	PCDHA9, PCDHA10 (SFARI score 3)	Also Hom in father	Y	Y

IAU79	Recessive	20:1,559,259-1,559,330	20p13	0.072	0	171.6751	SIRPB1 exon 2	Partial OVERLAP WITH COMMON CNV (0.0894)	Y	Y
IABB2	Recessive	17:3513856- 3514054	17p13.2	0.198	0	318.3258	SHPK exon 7	Partial OVERLAP WITH 80Kb gssvL47583 (0.00274)	y	Y
IABB14	Recessive	15:85186877-85186894	15q25.2	0.018	0	55.68398	WDR73 exon 8$	rs760394400; gnomAD MAF 0.1672	Y	Y
IABB14	Recessive	19:41356319-41386033	19q13.2	29.714	0	579.0146	CYP2A7 6 exons	gssvL58578 0.002 Plus gssvL58575 0.0017	Y	Y
IABB14	Recessive	19: 51330951-51331063	19q13.33	0.112	0	579.0146	KLK15 exon 2	Overlap with gssvL59230 0.02	Y	Y
SMPA42	X-linked	X:47912041-47970683	Xp11.23	58.643	0	1665.299	ZNF630; SSX6	Deletions in ZNF630 do not segregate with ID; PMID 20186789	Y	
SMPA9	X-linked	X:47917867-47970683	Xp11.23	52.817	0	1558.271	ZNF630; SSX6	Ditto	Y	

$ see Riazuddin et al, 2017; PMID 30171209)

Table 6: CNV summary: likely pathogenic, high confidence CNVs. Coordinates given using hg19. Comparison with control population CNVs used the Database of Genomic Variants (DGV) Gold, accessed through the Decipher genomic browser (www.deciphergenomics.org).

Family	**Mode of Inheritance**	**(hg19)**	**Cytoband**	**Size (Kb)**	**CN**	**CNV confidence**	**Genes**	**DGV gold standard**	**PCR/ RT-PCR**	**Comments**
SMPA80	Paternal	9:370,244-13,398,587	9p24.3-p23	13,028	1	809.7327	Many	N	Y	
IAU25	De novo	17:16,778,164-20,364,207	17p11.2	3,586	1	693.2049	Many	N	Y	Unaffected sibling also tested negative
IAU74	De novo	2:235,370,826-243,034,519	2q37.1-37.3	7,664	1	4781.854	Many	N	Y	
IAU7	De novo	1:249357-699532	19p13.3	450.176	1	547.5059	Many	N	Y	Distal probe (PPAP2C) validated, proximal

										probe (HCN2) not so clear cut.
SMPA15	Paternal	15:91,290,633-91,347,451	15q26.1	57	1		BLM		Y	
SMPA15	Paternal	15:40,477,419-41,026,994	15q15.1	550	1		Many		Y	CASC5: Y PAK6:Y

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.