Biallelic variants for autism....

1 Autism spectrum disorder trios from consanguineous populations are enriched

2 for rare biallelic variants, identifying 32 new candidate genes

3 Running title: Biallelic variants for autism and consanguinity

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

- 13 ¹Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health
- 14 Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada;
- ¹⁵²Institute of Medical Science, University of Toronto, Toronto, ON, Canada;
- ¹⁶ ³University of Health Sciences, & NUR International University, Lahore, Pakistan;
- 17 ⁴CMH Lahore Medical College & Institute of Dentistry, Lahore (affiliated with National
- 18 University of Medical Sciences, Rawalpindi), Pakistan;
- ⁵Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan;

- ⁶Sana Medical Genetics Laboratory, Qazvin, Iran; Reference Laboratory of Qazvin Medical
- 21 University, Qazvin, Iran;
- ⁷Science and Environment Sector, ICESCO, Rabat, Kingdom of Morocco;
- ²³ ⁸Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON,
- 24 Canada;
- ⁹Department of Psychiatry, State University of New York at Buffalo; Jacobs School of Medicine
- and Biomedical Sciences, Buffalo, NY, USA;
- ¹⁰Department of Physiology, College of Medicine, King Saud University, Riyadh, Saudi Arabia;
- ¹¹Department of Educational Sciences, Farhangian University, Tehran, Iran;
- ¹²Center for Psychiatric Research, Maine Medical Center Research Institute, 509 Forest Avenue,
- 30 Portland, ME, 04101, USA;
- ¹³Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA, 02110, USA;
- ¹⁴Department of Psychiatry, Maine Medical Center, 22 Bramhall Street, Portland, ME, 04102,
- 33 USA;
- ¹⁵Autism Research & Treatment Center, Department of Physiology, Faculty of Medicine,
- 35 National Plan for Science and Technology (NPST), King Saud University, Riyadh, Saudi Arabia;
- ¹⁶Department of Psychiatry, Harvard Medical School, Boston, MA, USA;
- ¹⁷The University of Child Health Sciences, Children's Hospital, Lahore, Pakistan;
- ¹⁸Human Genetics Institute of New Jersey and Department of Genetics, Rutgers University,
- 39 Piscataway, NJ, USA;
 - ¹⁹Department of Psychiatry, Queens University Kingston, Kingston, ON, Canada;

Biallelic variants for autism....

²⁰Division of Psychiatry, University College London, London, UK;

- 40 ²¹Department of Psychiatry, University of Toronto, Toronto, ON, Canada.
- 41 §Equal contribution
- 42 *Correspondence should be addressed to:
- 43 Dr. John B. Vincent, Brain Molecular Science, Campbell Family Mental Health Research Institute,
- 44 Centre for Addiction and Mental Health (CAMH), 250 College Street, Toronto, ON, Canada, M5T
- 45 1R8; Email: john.vincent@camh.ca
- 46 Dr. Muhammad Ayub, Division of Psychiatry, University College London, 149 Tottenham Ct Rd,
- 47 London W1T 7NF, United Kingdom. Email: <u>m.ayub@ucl.ac.uk</u>

Biallelic variants for autism....

49 Abstract

50	Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects
51	about 1 in 36 children in the united states, imposing enormous economic and socioemotional
52	burden on families and communities. Genetic studies of ASD have identified de novo copy
53	number variants (CNVs) and point mutations that contribute significantly to the genetic
54	architecture, but the majority of these studies were conducted in populations unsuited for
55	detecting autosomal recessive (AR) inheritance. However, several ASD studies in
56	consanguineous populations point towards AR as an under-appreciated source of ASD variants.
57	Methods: We used whole exome sequencing to look for rare variants for ASD in 115 proband-
58	mother-father trios from populations with high rates of consanguinity, namely Pakistan, Iran,
59	and Saudi Arabia. Consanguinity was assessed through microarray genotyping.
60	Results: We report 84 candidate disease-predisposing single nucleotide variants and indels,
61	with 58% biallelic, 25% autosomal dominant/ <i>de novo</i> , and the rest X-linked, in 39 trios. 52% of
62	the variants were loss of function (LoF) or putative LoF (pLoF), and 47% nonsynonymous. We
63	found an enrichment of biallelic variants, both in sixteen genes previously reported for AR ASD
64	and/or intellectual disability (ID) and 32 previously unreported AR candidate genes (including
65	DAGLA, ENPP6, FAXDC2, ILDR2, KSR2, PKD1L1, SCN10A, SHH, and SLC36A1). We also identified
66	eight candidate biallelic exonic loss CNVs.

67 Conclusions: The significant enrichment for biallelic variants among individuals with high F_{roh}
 68 coefficients, compared with low F_{roh}, either in known or candidate AR genes, confirms that

Biallelic variants for autism....

- 69 genetic architecture for ASD among consanguineous populations is different to non-
- consanguineous populations. Assessment of consanguinity may assist in the genetic diagnostic
- 71 process for ASD.
- 72 Keywords: Autism spectrum disorder; Trio; Autosomal recessive; Consanguinity; Biallelic;
- 73 Candidate Genes
- 74

Biallelic variants for autism....

76 Introduction

- 77 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
- 80 a recent meta-analysis of longitudinal studies (2). Phenotypically, individuals with ASD are
- 81 heterogeneous in their level of intellectual functioning, and co-morbid psychiatric and
- 82 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%
- 84 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,
- 86 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), and increasing to one in 36 in the most recent report (11). This increase is partly due to changes in diagnostic criteria, reporting practices, and increased awareness (12–15).

A systematic review of literature from South Asian countries reported a prevalence of ASD in the range of 0.09-1.07 % (16). 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% (17–

Biallelic variants for autism....

97	20). Clinical services and special schools for ASD are available in Pakistan, but are typically limited
98	to major cities (21). The literature on clinical presentation is limited. One study with a small
99	sample size shows symptoms consistent with studies elsewhere in the world (22).
100	Consanguineous marriages lead to a marked increase in the frequency of rare recessive
101	disorders (23,24). Many genetic variants will only be pathogenic in recessive form; this includes
102	variants for ID, and almost certainly for ASD too. Populations with a high proportion of
103	consanguinity have been important for describing autosomal recessive genes in ID in Pakistan
104	(25,26), Iran (27,28), Syria (29), and Saudi Arabia (30), yet, to date, relatively few reports of AR
105	genes underlying ASD in consanguineous populations have been published (31–33). Previous
106	work has highlighted recessive inheritance as an important component of the genetic
107	architecture of ASD. For instance, a large study of consanguineous versus non-consanguineous
108	families in India concluded that consanguinity increases the risk for ASD with an odds ratio of
109	3.22 (34). Morrow and colleagues used SNP microarrays to map homozygous loci in 104 small
110	ASD families from the Middle East, Turkey and Pakistan, finding homozygous deletions
111	implicating SLC9A9, PCDH10, CNTN3 and others (31).
112	Research in outbred populations also supports AR inheritance as an important piece of
113	the genetic puzzle for ASD. For example, it has been estimated that loss-of-function (LoF)
114	recessive mutations contribute 3% of ASD genetics in two US-based case-control cohorts (35).
115	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

117 data are non-optimal for discovering compound heterozygous mutations.

Biallelic variants for autism....

118	Despite the higher prevalence of recessive inheritance in consanguineous populations,
119	data from studies of developmental disorder (DD) suggest that <i>de novo</i> variants are also
120	prominent, albeit in a lower proportion e.g. in a UK study of 6040 families from the Deciphering
121	Developmental Disorders (DDD) study, individuals of Pakistani ancestry had 29.8% de novo
122	compared with 49.9% in the European ancestry UK population (38). In an Iranian ID study, the
123	<i>de novo</i> rate was 27.86% (39).
124	Describing the genetic architecture of ASD is an important step towards
125	understanding the pathogenesis of the disorder. There has been significant progress to date,
126	and studies of ASD have shown that <i>de novo</i> copy number variants (CNVs) and point mutations
127	play a prominent role, however, the majority of these studies have been conducted in outbred
128	populations (40). Here, we report on a study of 84 genetic and 11 genomic variants (eight
129	homozygous; 3 heterozygous) from a cohort of 115 ASD trios from three countries with a high
130	frequency of consanguineous marriages (Pakistan, Iran, and Saudi Arabia).
131	Materials and Methods
132	Trio Family Ascertainment
133	Institutional Research Ethics Board approval was received from the Centre for Addiction and

Mental Health (CAMH) and the recruitment sites. 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.

Biallelic variants for autism....

138 Whole Exome Sequencing, Alignment and Variant Calling

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

149 Microarray genotyping and analysis.

150 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 151 152 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 153 generate an F-coefficient of consanguinity for the probands (and affected siblings), F_{ROH}, as 154 described in McQuillan et al, 2008 (41), where $F_{roh} = \sum L_{roh}/L_{auto}$. Autosome size L_{auto} was 155 156 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 157 were run on both Illumina CoreExome and Affymetrix CytoScanHD arrays (probands from trios 158

Biallelic variants for autism....

- 159 IABB2, IABB3, IABB4, and IABB5), we were able to generate both Froh and LoH scores, which, as
- the relationship was almost linear (Pearson $R^2 = 0.9971$) allowed us to convert LoH scores from
- 161 CytoScanHD microarrays to F_{roh} (Supplementary Materials). Comparison statistics were
- 162 performed using GraphPad (https://www.graphpad.com/).

163 **Copy Number Variant (CNV) analysis**

- 164 CNV analysis of the CytoScanHD array data was performed using ChAS and PennCNV; for
- 165 CoreExome data by the Illumina Genome Suite/CNVpartition and PennCNV. Evidence of
- 166 homozygous loss CNVs was cross-references with WES data, using the Integrated Genome
- 167 Viewer v2.3.5 (IGV: <u>https://software.broadinstitute.org/software/igv/</u>; (42), and CNVs
- 168 corroborated in this manner were then checked by PCR. Further
- analysis/prioritization/validation of CNVs, including using WES data, is given in Supplementary
- 170 Materials.

171 Cross-referencing with other datasets

- 172 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.
- 175 *De novo*, 2. Heterozygous LoF but status either de novo or unknown; 3. Biallelic (i.e.
- 176 homozygous); 4. Putative biallelic (possibly compound heterozygous, but phasing unknown); 5.
- 177 X-linked in males, maternally inherited. Data are shown in Supplementary File 1.

178	MSSNG: More than 13,000 individuals from the Autism Genetic Research Exchange (AGRE)
179	repository and other cohorts have been whole-genome sequenced, through Autism Speaks, and
180	data made available through the MSSNG database (https://research.mss.ng; db7 (43). MSSNG
181	data was accessed in May 2024.
182	SFARI: SSC biallelic: The Simons Simplex Collection (SSC) includes 2,600 simplex autism or
183	related developmental disorder families. WES data is available for ~2,500 of these families,
184	biallelic variants for our gene list were searched using the GPF browser
185	(https://gpf.sfari.org/gpf19/datasets/SSC/browser). Putative compound heterozygous variants
186	criteria for minor allele frequency (MAF) in SSC exome, gnomAD exome and genome
187	frequencies, of <0.001, or, for missense variants, MCP scores >1 and CADD scores >18. SVIP
188	biallelic: the autism Simons Variation in Individuals Project (SVIP) dataset was searched for
189	putative biallelic variants through the browser gpf.sfari.org/gpf19/datasets/SVIP/browser.
190	Putative compound heterozygous variants checked for criteria including minor allele frequency
191	(MAF) in SSC exome, gnomAD exome and genome frequencies, of <0.001, or, for missense
192	variants, MCP scores >1 and CADD scores >18.
193	Deciphering Developmental Disorders (DDD) study: includes sequence data for ~14,000
194	children; this dataset was searched for variants in genes overlapping with our study through the
195	known developmental genes list (https://decipher.sanger.ac.uk/ddd#ddgenes), as well as the
196	list of research variants (https://decipher.sanger.ac.uk/ddd#research-variants), which are 2,723
197	variants of unknown significance from 4,293 children. For <i>de novo</i> variants, all are constitutive,
198	unless specifically noted as mosaic here. CNV variants are not included here.

Biallelic variants for autism....

199	Autism Sequencing Consortium (ASC): genes with <i>de novo</i> variants, either in the control set or
200	case set, from ASC exome analysis browser (https://asc.broadinstitute.org/). This dataset lists
201	numbers of <i>de novo</i> protein-truncating variants, as well as <i>de novo</i> missense variants with MCP
202	scores either 1-2 or \geq 2. The ASC dataset includes 6,430 probands, and includes the SSC data.
203	DeRubeis: genes with de novo variants reported in the DeRubeis et al, 2014 study, with either
204	LoF or missense variants (44). As subject IDs were not reported, overlap with other studies
205	recorded here is not possible.
206	AutismKB: additional studies where variants have been reported for the genes are taken from
207	the Autism Knowledge Base (http://db.cbi.pku.edu.cn/autismkb_v2/quick_search.php), (45).
208	Neuroanatomical Enrichment Analysis
209	The Allen Adult Human Brain Atlas (46), Brainspan (47), and a single-cell atlas of the mouse
210	nervous system (48) were used to test for neuroanatomically specific expression. For each
211	gene, expression was standardized across regions or cell types. For these compartments, each
212	gene was then ranked from most specific to most depleted. The area under the receiver
213	operating statistic (AUC) was used to quantify specific expression for the genes of interest
214	within a region or cell cluster. The Mann-Whitney U test was used to test statistical significance,
215	and the Benjamini–Hochberg procedure (49) was used to correct for multiple tests. In addition,
216	in order to explore whether our candidate genes may contribute to electrophysiological and
217	morphological phenotypes in neurons we cross-referenced gene expression correlates with
218	electrophysiological phenotypes through a dataset published by Bomkamp et al, 2019 (50).

Biallelic variants for autism....

220 Results

- 221 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
- 224 (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,
- 226 providing data for CNV and ROH analyses.
- 227 Whole Exome Sequencing: We discovered 84 SNVs or indels through our analysis pipeline,
- including 49 homozygous autosomal variants, 20 autosomal de novo (and 1 inherited) variants,
- and 5 *de novo* and 9 maternally inherited X-linked variants.
- 230 Biallelic mutations (Table 2): We identified homozygous variants (7 LoF or pLoF; 9
- 231 nonsynonymous; 1 in-frame del) in 17 genes that have been previously reported to be
- associated with autosomal recessive non-syndromic ID, including BTN3A2 (30), CC2D1A (MIM
- 233 608443; MRT3; (51)), LINS1 (28), MADD (30), MTHFR (29), RSRC1 (52,53), TECPR2 (30) and
- 234 ZNF335 (27), or for syndromic or metabolic forms of autosomal recessive ID, including AGA
- 235 (aspartylglucosaminuria), ASL (arginosuccinic aciduria), ASPA (aspartoacylase
- 236 deficiency/Canavan disease), and HTRA2 (3-methylglutaconic aciduria). Three genes, CC2D1A,
- 237 DEAF1, and VPS13B have also previously been associated with ASD or autistic features (94–96;
- 238 (57)). Biallelic mutations in *DEAF1* are known to cause neurodevelopmental disorder with
- hypotonia, impaired expressive language, with or without seizures (MIM 617171). LoF variants
- 240 were reported for new candidate genes ENPEP, DAGLA, FAXDC2, GIMAP8, HRNR, ILDR2,

Biallelic variants for autism....

241	SLC36A1, SCN10A, VPS16, ZNF766, and pLoF (such as canonical splice site variants) in CNPY4,
242	DENND1B, LRRC34, PKD1L1, and PPP1R36. VPS16 has previously been reported for autosomal
243	dominant dystonia (MIM 608559), and PKD1L1 for AR visceral heterotaxy 8 (MIM 617205).
244	Heterozygous missense and LoF mutations in SHH are known to be involved in autosomal
245	dominant holoprosencephaly 3 (HPE3; MIM 142945), and so it is somewhat surprising to find a
246	homozygous missense variant associated with a milder phenotype and with no obvious HPE3-
247	related dysmorphic features. The Asn69Ile variant in SHH substitutes an asparagine for
248	isoleucine at a residue that is conserved across vertebrates, is predicted to be damaging, and is
249	found in heterozygous form in only 21 out of 1.6E6 gnomAD (v4.1.0) alleles, and no
250	homozygotes.
251	In one trio (trio SMPA21) we report a homozygous stop-gain variant in DAGLA, for which
251 252	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 neurodevelopmental disorders,
251 252 253	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene
251 252 253 254	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; <u>https://gene.sfari.org/</u>). We also report homozygous missense variants in novel
251 252 253 254 255	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; <u>https://gene.sfari.org/</u>). We also report homozygous missense variants in novel candidate genes <i>ANO10</i> , <i>CABP2</i> , <i>CAPSL</i> , <i>CLCA4</i> , <i>CPM</i> , <i>DNAH8</i> , <i>ERMP1</i> , <i>KSR2</i> , <i>RANBP9</i> , <i>RASAL2</i> ,
251 252 253 254 255 256	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; <u>https://gene.sfari.org/</u>). We also report homozygous missense variants in novel candidate genes <i>ANO10</i> , <i>CABP2</i> , <i>CAPSL</i> , <i>CLCA4</i> , <i>CPM</i> , <i>DNAH8</i> , <i>ERMP1</i> , <i>KSR2</i> , <i>RANBP9</i> , <i>RASAL2</i> , <i>TMEM25</i> , <i>TRIM3</i> , <i>VPS16</i> , and <i>WDR90</i> , plus in <i>SGSM3</i> and <i>EPHB1</i> , for which <i>de novo</i> variants in
251 252 253 254 255 256 257	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). 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 <i>ANO10</i> , <i>CABP2</i> , <i>CAPSL</i> , <i>CLCA4</i> , <i>CPM</i> , <i>DNAH8</i> , <i>ERMP1</i> , <i>KSR2</i> , <i>RANBP9</i> , <i>RASAL2</i> , <i>TMEM25</i> , <i>TRIM3</i> , <i>VPS16</i> , and <i>WDR90</i> , plus in <i>SGSM3</i> and <i>EPHB1</i> , for which <i>de novo</i> variants in have been reported for ASD (59), and are listed respectively as strong candidate and suggestive
251 252 253 254 255 256 257 258	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; <u>https://gene.sfari.org/</u>). We also report homozygous missense variants in novel candidate genes <i>ANO10</i> , <i>CABP2</i> , <i>CAPSL</i> , <i>CLCA4</i> , <i>CPM</i> , <i>DNAH8</i> , <i>ERMP1</i> , <i>KSR2</i> , <i>RANBP9</i> , <i>RASAL2</i> , <i>TMEM25</i> , <i>TRIM3</i> , <i>VPS16</i> , and <i>WDR90</i> , plus in <i>SGSM3</i> and <i>EPHB1</i> , for which <i>de novo</i> variants in have been reported for ASD (59), and are listed respectively as strong candidate and suggestive evidence in SFARI (scores of 2 and 3, respectively).
251 252 253 254 255 256 257 258 259	In one trio (trio SMPA21) we report a homozygous stop-gain variant in <i>DAGLA</i> , for which rare heterozygous variants were reported to be associated with neurodevelopmental disorders, including autism (58). SFARI also lists this as a strong candidate for a ASD disease gene (gene score of 2; <u>https://gene.sfari.org/</u>). We also report homozygous missense variants in novel candidate genes <i>ANO10</i> , <i>CABP2</i> , <i>CAPSL</i> , <i>CLCA4</i> , <i>CPM</i> , <i>DNAH8</i> , <i>ERMP1</i> , <i>KSR2</i> , <i>RANBP9</i> , <i>RASAL2</i> , <i>TMEM25</i> , <i>TRIM3</i> , <i>VPS16</i> , and <i>WDR90</i> , plus in <i>SGSM3</i> and <i>EPHB1</i> , for which <i>de novo</i> variants in have been reported for ASD (59), and are listed respectively as strong candidate and suggestive evidence in SFARI (scores of 2 and 3, respectively). Although not focused on consanguineous populations, WGS data from the MSSNG

261 genes (see Supplementary File 1), with homozygous rare variants in DNAH7 (MSSNG), DNAH8

Biallelic variants for autism....

262	(MSSNG), CAPSL (MSSNG), and CYP2A7 (MSSNG), and putative compound heterozygous
263	damaging rare variants in DNAH8 (in three MSSNG individuals), SCN10A (DDD), and WDR90
264	(DDD)(see Supplementary File 1). There was also evidence of disease-causing biallelic variants
265	in known AR neurodevelopmental genes MTHFR (three homozygous; DDD), AGA (one
266	homozygous stop gain; DDD), DEAF1 (two homozygous stop gain; MSSNG and DDD), ZNF335
267	(one homozygous missense, DDD; one putative compound heterozygous, DDD), ASL (one
268	homozygous missense; MSSNG), CC2D1A (one homozygous missense, MSSNG; two
269	homozygous LoF DDD), and VPS13B (six homozygous, nine putative compound heterozygous).
270	De novo mutations (Table 3 & 4): De novo dominant mutations have been shown to be a major
271	causal factor in ASD etiology. We identified 25 <i>de novo</i> variants (20 autosomal and 5 X-linked).
272	Among these genes, variants in four of them have been identified in previous studies (MECP2,
273	MYT1L, SCN2A, ZNF292). 14 of the 25 variants were LoF, three were putative LoF (splice site),
274	and the remainder were nonsynonymous. The mutations include genes that have been
275	implicated in other recent studies of ASD and/or ID. For instance, we report a <i>de novo</i>
276	hemizygous nonsynonymous mutation in CLTRN (on chromosome X). Hemizygous mutations in
277	CLTRN have recently been reported in several males with neutral aminoaciduria accompanied
278	with autistic features, anxiety, depression, compulsions and motor tics (60)- a clinical picture
279	reminiscent of the autosomal recessive Hartnup disorder (MIM 234500). There were also
280	several other genes with <i>de novo</i> variants that have not been associated with ASD (or ID)
281	previously, and may represent novel putative genes for ASD, and for other trios there are

Biallelic variants for autism....

variants in genes linked with other nervous system disorders, such as *DRP2* in Charcot-MarieTooth disease (61).

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 (62).

288 Cross-referencing with other datasets, there was also support for new candidate

dominant/de novo autosomal genes, including ECM1, SLAMF7, NCL, FAM53C, ADGRF2,

290 TANGO2, DGKZ, ATP2B1 (5 individuals), CBFA2T3, RETN, and PPIL2 (see Supplementary File 1).

291 There was support for X-linked recessive (in males, maternally inherited or *de novo*) candidate

292 gene NRK (three individuals), MAGEB2 (three individuals), MAGEC1 (11 individuals), MSN (five

individuals), and *ZNF185* (one individuals).

294

295 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).

297 Splice site variants

298 In addition to the detection of variants at canonical splice donor and splice acceptor sites

299 (Tables 2-4), use of SpliceAI suggests that, for the biallelic nonsynonymous variant in the known

ASD/ID gene ASL (trio SA7), an alternative explanation could be the generation of an alternative

301 splice donor by the cytosine to thymine transition at this site. However, other splice algorithms

Biallelic variants for autism....

302 such as https://www.fruitfly.org/seq tools/splice.html do not support this, and molecular

- 303 experimental procedures may be needed to corroborate or refute this prediction.
- 304 Copy Number Variants (Table 5 & 6)

305 Eight autosomal biallelic loss CNVs were identified among five trios using microarray data, that

306 were subsequently validated through inspection of the WES data using IGV, and through

- 307 molecular methods, using PCR and/or Sanger sequencing. Of these implicated genes,
- 308 mutations/knock-outs of both DHRS4 and KLK15 have a neurobehavioral phenotype in mouse
- 309 models (Supplementary File 1). DNAH7 is highly expressed in brain, and both SHPK
- 310 (sedoheptulokinase) and WDR73 have high expression in cerebellum. Of the biallelic CNVs
- identified, WDR73 has previously been associated with the autosomal recessive Galloway-
- 312 Mowat syndrome 1 (MIM 251300), which includes microcephaly, delayed psychomotor
- 313 development and cerebellar atrophy. Determining the ASD causative genetic variation in many

of the probands with CNVs is difficult. The proband with the *WRD73, CYP2A7*, and *KLK15* CNVs

315 (trio IABB14) also has a biallelic splice mutation in the Canavan disease gene ASPA. Additionally,

- the proband with the SHPK CNV (trio IABB2) also has a biallelic LoF mutation in the known ID
- 317 gene *DEAF1*, and the proband with the *SIRPB1* CNV (trio IAU79) has a biallelic non-frameshift

deletion in the known ID/spastic paraplegia gene *B4GALNT1*, complicating identification of the

319 molecular causation (see Supplementary File 1). A biallelic CNV loss implicates the gene SEMG1

- 320 in trio SMPA4, however expression of this gene is restricted to the seminal vesicle (see
- 321 Supplementary File 1), and thus unlikely to be related to the ASD phenotype.

Biallelic variants for autism....

- 322 There are also three large, multi-genic *de novo* CNV losses among the trios, which may
- also be pathogenic (Supplementary Table S1).
- 324 Neuroanatomical Enrichment Analysis
- 325 Testing for regional and cell-type specific expression did not indicate clear anatomical
- 326 targets with higher expression of the candidate genes. Enrichment in the cerebellum and
- 327 cerebral cortex was observed in the BrainSpan human developmental atlas, but this was not
- 328 supported by either the mouse or adult human atlases. We found no consistent
- 329 neuroanatomical expression pattern for the identified genes, suggesting heterogeneity of
- 330 neural circuits disrupted.

331 High consanguinity coefficient correlates with identification of biallelic variants

- 332 Using microarray data we calculated the consanguinity coefficient F_{RoH} using autosomal RoH for
- 333 104 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).

335 Comparison of the degree of consanguinity (using F_{ROH}) between 33 trios with biallelic variants

versus 20 trios with *de novo* and X-linked variants indicated a strongly significant shift (unpaired

- 337 *t* test (two-tailed): *p* < 0.0001; df=51; mean F_{RoH} for biallelic variants= 0.09003, S.E.M.=0.00672,
- 338 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 biallelic
- 340 0.08146, N=13, S.E.M.=0.01019; mean F_{RoH} for autosomal *de novo* and X-linked = 0.03282,
- 341 N=11, S.E.M.=0.01159).(Figure 2A). Comparison also showed clear differences in F_{ROH} levels
- 342 between different cohorts, and country of ascertainment (Figure 2B & C).

Biallelic variants for autism....

343

344 Discussion

345	Many recent studies involving NGS in ASD have involved large cohorts, focusing predominantly
346	on dominant/ <i>de novo</i> inheritance. This focus is largely driven by the fact that autosomal
347	recessive variants in novel genes are more difficult to identify in the outbred population.
348	Identification of candidate AR ASD genes through the study of consanguineous families should
349	help streamline the identification of biallelic mutations through clinical genetic screening in
350	outbred populations. Using whole exome trio analysis in consanguineous families, we have
351	enriched for recessive variants to assess the role these variants play in ASD in populations
352	where endogamy is common. This study of 115 trios has identified WES variants for 32 trios in
353	genes previously associated with ASD or other neurodevelopmental disorders, resulting in a
354	diagnostic yield of 28% (18 autosomal recessive, seven autosomal de novo plus one maternally
355	inherited, and six X-linked variants in known genes). With a further three trios with large, multi-
356	genic, loss CNVs validated experimentally as <i>de novo</i> and putatively pathogenic (Table 6), this
357	yield increases to 30%. There were many other variants identified that met filtering criteria
358	which could potentially represent novel ASD genes or targets. We present 32 new candidate
359	autosomal recessive variants/genes, including a biallelic variant in SHH, a gene associated with
360	autosomal dominant holoprosencephaly. Within this cohort we identified variants in five genes
361	that are associated with known metabolic syndromes (4.3% of the diagnosis cohort): AGA, ASL,
362	HTRA2, ASPA, and MTHFR. These genes may represent better clinical management
363	opportunities for patients, and potentially better therapies.

Biallelic variants for autism....

364	A number of the known ID genes identified among the 115 ASD trios have also previously
365	been reported for ASD or ASD-like features, e.g. CC2D1A (55,62,63), VPS13B (33), DEAF1 (54),
366	ZNF335 (64), ZNF292 (65,66), MYT1L (66,67), and SCN2A (59,68,69). Of the candidate genes
367	identified here, MSSNG, SFARI, and other datasets provide putative support for the biallelic
368	genes, with homozygous variants in DNAH7, ANO10, DNAH8, CAPSL, CYP2A7, and SIRPB1, and
369	putative compound heterozygous variants in DNAH7, DNAH8, SCN10A, and WDR90, and for de
370	novo/dominant genes NCL, SLAMF7, ADGRF2, DGKZ, ATP2B1, CBFA2T3, RETN, and PPIL2
371	(Supplementary File 1).
372	For three of the known ASD or ID genes reported here, namely ZNF292, SCN2A, and MECP2,
373	mutations have been identified in two unrelated trios from our cohort. Mutations in the known
374	ASD gene SCN2A (68) were identified in two trios- both <i>de novo</i> . In addition, mutations were
375	identified in three other voltage-gated sodium channel members, including a biallelic stop gain
376	mutation in SCN10A, in which heterozygous gain of function missense mutations have
377	previously been linked to familial episodic pain syndrome 2 (FEPS2; MIM 615551)(70), as well as
378	de novo missense mutations in SCN5A and SCN8A. The SCN8A variant is predicted as damaging by
379	all algorithms tested, and mutations in this gene have previously been linked to AD developmental and
380	epileptic encephalopathy 13 (DEE13; MIM 614558). The SCN5A variant was predicted as benign by
381	several of the algorithms and is less likely to be related to ASD in this trio. The Arg267* mutation
382	identified in MECP2 in trio SMPA78 has been frequently reported in cases of Rett syndrome. However
383	more C-terminal truncating mutations such as the Thr412Asnfs*5 mutation in trio IAU54 typically have a
384	significantly milder phenotype(71).

385	For some of the autosomal recessive candidate genes, there is additional support from
386	animal models. For instance, Rasal2 (MGI:2443881) and Vps16 (MGI:2136772) knockout mice
387	have behavioral/neurological and nervous system phenotypes
388	(http://www.informatics.jax.org/; Supplementary File 1). Additional discussion of mouse
389	models relevant to the candidate genes is provided in Supplementary Materials.
390	There are several trios for which there are two or more candidate variants (Supplementary
391	File 1). For some of these, there may be a variant that clearly delivers a more plausible
392	etiopathological explanation. For instance, for trio SMPA3 the biallelic change in MTHFR; for
393	trio IAH2, the biallelic LoF change in TECPR2 (SPG49; MIM 615031; ID (30); hereditary sensory
394	neuropathy with ID (72)); for trio IABB2, the biallelic LoF variant in (NEDHELS; MIM 617171); for
395	trio IABB14, the biallelic splice mutation in ASPA (Canavan disease; MIM 271900).
396	Since our initial methodology used relatively stringent MAF cut-off criteria for biallelic
397	variants (<0.001), we attempted to justify this level by reiterating the analysis but with lower
398	thresholds (<0.01). Using the more relaxed criteria resulted in just two additional candidate
399	variants, in the genes CLCA4 (gnomAD South Asian MAF=1.72E-3), and RASAL2 (gnomAD South
400	Asian MAF=6E-3), but none in known neurodevelopmental genes. Our methodology was
401	particularly stringent regarding de novo nonsynonymous variants, owing to the large number of
402	likely spurious calls, and the inclusion of only variants predicted as damaging by all algorithms
403	used may have excluded some bona fide mutations. Other variants, particularly intronic or
404	intergenic, would also likely be missed by our approach. Also, a proportion of exons, and
405	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 408 409 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 410 Isles (PABI), which is most like our cohort (38). The PABI cohort had 356 (333 undiagnosed) 411 probands, 110 of whom (30.9%) had biallelic coding candidate variants, approximately half of 412 which are in known DD genes and half in novel candidates (38). For comparison, our study 413 414 reports very similar yields for recessive variants, with 15% of probands with biallelic variants in known ASD/ID genes (17/115), or 30% including those with biallelic variants in novel candidate 415 genes (35/115). We report 28 de novo variants (20 autosomal, five X-linked, three CNVs) in 17 416 of 115 individuals (15%), which is somewhat lower than those reported for the PABI cohort 417 (29.8%). This difference is most likely due to differences in the methodologies or stringency 418 419 used for calling *de novo* variants in the respective studies.
- Comparison of consanguinity (F_{ROH}) between trios with biallelic 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 biallelic rare variants is strongly associated with
 higher consanguinity, and should be a consideration for diagnostic genetic analysis.
- 424 Overall, the findings demonstrate the importance of autosomal recessive mutations in
 425 ASD in countries, or in cohorts, with high rates of consanguinity. In countries where
 426 consanguinity is rare, when screening ASD individuals' genomes attention should still be paid to

427	the relatedness of the parents, and, for those with a high F-coefficient, attention should be
428	focused on biallelic variants. Additionally, the candidate genes reported here should be
429	examined for both autozygous and allozygous mutations in outbred populations, just as would
430	be done for more established autosomal recessive gene disorders.
431	
432	Declaration of Interests Statement
433	L.F. owns shares in Quince Therapeutics and has received consulting fees from PeopleBio Co., GC
434	Therapeutics Inc., Cortexyme Inc., and Keystone Bio. M.A. is co-founder and director of Institute of
435	Omics and Health Research, Lahore, a research organization. The organization was not involved
436	in the research on which this manuscript is based. All other authors declare no competing
437	interests.
438	
439	
440	Acknowledgements
441	The authors thank the families for their participation in this study. We acknowledge the
442	assistance of Natalie Freeman with PLINK microarray analysis, and of Stephen Pastore with the
443	preparation of violin plots using R software. This work was supported by grants from the

- 444 Canadian Institutes of Health Research to JBV (#MOP-102758 and #PJT-156402), also a Brain &
- 445 Behavior Research Foundation (BBRF/NARSAD) Young Investigator Award to NV. RH was
- 446 supported by a Peterborough K.M. Hunter Charitable Foundation Graduate Scholarship. AR and
- 447 AH were supported by International Research Fellowship Program scholarships from the

Biallelic variants for autism....

448	Pakistani Higher Education	Commission. C	S was supported by a	Summer University Research
-----	----------------------------	---------------	----------------------	----------------------------

- 449 Program award from the Institute of Medical Science, University of Toronto. Computations
- 450 were performed on the CAMH Specialized Computing Cluster. The SCC is funded by The Canada
- 451 Foundation for Innovation, Research Hospital Fund. Additional acknowledgements are included
- 452 in Supplementary Materials

453 Data and code availability

- 454 New variants reported here are available through ClinVar SUB14665852, and SCV accession
- 455 numbers are included in Supplementary File 1. Microarray genotype data and whole exome sequence
- 456 data will be made available from the corresponding author upon request.
- 457

458 References

- 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.
- 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.
- 464 3. Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring
 465 mental health diagnoses in the autism population: a systematic review and meta-analysis.
 466 Lancet Psychiatry. 2019 Oct;6(10):819–29.
- 467 4. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and
 468 global burden of autism spectrum disorders. Psychol Med. 2015 Feb;45(3):601–13.
- Elsabbagh M, Divan G, Koh YJ, 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.
- 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.

475 476 477 478 479	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.
480 481	8.	Kim YS, Leventhal BL, Koh YJ, Fombonne E, Laska E, Lim EC, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry. 2011 Sep;168(9):904–12.
482 483 484	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.
485 486 487 488	10.	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
489 490 491 492	11.	Maenner MJ. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. MMWR Surveill Summ [Internet]. 2023 [cited 2024 Sep 16];72. Available from: https://www.cdc.gov/mmwr/volumes/72/ss/ss7202a1.htm
493 494	12.	Fombonne E. Editorial: The rising prevalence of autism. J Child Psychol Psychiatry. 2018;59(7):717–20.
495 496 497	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.
498 499	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.
500 501 502	15.	Rødgaard EM, Jensen K, Vergnes JN, 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.
503 504 505	16.	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.
506 507 508	17.	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.
509 510	18.	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

- 511 19. Tareen A, Mirza I, Minhas A, Minhas F, Rahman A. Developing a child and adolescent
 512 mental health service in a low-income country: a global partnership model. Psychiatr Bull.
 513 2009 May;33(5):181–3.
- 514 26. Sarwat A, Ali SMI, Ejaz MS. Mental health comorbidity in children: a hospital based study
 515 in child psychiatry clinic. Pak J Med Sci. 2009;25(6):982–5.
- 516 21. Minhas A, Vajaratkar V, Divan G, Hamdani SU, Leadbitter K, Taylor C, et al. Parents'
 517 perspectives on care of children with autistic spectrum disorder in South Asia Views from
 518 Pakistan and India. Int Rev Psychiatry. 2015 May 4;27(3):247–56.
- 22. 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.
- 522 23. Bundey S, Alam H. A five-year prospective study of the health of children in different ethnic
 523 groups, with particular reference to the effect of inbreeding. Eur J Hum Genet EJHG.
 524 1993;1(3):206–19.
- 525 24. Modell B, Darr A. Science and society: genetic counselling and customary consanguineous
 526 marriage. Nat Rev Genet. 2002 Mar;3(3):225–9.
- 527 25. Harripaul R, Vasli N, Mikhailov A, Rafiq MA, Mittal K, Windpassinger C, et al. Mapping
 528 autosomal recessive intellectual disability: combined microarray and exome sequencing
 529 identifies 26 novel candidate genes in 192 consanguineous families. Mol Psychiatry.
 530 2018;23(4):973–84.
- 26. 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.
- 534 27. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, et al. Genetics of intellectual
 535 disability in consanguineous families. Mol Psychiatry. 2018 Jan 4;
- 28. 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.
- 29. 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.
- 30. 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.
- 31. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. Science. 2008 Jul 11;321(5886):218–23.

- 32. 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.
- 549 33. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al.
 550 Using whole-exome sequencing to identify inherited causes of autism. Neuron. 2013 Jan 23;77(2):259–73.
- 34. 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.
- 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.
- 36. 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.
- 37. Kaiser VB, Svinti V, Prendergast JG, Chau YY, 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.
- 38. 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.
- 39. 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.
- 40. Ramaswami G, Geschwind DH. Genetics of autism spectrum disorder. Handb Clin Neurol.
 2018;147:321–9.
- 41. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et
 al. Runs of homozygosity in European populations. Am J Hum Genet. 2008 Sep;83(3):359–
 72.
- 42. 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.
- 43. 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.
- 44. 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.

- 45. 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.
- 46. 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.
- 47. 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).
- 48. 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.
- 49. 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.
- 50. 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.
- 51. 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 nonsyndromic mental retardation. J Med Genet. 2006 Mar;43(3):203–10.
- 52. 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.
- 53. 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.
- 54. 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.
- 55. Manzini MC, Xiong L, Shaheen R, Tambunan DE, Di Costanzo S, Mitisalis V, et al.
 614 CC2D1A regulates human intellectual and social function as well as NF-κB signaling
 615 homeostasis. Cell Rep. 2014 Aug 7;8(3):647–55.
- 56. Douzgou S, Petersen MB. Clinical variability of genetic isolates of Cohen syndrome. Clin
 Genet. 2011 Jun;79(6):501–6.

- 57. 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.
- 58. Smith DR, Stanley CM, Foss T, Boles RG, McKernan K. Rare genetic variants in the
 endocannabinoid system genes CNR1 and DAGLA are associated with neurological
 phenotypes in humans. PloS One. 2017;12(11):e0187926.
- 59. 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.
- 60. 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.
- 61. 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.
- 633 62. Mirzaa GM, Chong JX, Piton A, Popp B, Foss K, Guo H, et al. De novo and inherited
 634 variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum
 635 disorder. Genet Med. 2020 Mar;22(3):538–46.
- 636 63. Sener EF, Onal MG, Dal F, Nalbantoglu U, Ozkul Y, Canatan H, et al. Novel alterations of
 637 CC2D1A as a candidate gene in a Turkish sample of patients with autism spectrum disorder.
 638 Int J Neurosci. 2020 Dec 28;1–8.
- 64. 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.
- 642 65. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo
 643 mutations in autism/developmental delay risk genes suggest a multifactorial model. Mol
 644 Autism. 2018;9:64.
- 66. 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.
- 647 67. Coursimault J, Guerrot AM, Morrow MM, Schramm C, Zamora FM, Shanmugham A, et al.
 648 MYT1L-associated neurodevelopmental disorder: description of 40 new cases and literature
 649 review of clinical and molecular aspects. Hum Genet. 2021 Nov 8;
- 68. 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.

- 653 69. Jiang Y hui, Yuen RKC, Jin X, Wang M, Chen N, Wu X, et al. Detection of clinically
 654 relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J
 655 Hum Genet. 2013 Aug 8;93(2):249–63.
- Faber CG, Lauria G, Merkies ISJ, Cheng X, Han C, Ahn HS, 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.
- 658 71. Bebbington A, Percy A, Christodoulou J, Ravine D, Ho G, Jacoby P, et al. Updating the
 659 profile of C-terminal MECP2 deletions in Rett syndrome. J Med Genet. 2010
 660 Apr;47(4):242–8.
- 72. 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.
- 564 73. 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.
- 667 74. 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.
- 670 75. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Autism diagnostic
 671 observation schedule: a standardized observation of communicative and social behavior. J
 672 Autism Dev Disord. 1989 Jun;19(2):185–212.
- 673 76. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders,
 674 4th Edition, Text Revision. 4th edition. Washington, DC: American Psychiatric Association;
 675 2000. 943 p.
- 576 77. Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. The
 677 developmental, dimensional and diagnostic interview (3di): a novel computerized assessment
 678 for autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2004 May;43(5):548–
 679 58.
- 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
- 683 79. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders
 684 [Internet]. World Health Organization; 2010. Available from:
 685 https://www.who.int/classifications/icd/en/bluebook.pdf
- 80. 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.
- 689

Biallelic variants for autism....

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)(73) 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 (74) and ADOS (75) (DSM-IV-TR) (76); SCQ (50)	Certified Physicians
IABB (Dr. Bita Bozorgmehr)	Shahid Beheshti Medical University, Teheran	Iran	10 (30)	9:2	Autism Spectrum Screening Questionnaire (ref) DSM IV	Dr. Bita Bozorgmehr
SA (Laila Al Ayadhi)	Autism Treatment Centre, King Saud University, Riyadh	Saudi Arabia (2-12)	13 (39)	12:1	ADI-R (74), Autism Diagnostic Observation Schedule (ADOS) (75) and 3DI (Developmental, dimensional diagnostic interview) scales (77) Childhood Autism Rating Scale (CARS)(73); DSM IV	Laila Al Ayadhi
Autism (Raheel Qamar, Maleeha Azam and Zehra Agha)	COMSATS University, Islamabad, and National Rural Support Program	Pakistan	4 (12)	4:4	Childhood Autism Rating Scale (CARS)(73)	Who?

RQPA (Raheel Qamar, Maleeha Azam and Zehra Agha)	and the Ghazi Barotha Taraqiata iDara		4 (12		DSM-5 (78) and ICD-10 criteria (79)	Who?
IAH (Abolfazl Heidari)	Qazvin Medical University, Qazvin	Iran	3 (9)	2:1	DSM IV	Dr Hossein Mojdehipanah
Total			115 (345)	91:24		

- 694 **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: <u>https://omim.org/</u>; SFARI ASD genes <u>https://gene.sfari.org/</u>; DDD gene2phenotype
- 696 genes: <u>https://www.ebi.ac.uk/gene2phenotype/</u>; Geisinger developmental brain disorder gene database:

697 <u>https://dbd.geisingeradmi.org/</u>; NP*denovo*: <u>http://www.wzgenomics.cn/NPdenovo/</u>; Gene4denovo:

- 698 <u>http://www.genemed.tech/gene4denovo/search</u>. Genes previously unreported for ASD or ID are in bold type. Control population
- 699 frequencies show are from gnomAD (all, and for South Asian sub-population). All variants were validated by IGV (Integrated
- 700 Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in Supplementary Materials. Allele frequencies
- are from gnomAD v2.1.1 (www. <u>https://gnomad.broadinstitute.org/</u>; accessed June 2022).

Variant	Family	Chi	^r Coordinate	Ref.	Alt.	Gene	Variant Type	cDNA/Amino Acid Change	gnomAD A	llgnomAD S.
			s (hg19)						Freq.	Asian Freq.
1	SMPA3	1	11855404	с	A	MTHFR (29)(MIM 236250; SFARI score 2)	nonsynonymous	NM_001330358.2: exon6: c. 905G>T: p. Gly302Val	3.99E-06	3.27E-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	GACGTTTC TG	-	HRNR	frameshift del	NM_001009931.3: exon3: c.423_432del:p.Arg142Glufs*15	4.37E-05	3.59E-4
4	SMPA76	1	166891855- 166891856	СТ	-	ILDR2	frameshift del	NM_199351.3: exon8: c.1187_1188del: p. Glu396Glyfs*39	3.98E-06	3.27E-05
5	SMPA3	1	178425880	С	A	RASAL2	Nonsynonymous	NM_004841.3: exon11: c.1813C>A: p. Leu605lle	6.72E-4	6.04E-3
6	SMPA40	1	197621364	С	A	DENND1B	Splicing	NM_001195215.1: exon7: c.447+1G>T	0	0
7	Autism8	2	74759786	G	Т	HTRA2 (MIM 617248)	Nonsynonymous	NM_013247.3: exon7: c.1156G>T: p. Asp386Tyr	7.95E-06	6.53E-05
8	SMPA19	3	38802185	С	A	SCN10A	Stopgain	NM_001293306.2: exon7: c.937G>T; p. Gly313*	3.98E-06	3.27E-05
9	SMPA76	3	43593070	С	A	ANO10 (spinocerebellar ataxia 10, MIM 613728)	Nonsynonymous	NM_001346463.2:exon11:c.1679G>T:p.Cys560Phe	0	0
10	SA15	3	47164429	G	с	SETD2 (LLS (AD) MIM 616831; SFARI score 1; Geisinger Tier 1)	Nonsynonymous	NM_001349370.3: exon2:c.1565C>G: p. Ser522Cys	0	0

11	SMPA27	3	134911657	Т	С	EPHB1 (SFARI score 3)	Nonsynonymous	NM_004441.3: exon11: c.2122T>C: p. Phe708Lys	4.01E-06	3.27E-05
12	Autism3	3	157840002	с	Т	RSRC1 (MRT70; MIM 618402; SFARI score S)	Stopgain	NM_001271834.2: exon2:c.109C>T: p. Arg37*	0	0
13	SMPA8	3	169524635	С	Т	LRRC34	Splicing	NM_001172779.1: exon5: c.528+1G>A	4.08E-06	03.44E-05
14	SMPA40	4	111480831	С	Т	ENPEP	Stop gain	NM_001977.4: exon19: c.2671C>T; p.Arg891*	1.791E-4	1.34E-03
15	RQPA12	4	178360822	G	A	AGA (MIM 208400)	Nonsynonymous	NM_000027.3: exon3: c.302C>T: p. Ala101Val	3.98E-06	3.27E-05
16	SMPA35	4	185033930	САТ	С	ENPP6	frameshift del	NM_153343.3: exon6: c.886_887del: p.Met296Aspfs*28	0	0
17	IABB11	5	35910583	С	Т	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.487_490dup: p. Asp164Glyfs*2	8.01E-06	6.6E-05
19	IAU56	5	154199963	С	-	FAXDC2	frameshift del	NM_032385.3: exon9: c.915delG: p.Thr306Profs*100	5.28E-05	3.93E-04
20	SMPA48	6	13634723	С	G	RANBP9	Nonsynonymous	NM_005493.3: exon11: c.1735G>C: p. Gly579Arg	2.89E-05	1.06E-04
21	IABB8	6	38851674	A	Т	DNAH8	Nonsynonymous	NM_001206927.2: exon56: c.8159A>T: p. Asp2720Val	1.24E-05	1.02E-04
22	IAU7	7	47944924	Т	с	PKD1L1 (AR laterality defects; MIM 616205)	Splicing	NM_138295.3: exon11: c.1523-2A>G	0	0
23	SA7	7	65551627	С	Т	ASL (MIM 207900)#	Nonsynonymous	NM_001024943.2: exon6:c.502C>T: p. Arg168Cys	7.98E-06	0
24	IABB7	7	99720009	G	С	CNPY4	Splicing	NM_152755.1: exon2: c.245+1G>C	2.12E-05	0
25	IAU79	7	150174821	тсс	Т	GIMAP8	frameshift del	NM_175571.3: exon5: c.1952_1953del: p. Gln652Serfs*6	0	0
26	SMPA11	7	155604611	Т	A	SHH (HPE3 (AD) MIM 142945)	Nonsynonymous	NM_000193.4: exon1: c.206A>T: p. Asn69lle	2.39E-06	1.96E-4
27	IABB3	8	100155379	AG	A	VPS13B (COH1 MIM 216550; SFARI score 1; Geisinger AR)	frameshift del	NM_015243.3: exon13: c.1829delG: p. Ser610Thrfs*3	0	0
28	SA4	9	5805613	т	С	ERMP1	Nonsynonymous	NM_024896.3: exon9: c.1721A>G: p.His574Arg	0	0
29	IABB2	11	679701	G	GA	DEAF1 (MIM 617171; SFARI score 1; Geisinger AR)	frameshift ins	NM_001293634.1: exon6: c.845dupT: p. Ala283Argfs*20	0	0

30	SMPA11	11	6477889	с	A	ТКІМЗ	Nonsynonymous	NM_033278.3:exon6:c.1067G>T:p.Arg356Leu	0	0
31	SMPA20	11	47305749	A	G	MADD (27,30)	Nonsynonymous	NM_001135943.2: exon11: c. 1883A>G: p. Tyr628Cys	1.99E-05	1.63E-04
32	SMPA21	11	61495635	с	G	DAGLA (SFARI score 2)	Stopgain	NM_006133.3: exon7:c.647C>G: p. Ser216*	0	0
33	SMPA80	11	67290028	с	Т	CABP2 (Deafness, MIM 607314)	Nonsynonymous	NM_016366.3: exon2: c.202G>A: p. Ala68Thr	5.70E-05	3.48E-04
34	SMPA8	11	118403863	A	G	TMEM25	Nonsynonymous	NM_001144036.2: exon3: c. 302A>G: p. Asn101Ser	3.64E-05	2.62E-04
35	IAU79	12	58021514_ 58021534	GGGAAGCCGAC GAGCTCGTGGT	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
36	IABB10	12	69263140	с	G	СРМ	Nonsynonymous	NM_001005502.2: exon6: c.752G>C: p. Gly251Ala	0	0
37	IABB11	12	118199039	G	A	KSR2	Nonsynonymous	NM_173598.6:exon4:c.763C>T:p.Arg255Trp	2.196E-05	0
38	SMPA8	14	65032074	G	С	PPP1R36	Splicing	NM_172365: exon5: c.270-1G>C	1.21E-05	9.99E-05
39	IAH2	14	102963429	TGTGGGGACCGC CTG	Т	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
40	IABB4	15	101113962	Т	-	LINS1 (MRT27; MIM 614340; Geisinger AR)	frameshift del	NM_001040616.3: exon5: c.1116delA: p. Glu372Aspfs*9	0	0
41	SMPA76	16	716004	с	Т	WDR90	Nonsynonymous	NM_145294.4: exon36: c.4489C>T: p. Arg1497Trp	3.72E-06	8.04E-06
42	IABB14	17	3392637	G	Т	ASPA (Canavan disease; MIM 271900)	Splicing	NM_001128085: exon5: c.634+1G>T	1.06E-05	3.28E-05
43	SA2	19	14024452	G	т	CC2D1A (MRT3; MIM 608443; SFARI score 2; Geisinger AR)	Splicing	NM_017721: exon6: c.748+1G>T	0	0
44	IABB16	19	50321659	G	A	MED25 (BVFYS; MIM 616449)	Nonsynonymous	NM_030973.3: exon1: c.61G>A: p. Val21Met	0	0

45	SA11	19	58265432	С	-	ZNF766	Frameshift del	NM_173632.4:c.934delC: p.Arg312Glufs*80	3.266E-5	1.55E-5
46	IABB16	20	2847201	с	A	VPS16 (Dystonia (AD) MIM 619291)	Nonsynonymous	NM_080413.3: exon20: c.2069C>A: p. Ala690Glu	1.99E-05	6.53E-05
47	IABB16	20	16360737	A	G	KIF16B (PMID 29736960)	Nonsynonymous	NM_001199865.2: exon19: c. T1910T>C: p. Val637Ala	0	0
48	IABB5	20	44578999	с	Т	ZNF335 Hu et al, 2019; MCPH10; MIM 615095; Geisinger Tier 4)	Nonsynonymous	NM_022095.3: exon22: c.3346G>A: p. Gly1116Arg	1.26E-5	0
49	IABB11	22	40803252	G	С	SGSM3 (SFARI score 2; Geisinger Tier 4)	Nonsynonymous	NM_001350039.2:c.1288G>C:p.Glu430Gln	1.99E-5	1.63E-4

702 # SpliceAI predicts the generation of a splice donor site.

- 704 **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: <u>https://omim.org/</u>; SFARI ASD genes <u>https://gene.sfari.org/</u>; DDD gene2phenotype genes:
- 706 <u>https://www.ebi.ac.uk/gene2phenotype/</u>; Geisinger developmental brain disorder gene database: <u>https://dbd.geisingeradmi.org/</u>;
- 707 NP*denovo*: <u>http://www.wzgenomics.cn/NPdenovo/;</u> Gene4denovo: <u>http://www.genemed.tech/gene4denovo/search</u>. 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 v2.1.1 (www. <u>https://gnomad.broadinstitute.org/</u>; accessed June 2022).
- All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in
- 711 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	Т	ECM1 (DDDG2P: lipoid proteinosis, AR)	Stopgain	NM_001202858.2: exon7: c.823G>T: p. Glu275*	7.96E-06	6.53E-05
4	RQPA20	1	160722912	С	A	SLAMF7	Stopgain	NM_001282589.2: exon5:c.560C>A: p. Ser187*	0	0
5	IAU29	2	1946866	С	-	MYT1L (MRD39: MIM 616521; SFARI score 1; MYT1L)	frameshift del	NM_001303052.2: exon9: c.394delG: p. Glu132Argfs*42	0	0
6	SMPA53	2	166166958	С	Т	SCN2A (DEE11: MIM 6137219; SFARI score 1; Geisinger Tier 1)	Stopgain	NM_001040143.2: exon6:c.823C>T: p. Arg275*	3.98E-06	4.62E-05
7	SMPA38	2	166201156	С	A	SCN2A (DEE11: MIM 6137219; SFARI score 1; Geisinger Tier 1)	Nonsynonymous	NM_001040143.2:exon15:c.2654C>A:p.Thr885Asn	0	0

8	IAU66	2	232320177	С	-	NCL	frameshift del	NM_005381: exon13: c.1991delG: p. Gly664Glufs*70	0	0
9	SA1	3	38645393	A	Т	SCN5A (BRGDA1: MIM 601144)	Nonsynonymous	NM_198056.3:exon12:c.1700T>A: p.Leu567Gln	8.031E-6	0
10	SMPA75	5	137680632	-	Т	FAM53C	frameshift ins	NM_001135647.2: exon4: c.256dupT: p. Ser86Phefs*14	0	0
11	SMPA75	6	47646871	G	A	ADGRF2	Splicing	NM_153839.7: exon5: c.267+1G>A	4.01E-06	3.27E-05
12	Autism10	6	87969507- 87969508	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1; Geisinger Tier 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p. Glu2054Lysfs*14	0	0
13	IAU65	6	87969506- 87969507	AG	-	ZNF292 (MRD64: MIM 619188; SFARI score 1; Geisinger Tier 1)	frameshift del	NM_015021.3: exon8: c.6159_6160del: p. Glu2054Lysfs*14	0	0
14	RQPA20	11	46389550	А	С	DGKZ	Splicing	NM_201533: exon4: c.379-2A>C	0	0
15	Autism9	12	52188238	G	A	SCN8A (DEE13: MIM 614558; SFARI score 1)	Nonsynonymous	NM_014191.4:c.4608G>A: p.Met1536lle	0	0
16	SMPA73	12	89998008	G	С	ATP2B1 (MRD66: MIM 619910; SFARI score 3S)	Stopgain	NM_001001323.2: exon15:c.2558C>G: p. Ser853*	0	0
17	RQPA20	16	89006655	G	А	LOC100129697/CBFA2T3	Nonsynonymous	NM_001290330.2: exon1: c.53G>A: p. Arg18Lys	0	0
18	SMPA75	19	7734220- 7734221	СТ	-	RETN	frameshift del	NM_001193374.1: exon2: c.8_9del: p. Cys5Serfs*13	4.37E-05	3.27E-04
19	RQPA20	22	20049060	с	G	TANGO2 (Geisinger AR)	Stop gain	NM_152906.7:exon7:c.459C>G: p.Tyr153*	0	0
20	RQPA20	22	22051034	A	G	PPIL2	Splicing	NM_148175: exon21: c.1570-2A>G	0	0

- 713 **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: <u>https://omim.org/</u>; SFARI ASD genes <u>https://gene.sfari.org/</u>; DDD gene2phenotype genes:
- 715 <u>https://www.ebi.ac.uk/gene2phenotype/</u>; Geisinger developmental brain disorder gene database: <u>https://dbd.geisingeradmi.org/</u>;
- 716 NP*denovo*: <u>http://www.wzgenomics.cn/NPdenovo/</u>; Gene4denovo: <u>http://www.genemed.tech/gene4denovo/search</u>. 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 v2.1.1 (www. <u>https://gnomad.broadinstitute.org/</u>; accessed June 2022).
- All variants were validated by IGV (Integrated Genomics Viewer; Broad Institute) and by Sanger sequencing. Validations are shown in
- 720 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	SMPA3	Recessive	30236780	С	-	MAGEB2	frameshift del	NM_002364.4:c.83delC:p.Pro28Leufs*84	0	0
2	IAU66	Recessive	48672948	A	G	HDAC6	Nonsynonymous	NM_006044.4:c.908A>G: p.Tyr303Cys	0	0
3	IAU69	Recessive	55779942	Т	С	RRAGB	Nonsynonymous	NM_006064.4: exon7: c.730T>C: p. Phe244Leu	5.49E-06	1.23E-05
4	IAU45	Recessive	64947718	С	A	MSN	Nonsynonymous	NM_002444.3: exon3: c.139C>A: p. Leu47Met	0	0
5	SMPA15	Recessive	105193588	С	Т	NRK	Stopgain	NM_198465.4: exon27: c.4375C>T: p. Gln1459*	0	0
6	IAU47	Recessive	107170098	A	G	MID2 (MRX101; MIM 300928)	nonsynonymous	NM_012216.4:c.2003A>G:p.Tyr668Cys	0	0
7	IAU24	Recessive	129290446	С	Т	AIFM1 (Cowchock	nonsynonymous	NM_001130847.4: exon2: c. 238G>A: p. Ala80Thr	0	0

						syndr.; MIM 310490)				
8	IAU51	Recessive	140994661	-	A	MAGEC1	frameshift ins	NM_005462.3:c.1471_1472insA; p.Leu491Tyrfs*10	4.37E-05	0
9	IABB8	Recessive	152106616	A	G	ZNF185	Splicing	NM_001178110: exon13: c.942-2A>G	0	0
10	SA1	De Novo	15677205	Т	A	CLTRN (TMEM27)	nonsynonymous &	NM_020665.3:exon3:c.137A>T:p.Glu46V al	0	0
11	SMPA71	De Novo	62926209	G	A	ARHGEF9 (DEE8; MIM 300607; SFARI score 1S; Geisinger Tier1)	nonsynonymous	NM_015185.3: exon3:c.310C>T: :p.Arg104Trp	0	0
12	SMPA84	De Novo	100507601	A	С	DRP2 (CMTX1 MIM 302800; PMID 26227883)	nonsynonymous	NM_001171184.2: exon15: c. 1639A>C: p. Ser547Arg	0	0
13	IAU54	De Novo	153296081	-	Т	MECP2 (RTT; MIM 312750; SFARI score 1; Geisinger Tier1)	frameshift ins	NM_001110792.2:exon3:c.1234dup:p.Th r412Asnfs*5	0	0
14	SMPA78	De Novo	153296516	G	A	MECP2 (RTT; MIM 312750; SFARI score 1' Geisinger Tier1)	Stopgain	NM_001110792.2:exon3:c.799C>T:p. Arg267*	0	0

721 [&]MutationTaster predicts splice donor created

- 723 **Table 5: CNV summary:** validated biallelic loss CNVs called by CNVpartition or ChAS. Biallelic losses called from microarray data by
- 724 PennCNV and from WES data by CLAMMS (80)) 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

726 (<u>www.deciphergenomics.org</u>), 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 2)	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 2)	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	

727 \$ see Riazuddin et al, 2017; PMID_30171209)

Biallelic variants for autism....

728 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.

734	Figure 2: Violin plots comparing consanguinity coefficient Froh, for A: different categories of
735	variant identified: All biallelic variants (N=27 observations), variants in known AR ID and/or ASD
736	genes (N=10 observations), variants in novel candidate genes (N=17 observations), all X-linked
737	(XL) variants plus <i>de novo</i> autosomal dominant (AD) variants (N=18 observations), variants in
738	known XL and AD genes (N=11 observations), and variants in novel candidate XL and AD genes
739	(N=7 observations). Unpaired t test (2-tailed) for comparison of means showed comparison for
740	i) all biallelic versus all <i>de novo</i> autosomal plus X-linked to be extremely significant (p<0.0001,
741	<i>t</i> =5.8309, d.f.=51), ii) known AR versus AD ID and/or ASD genes to be strongly significant
742	(p=0.0045, t=3.1642, d.f.=22), iii) novel candidate AR versus AD genes to be strongly significant
743	(p<0.0001, <i>t</i> =5.4795, d.f.=26). No significant difference in mean F _{roh} was seen for comparisons
744	between known and novel candidate AR genes, or between known and novel AD plus XL genes;
745	B: different cohorts, categorized by last name of principal investigator for each collection (Agha,
746	N=3, Ansar, N=52, Heidari, N=2, Al Ayadhi, N=13, Bozorgmehr, N=10, Sasanfar, N=24
747	observations). While F _{roh} distribution was similar for most cohorts, the Iranian cohort from
748	Sasanfar had a significantly lower mean than most other cohorts; C: grouped by country of

Biallelic variants for autism....

- origin (Pakistan, N=55, Iran, N= 36, Saudi Arabia=13 observations). Comparison of mean Froh
- 750 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
- 753 deviation, and outliers marked by "X".



