Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deleterious coding variation in SFARI-associated genes among Indian individuals with Autism Spectrum Disorders: an exploratory exome-sequencing study

Meghana Kommerahalli RameshRaju
University of Mysore
Ashitha S. Niranjana Murthy
National Institute of Mental Health and Neurosciences
Snijesh Valiya Parambath
St. John's National Academy of Health Sciences
Durbagula Srividhya
St. John's National Academy of Health Sciences
Chaithra Shanthithadda
University of Mysore
Aruna Korlimarla
St. John's National Academy of Health Sciences
Naveen Kumar Gowda
St. John's National Academy of Health Sciences
Vijaya Raman
St. John's National Academy of Health Sciences
Chetan Ghati Kasturirangan
National Institute of Mental Health and Neurosciences
Gautham Arunachal Udupi
National Institute of Mental Health and Neurosciences
Ramachandra Nallur Basappa
nallurbr@gmail.com

University of Mysore

Research Article

Keywords: Autism Spectrum Disorders (ASD), Whole Exome Sequencing (WES), Protein Truncating Variants (PTVs), common and rare variants, Minor Allele Frequency (MAF)

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11281865/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Tables 1 to 4 are available in the Supplementary Files section.

Deleterious coding variation in SFARI-associated genes among Indian individuals with Autism Spectrum Disorders: an exploratory exome-sequencing study

Meghana Kommerahalli Rameshraju .[1,#], Ashitha S.Niranjana Murthy.[1,4,#], Snijesh Valiya Parambath.[2,7], Durbagula Srividhya.[2,6], Chaithra Shanthithadda.[1], Aruna Korlimarla.[2,5], Naveen Kumar Gowda.[2], Vijaya Raman.[2], Chetan Ghati Kasturirangan .[3], Gautham Arunachal Udupi.[3], and Ramachandra Nallur Basappa [1*]

1. Department of Studies in Genetics and Genomics, University of Mysore, Manasagangothri, Mysuru- 570 006, Karnataka, India
2. Centre for Advance Research and Excellence in Autism and Developmental Disorders (CAREADD), St. John's Research Institute, St. John's National Academy of Health Science, Bengaluru -560034, Karnataka, India
3. Department of Human Genetics, National Institute of Mental Health and Neuro Sciences, Bengaluru Urban- 560029, Karnataka, India
4. Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bengaluru Urban- 560029, Karnataka, India
5. Department of Research, Sri Shankara Cancer Hospital and Research Center, Bengaluru - 560004, Karnataka, India
6. Department of Studies in Biotechnology, University of Mysore, Manasagangothri, Mysuru- 570 006, Karnataka, India
7. Division of Molecular Medicine, St. John's Research Institute, St. John's National Academy of Health Sciences, Bengaluru 560034, Karnataka, India.

Shared first authorship

***Corresponding author:**
DR. N. B. RAMACHANDRA,
DAE- Raja Rammanna Chair
DoS in Genetics and Genomics
University if Mysore-Mansagangothri
Mysuru-570006
Email: nallurbr@gmail.com ramachandra@genetics.uni-mysore.ac.in
ORCID ID: 0000-0002-0919-7781

ABSTRACT:

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with impairment in social interaction, communication and restrictive, repetitive behaviours and interests. Considering limited genetic research on ASD in India, we are uncertain of whether its architecture is similar or dissimilar to what is observed globally. Therefore, utilizing SFARI- a well-curated database of ASD-associated genes, we investigated their mutation burden among Indian ASD subjects and studied the impact of variant stratification across populations. For this purpose, Whole Exome Sequencing and analysis was performed for 25 Indian ASD subjects. The findings revealed that only 15% of SFARI genes carried 241 common and rare deleterious variations in our ASD cohort. The most frequent variants were Protein Truncating Variants (PTVs) (69.7%). SFARI score-1 genes had more missense variants than PTVs, hinting at their evolutionarily conserved nature. Seven PTVs in genes, *SON*, *KMT2C*, *EXOC6B, RBFOX1, CAPN12* and *KIZ* were found to be altered in multiple ASD individuals. Notably, 32% of 241 deleterious variations identified were common across populations. The Asian population had significantly greater proportion of rare PTVs than the global population (p-value <0.0071). Multidimensional functional analysis identified 101 genes with neurodevelopmental functions, of which, six genes were altered in unrelated ASD subjects globally which could be potential new ASD-candidate genes. Overall, our study provides evidence that an increased rate of PTVs burdens Indian ASD subjects and common variants levy additional genetic risk. Further, using population-specific human genome reference is essential to obtain meaningful insights into population-specific risks for disease.

Keywords: Autism Spectrum Disorders (ASD), Whole Exome Sequencing (WES), Protein Truncating Variants (PTVs), common and rare variants, Minor Allele Frequency (MAF).

1. Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that is behaviourally defined by challenges in social skills restricted repetitive behaviours, and communication deficits, according to the Diagnostic and Statistical Manual of Mental Disorders-V (DSM-V) (American Psychiatric Association, 2013). The Centers for Disease Control and Prevention (CDC) reported that 1% of the global population is diagnosed with ASD (Baxter et al., 2015); similarly, in India, 1 in 100 children under the age of 10 are affected by ASD (Arora et al., 2018).

ASD has a very high heritability rate (Sandin et al., 2017), and its genetic architecture is being deciphered through various large-scale sequencing efforts. Using Whole Exome/Genome Sequencing (WES/WGS) is a reliable tool to identify causal genes associated with a disease in relevant pathways at a molecular level; over 100 risk-conferring ASD genes have been identified in the last ten years, representing significant progress in the field (O'Roak et al., 2012). Research on molecular genetics has shown that a combination of common and rare variations influences the risk for ASD (Weiner et al., 2017). Recurrent identification of rare *de novo* variations in the same genes across different ASD cohorts lead to the discovery of high-risk genes for ASD. Further, *de novo* and inherited rare protein-truncating variants (PTVs) were also found as predominant risk factors for ASD- prompting the conclusion that rare variants dominate the ASDs allelic spectrum (Iossifov et al., 2012; Sanders et al., 2012).

Nonetheless, research has also indicated that common variations significantly increase the risk of ASD, and individuals with ASD exhibit an enormous burden of common variations (Klei et al., 2021). Therefore, genetic risk for ASD may be conferred by a complex combination of rare variations of large effect and common variations of small effect size.

The Simons Foundation Autism Research Initiative (SFARI) is a publicly available, constantly evolving, curated database of candidate genes for ASD that integrates genomic information from multiple large-scale sequencing studies across the globe. SFARI also provided detailed annotations of the involvement of candidate genes in ASD. To make it simple for users to determine which genes are most likely to remain associated with an increased risk of autism, they classified genes into four groups: score 1, score 2, score 3, and syndromic. The 'syndromic' group comprises variations that are associated with a significant level of higher risk and consistently related to additional traits not necessary for an ASD diagnosis. Score1 genes are considered to be certainly implicated in ASD due to the presence of at least three *de novo* likely-gene-disrupting variations by literature. Score2 genes are reported with two *de novo* likely-gene-disrupting variations, while score3 genes are found to be reported with one *de novo* likely-gene-disrupting variation (Banerjee-Basu & Packer, 2010). This systematic classification of ASD-candidate genes by SFARI enabled us to prioritize genes and screen for the genetic burden of the disease in unrelated populations.

The most extensive ASD exome sequencing study involving 11,986 ASD subjects from the American and European populations identified 102 ASD-associated genes with disruptive *de novo* and rare variants (Satterstrom et al., 2020). These genes were regarded as high-risk, score 1 ASD genes within the SFARI database. Although it integrates ASD-linked genomic variations from studies across the globe, the Indian population is not represented among them.

Until recently, the scope of ASD research in India was restricted to the screening of a limited number of single nucleotide polymorphisms (SNPs) in genes like *OXTR, DNMTs, RELN, MTHFR, HTR2A, EN2, ITGB3, MAOA, TPH2, and DNMTs* (Mandamkulathil et al., 2019; Patra & Kar, 2021). To the best of our knowledge, only one WES study has been published on Indian ASD subjects, which reported rare variants in 22 genes, of which 15 were SFARI genes (Sheth et al., 2023). Therefore, to study the mutation burden levied by these well-established SFARI genes for ASD, we performed WES of 25 Indian ASD subjects to investigate the interplay of common and rare variations in SFARI genes and identify any population stratification among them through their Minor Allele Frequency (MAF) values. Additionally, we found genes outside the SFARI gene list with plausible roles in neurodevelopment and/or ASD.

2. Materials and Methods:

2.1 Recruitment and Assessment of Subjects:

A total of 25 ASD subjects were recruited for this study. Of these, 15 subjects were recruited from the All-India Institute of Speech and Hearing (AIISH), Manasagangotri, Mysore. The Institutional Ethics Committee approved the study (IHEC-UOM No. 135/Ph.D/2016-17). Differential Diagnostic Checklist for ASD (DDC-ASD), an in-house diagnostic tool developed by the ASD Unit experts at AIISH, was used for diagnosis. Additionally, the diagnostic tools used were the Childhood Autism Rating Scale (CARS), Autism Diagnostic Observation Schedule (ADOS), and Modified Checklist for Autism in Toddlers (M-CHAT). Delayed speech and language (DSL) was identified through the Receptive-Expressive Emergent Language (REEL) Test.

Additionally, 10 ASD subjects were recruited at the Centre for Advanced Research and Excellence in Autism and Developmental Disorders (CARE-ADD), St. John's Research

Institute, Bangalore, Karnataka, India, with Institutional Ethics Committee approval (IHEC No. IEC 210/2023). Clinical psychologists and psychiatrists utilized the INCLEN Diagnostic Tool for Autism Spectrum Disorder (INDT-ASD) to conduct a comprehensive clinical assessment. The severity of the condition was assessed using the Indian Scale for Assessment of Autism (ISAA). The degree of social and communication deficits was assessed using the parent-rated Social Responsiveness Scale (SRS). Children with ASD had their adaptive behaviours measured using the Vineland Adaptive Behavior Scale-2 (VABS-2). To test autistic features in first-degree relatives, parents of children with ASD were also given the Autism Spectrum Quotient (AQ). Clinical information, family medical history and demographics of ASD subjects were gathered. Written consent was taken from all families before peripheral blood sampling.

2.2 Whole Exome Sequencing (WES):

The 25 subjects in our cohort were sequenced in two batches. The first, 15 subjects recruited from AIISH were subjected to WES on Illumina HiSeq 2500 paired-end sequencing platform with 150*2 base pair read chemistry with 100X coverage using the Agilent SureSelect All Exon V6+UTRs library preparation kit following the manufacturer's protocol. WES for the next batch of 10 ASD subjects recruited in CARE-ADD was performed on Illumina NextSeq 550 platform with 150 base pair read chemistry with >70X coverage using Twist library preparation EF kit as per manufacturer’s protocol.

2.3 Variant Calling and Annotation:

Variant calling was performed using Genome Analysis ToolKit (GATK) best practices. In brief, it involved a series of steps to ensure high-quality and accurate variant identification. Initially, the raw sequencing data underwent quality control using FastQC to assess overall data quality. Low-quality bases (phred score ≤ 20) and adapter sequences were then trimmed with

TrimGalore and CutAdapt to clean the data. The trimmed reads were aligned to the reference genome (GRCh38) using Burrows-Wheeler Aligner (BWA), ensuring accurate mapping of the reads. The next step involved identifying and marking duplicate reads using the MarkDuplicates module in GATK, which prevented the artificial inflation of variant calls. Following this, the BaseRecalibrator tool was used to recalibrate base quality scores, correcting any systematic errors made by the sequencer. The HaplotypeCaller tool was then employed to identify potential variant sites, leveraging sophisticated algorithms to call variants with high sensitivity and precision. After variant calling, variant quality score recalibration was conducted to filter out false positives. Finally, functional annotation of the variants was performed using Ensembl Variant Effect Predictor (VEP), providing insights into the potential biological impact of the identified variants. This comprehensive workflow ensured reliable and reproducible variant calling results.

2.4 Variant filtering and prioritization:

Read quality-based filtering was applied using stringent parameters like depth (DP ≥ 20) and Genotype Quality (GQ ≥ 20). We considered 'high impact' variants such as PTVs (frameshift InDels, stop gain, stop loss), Splice variants (splice donor and acceptors), and 'moderate' impact missense/nonsynonymous SNPs (nsSNP). Further, we only selected nsSNPs that were predicted to be deleterious by 10 out of 16 (60%) different "SNP effect prediction" tools such as SIFT, PolyPhen, CADD PHRED, REVEL, MutPred, ClinPred, DEOGEN2, FATHMM, LIST-S2, LRT, M-CAP, MetaLR, MutationAssessor, MutationTaster, PROVEAN, and BayesDel. This allowed for the separation of deleterious nsSNPs from benign/tolerant variations. Importantly, to obtain a holistic distribution profile of all the prioritized high-impact PTVs, Splice variants, and deleterious nsSNPs in SFARI genes, we included both common (MAF 0.01-1.0) and rare (MAF ≤ 0.01 and novel) variants for the entirety of this study.

2.5 Gene prioritization and characterization:

In order to holistically inspect the contributions of SFARI genes in the Indian ASD subjects, we included all variations identified in SFARI genes irrespective of their MAFs. Based on SFARI gene scoring, the prioritized genes were further categorized into 'high-confidence' scores 1 and 2 genes and 'suggestive-evidence' scores 3 and 0 genes. We considered genes that are not under any scoring category in SFARI as Score 'Zero'. To explore the distribution of prioritized genes across four different SFARI score categories, we further looked for genes that were altered in < 10% of the individuals, 10%–50% of the individuals, 50%–95% of the individuals, and > 95% of the individuals. The genes that are altered in > 95% of individuals were considered 'recurrently altered' genes.

In order to delineate population differences in the MAFs between different ethnicities, we compared MAFs of all prioritized variants across global populations as described in 1000 Genome Phase 3 (1000Gp3) and gnomAD databases with that of MAF in Asian and Indian populations as described in GenoAsia and IndiGen databases, respectively. The chi-sq test was used to determine if these differences were significant using a value cut-off of >0.05, and variants showing significant differences between populations in MAF were further visualized by plotting a heatmap using the ggboxplot function in the ggplot2 R package.

2.6 Identification of new candidate ASD genes among nonSFARI genes:

Using the Indigenome database as our reference genome set, we collected all genes carrying rare variations in ASD subjects. After excluding all the SFARI genes from this list, we analyzed the remaining 'nonSFARI' genes to explore if there were additional genes with neurodevelopmental functions and/or linked to ASD-related behaviours/functions. We referred to three well-curated databases for brain development genes: SysNDD, Developmental Brain Disorders (DBD) and Development Disorder Genotype-Phenotype Database (DDG2P). Our

WES study identified genes overlapping with genes in these three databases. We performed gene enrichment analysis using the tool EnrichR. Enrichment terms were selected with a cut-off p-value <0.01, and enrichment terms directly connected with autism phenotypes like social impairment, restrictive, repetitive behaviours and communication deficits were also selected. Terms with the highest combined scores were selected; it is computed by taking the log of the p-value from the Fisher exact test and multiplying that with the z-score of the deviation from the expected rank.

Further, protein-protein interaction (PPI) analysis was performed using the tool STRING-db. We used all active interaction resources, such as experimental evidence, co-expression, occurrence, gene fusion, text mining, databases, and neighbourhood. Clustering was performed using the DBSCAN method, which finds clusters based on the local node density.

3. Results:

3.1 Only 15% of SFARI genes carried deleterious variations in our ASD cohort

In order to holistically inspect the contributions of SFARI genes in the Indian ASD cohort, we included all common and rare deleterious variations identified within SFARI genes irrespective of their MAFs. A total of 170 out of 1,128 SFARI genes (15%) carried 241 unique deleterious variations in 25 ASD subjects (Fig. 1a, Table 1, Supplementary Table S1 and S2). Of the 170 SFARI genes carrying variations, 28 genes were SFARI score 1 genes. However, SFARI score 2 category genes had 120 genes carrying variations in our cohort which was the highest. Together these 148 (87% of 170 genes) score 1 and 2 genes hosting variations in our cohort were considered as 'high-confidence' ASD genes. The remaining 22 (13%) score 3 and 0 genes were considered 'suggestive-evidence' genes. Each ASD subject had an average of

1.64 unique variations in score 1 genes and an average of 8.6 variations in 'high-confidence' ASD genes (Table 1).

3.2 PTVs were most frequent in Indian ASD subjects, but the score 1 gene was intolerant to them

PTVs topped the list as the most frequent type of variation (69.7%) carried by ASD subjects of our cohort, followed by missense variations (21.4%) (Fig. 1b, Supplementary Table S2). However, when considering the unique variants, missense SNP was the most common type of variation (50.6%) identified in 103 SFARI genes, closely followed by PTVs (41.5%) in 72 genes and 19 splice variants in 17 genes. Therefore, the contribution of PTVs was more significant than missense variants towards an individual ASD subject's genetic susceptibility. The proportion of genes carrying unique missense variation (53.65%) was more than genes carrying PTVs (37.5%) and splice-site variants (8.85%) (Fig. 1b).

Overall, each ASD subject had an average of 30.4 variations per subject (including deleterious variants that are both common and rare) in an average of 28.12 genes (Fig. 1c, Supplementary Table S2), including 21.2 PTVs and 6.52 missense variations per ASD subject. Score 1 category genes had a significantly higher proportion of deleterious missense variations than PTVs. In contrast, there was a marginal difference in the proportion of PTVs and missense variations in other categories (Fig. 1d). We suspected that this is because score 1 genes are evolutionarily highly conserved. When we checked for their probability of loss of function intolerance score (pLI) and Loss-of-function Observed / Expected Upper bound Fraction (LOEUF) scores, we identified that 21 of the 28 score 1 category genes had pLI score ≥ 0.9 and LOEUF score ≤ 0.2 2 making them extremely intolerant to damaging variations.

3.3 Variations in 'high-confidence' ASD genes were less recurrently altered

On average, each ASD subject carried 1.64 and 6.96 variations in score 1 and score 2 'high confidence' genes, respectively (Table 1). When we checked if the same variation was altered in multiple ASD subjects, we observed that 110 of these 170 genes (64.7%) were altered in less than three (<10%) ASD subjects. In contrast, variations in the remaining 12 genes were altered in over half of the cohort (Fig. 1e, Supplementary Table S3A and S3B). 'High-confidence' ASD genes hosted variations that were not as recurrently altered in ASD subjects of our cohort as often as variations in 'suggestive evidence' genes (p-value 0.0008) (Table 2), reiterating that high-confidence genes are more resistant to molecular changes

3.4 Upto 32% of the deleterious variations identified were common across populations

To delineate the influence of allele frequencies on all the pathogenic variations identified in our ASD cohort, we utilized the MAF information from three different reference genome datasets, namely GnomAD, Indigenome and GenomeAsia, representing the global, Indian and South Asian populations, respectively. On categorizing all 241 pathogenic variations as common and rare variants based on the cut-off of MAF > 0.01 and MAF < 0.01, respectively, we noted significantly higher occurrences of rare variants in the Asian reference datasets compared to the global reference genome (p-value 0.0253) (Fig. 1f, Supplementary Table S4). Several variations designated as common by gnomAD turned out to be rare variants in GenomeAsia and Indigenome datasets. Interestingly, despite filtering for variants predicted to be highly deleterious by reputed SNP effect prediction tools, it is interesting that about 23-32% of the 241 variations in SFARI genes were common across populations (Supplementary Table S4). It may be possible that these common variants within 'high-confidence' SFARI genes- with the propensity of causing a deleterious effect on protein function- might be contributing to the increased prevalence of ASD across the globe.

3.5 When considering individual SNPs, the Indian ASD population carried rare PTVs in 'high-confidence' SFARI genes that were common variations in the global population

When we looked into individual SNPs, we identified 37 variants from 27 genes opposing MAF values in GnomAD and Indigenome datasets. Of these 27 genes, 24 were found to be high-confidence genes, and 3 were found to be suggestive-evidence genes. Interestingly, our analysis found an increased consensus in MAF values between databases compared to what was expected- p-value 3.98E-23 when considering variants (Supplementary Table S5). Among the 170 genes, variants in 106 genes (62%) were rare, and 37 genes were common in both databases. The 27 genes with discordant MAFs included score 1 genes like *EIF3G, GIGYF2, KMT2C* and *RAI1*, and 20 other score 2 genes. Variants in genes *KMT2C, CUL7, OR52M1* and *LRP2* were rare in GnomAD but common in the Indigenome database, whereas 16 genes, including score 1 *GIGYF2* and *RAI1* and 12 other score 2 genes, carried PTVs that were common in GnomAD and rare in Indigenome (Fig. 2a). This was despite our dataset showing significantly higher consensus in MAFs among rare variants than expected.

Further, when we looked into individual variants to explore the population stratification of their MAFs, we identified 33 variants that showed differences in frequency distribution. Variants in high-confidence genes (*PCDHA4*, *KMT2C, CNTN5* and *EIF3G*) and suggestive-evidence genes (*HOXA1, POLR2A* and *PABPC1*) were novel among African, American, East Asian, South Asian and European population, but common in Indian population. Variants in *HYDIN, TBCK, SBF1* and *ABCA7* were common among other sub-populations and rare for the African population, indicating that these variants could be population-specific. (Supplementary Table S11)

3.6 Prevalence of PTVs in SFARI genes differs significantly between populations

Comparison of common and rare variations concerning their SNP type indicated a significant difference in the frequency of PTVs between global and Asian populations (p-value 0.0071) but not in the frequencies of missense and splice SNPs (Supplementary Table S6). Reiterating our previous observation with statistical support, GnomAD had more common PTVs, and Indigenome had more rare PTVs than expected, indicating a difference in the prevalence of ASD-associated PTVs between populations (Fig. 2b). We further probed what contributed to this pattern and found that it was mostly by score 2 genes- 92% of rare PTVs in GnomAD were from score 2 genes and 82% for Indigenome (Fig. 2c).

3.7 Rare pathogenic variations in SFARI genes were often unique to an ASD subject

Variations recurrently altered in more than 50% of the 25 ASD subjects majorly composed of common SNPs, but an almost equal proportion of common and rare variants were recurrently altered in 10-50% of the cohort. Remarkably, variants that were altered in <10% of subjects were rare variants (94.5%) (Fig. 2d-f, Supplementary Table S6 S7). This difference was more pronounced in score 2 genes, as iterated in Fig. 1e.

3.8 Only 12% of SFARI genes hosted rare variants in Indian ASD subjects

Further, when we excluded all common variants from our study to include only rare variants with MAF <0.01, 163 variants in 132 genes (11.7% of SFARI genes) were retained according to GnomAD reference, whereas 180 variants in 137 genes remained when Indigenome reference database was considered (Fig. 2g). Although marginal, the choice of genome reference did show an influence on final variants and gene prioritization.

3.9 Seven PTVs were recurrently altered in our cohort

Seven variants in a total of six genes (*SON*, *KMT2C*, *EXOC6B, RBFOX1, CAPN12* and *KIZ*) out of 170 genes were found recurrently altered in our cohort. Of these, two were score 1

genes *SON* and *KMT2C*, three score 2 and one score 3 genes (Table 3). Interestingly, all these variants were PTVs.

3.10 Seven variants previously reported in global ASD subjects were also present in our cohort

We found seven previously reported SFARI variants in seven genes of our ASD subjects - *RELN, ABCA10, ASPM, BRCA2, MTHFR, FCRL6*, and *PEBP4* (Table 4). Their MAFs were concordant between populations and included two rare PTVs in genes *BRCA2* (found in three ASD subjects) and *FCRL6* and two rare missense variants in *RELN* (score 1) and *ASPM*, two common PTVs in *ABCA10* (in three ASD subjects) and *PEBP4* (in four ASD subjects), one common missense variant in *MTHFR* found in seven ASD subjects of our cohort. It is interesting to note that common variants listed in the SFARI database were also common within our cohort and altered in more than three ASD subjects.

3.11 101 nonSFARI genes showed the strong potential of being an ASD candidate gene

Since only 12% of SFARI genes hosted rare variations, we analyzed the remaining 1871 nonSFARI genes for new genes with plausible roles in neurodevelopment. First, we identified 153 of these genes as definitive neurodevelopmental genes by sysNDD, DBD and DDg2p databases. Of them, six genes, *ARCN1, DNM1, KIDINS220, NALCN, ZFHXA1* and *PDHA1*, were reported to be mutated in ASD subjects, according to Fu et al. (2022). Enrichment analysis identified strong neuronal functions in 101 genes; nine were expressed in brain tissues, and 22 genes had synaptic functions. PPI analysis showed significantly more interactions among 41 genes than expected (p-value 3.57e-11) and had 8 clusters in total (Fig. 3).

4. Discussion:

More than 1000 candidate genes have been identified for ASD according to the well-curated database for ASD genes, i.e., the SFARI database. However, this data is derived primarily from ASD studies in North American and European populations, where the Indian/South Asian population is not represented or is under-reported. Therefore, we aimed to study the genetic overlaps and differences between global and Indian ASD subjects. Although previous studies have utilized the SFARI database to prioritize genes and have also discussed the importance of common variants in ASD risks (Leblond et al., 2019), our study differs in its approach as it is designed to explore the population stratification of MAFs of variants in SFARI genes across ethnicities that allowed us to identify population-specific variants in well-established ASD genes.

In order to explore the cumulative contributions of SFARI genes in the Indian context, we included all pathogenic variations identified in SFARI genes irrespective of their MAFs. In total, 15% of SFARI genes carried 241 deleterious variations (both common and rare) in our ASD cohort. Of them, rare deleterious variations were present in only 12% of SFARI genes- a relatively small representation of overall SFARI genes. Other ASD studies in the South Asian populations were reviewed to examine the proportion of SFARI genes mutated. We found a very recent WES study on 101 Indian ASD trios, which reported variations in 22 genes, 15 (1.33%) of which were SFARI genes. These differences are primarily due to post-gross prioritization methods like filtering in for previously reported disease-causing variants in databases like ClinVar and HGMD (Sheth et al., 2023). Similarly, a study on 100 Vietnamese ASD trios identified variations in 17 genes, 8 (0.8%) of which were SFARI genes. Unlike our findings, missense was their study's most common variant type (61%). While we only included variants with MAF ≤ 0.0, the Vietnamese study restricted it to variants with MAF < 0.001, which could be the reason for the differences in the proportion of SFARI genes (Tran et al., 2020). Likewise, a Saudi trio-based study on 19 simplex ASD families identified variations in

47 genes, of which three (0.2%) were SFARI genes. This study also reported missense variants to be the most frequent (77%), followed by PTVs. Unlike our study, the variant prioritization was done by excluding variants based on multiple databases like dbSNP, 1000 Genomes project, and ExAC, as well as their Saudi reference database (SHGP 2379 exomes) with MAF ≥ 1%, which may have contributed to the observed differences (Al-Mubarak et al., 2017). Another South-Asian population study from Pakistan with 36 families identified variations in 15 genes and showed no overlap with SFARI genes despite reanalysis of biallelic variants with less stringent MAF values (0.001, 0.01) to check if any known ASD/ID genes were missed. This could be due to various parameters applied like exclusion by clinical variant search using ClinVar and ClinGen databases or dbSNP; differing minor allele frequency cut-offs like ≤ 0.001 for autosomal recessive, or ≤ 0.00001 for autosomal dominant/de novo (Khan et al., 2024).

When only rare variants with MAF < 0.01 were considered using the Indigenome reference database, we found 180 variants in 137 (12%) SFARI genes. Compared to other South Asian population studies, we found a greater proportion of SFARI genes, which could be due to using only one reference genome database for filtering variants. Other studies have used both global and indigenous genome references for variant prioritization, which could lead to the exclusion of more variants than our study has reported. Further, we have not used any clinical variant classification databases for the study, but we have used SNP effect prediction methods to prioritize our variants, a standard protocol. The techniques employed in each study are summarized in the Supplementary Table.S12.

PTVs are thought to levy a higher mutational burden than missense variations on protein structure and functions, suggesting that the average effect on liability caused by PTVs is greater than missense variations (He et al., 2013). In our study, when we considered common and rare variants, we observed a significantly higher frequency of PTVs (69.7%) than missense variations (21.4%). Even when considering only rare variants, we observed a higher proportion

of PTVs (60.9%) than missense variations (36.5%). Complementing our findings, there are reports that ASD-associated genes are dominated by *de novo* PTVs compared to damaging missense variants (Fu et al., 2022; Iossifov et al., 2014). Likewise, other comprehensive studies on ASD families also indicated a higher burden of rare inherited PTVs in ASD risk genes in ASD subjects (Cirnigliaro et al., 2023; Krumm et al., 2015). Missense variants were found to be less frequent due to smaller sample sizes, and as Sanders,(2019) mentioned in the review, larger sample sizes were required to reliably show a significant excess of *de novo* missense variants, independent of de novo PTVs.

In our study, SFARI score 1 category of genes had a higher proportion of missense variations than PTVs. When we looked into the pLI and LOEUF scores for these score 1 genes, we found that the majority of them had pLI score ≥ 0.9 and LOEUF score ≤ 0, rendering them intolerant to pathogenic PTVs but allowing the less harmful missense variants... Hence, score-1 genes are evolutionarily conserved. We also noticed that 'high-confidence' score 1 and 2 SFARI genes were not as recurrently altered as 'suggestive evidence' genes- which were more often altered in multiple ASD subjects. This could imply that 'suggestive evidence' genes might have a mild role in ASD-related processes or are less often evolutionarily conserved and thus may not be protected from acquiring more variants. Alternatively, they may be recognized as ASD genes when sufficient evidence accumulates with increased sequencing. At the same time, less recurrently altered 'high-confidence' genes could be well protected as they play a significant role in ASD-related fundamental biological processes.

Interestingly, our study found that about 32% of the 241 deleterious variations in SFARI genes were common variants that contribute substantially to ASD liability and add significant genetic burden on ASD individuals (Gaugler et al., 2014). Our observations questioned if common variants in prominent SFARI genes may have the propensity to cause deleterious effects on protein function and ASD phenotype and influence the increase in the prevalence of

ASD across the globe. Along these lines, common variants in the *EIF4E* gene were identified to be associated with etiology of ASD and behavioral parameters for repetitive and stereotypical behaviors (Waltes et al., 2014). Similarly, Alarcón et al. (2008) showed substantial support for a common variation in *CNTNAP2*'s role in language development in ASD. They provided a framework for linking this variation to the organization of brain circuits essential for human verbal communication. Matoba et al. (2020) highlight the regulatory effect of distal common SNP on *DDHD2* in neural progenitors of the developing cortex. Grove et al. (2019) reported that common risk variants are associated with and more prominent in high-functioning ASD in the Danish population. These findings reassert that a large number of deleterious common variants with small effect sizes may add additional burden to ASD etiology.

Various reports have mentioned the importance of allelic frequency in epidemiological and genetic studies (Taioli et al., 2004). Allele frequency of a SNP can vary drastically between ethnic groups (Goddard et al., 2000). In order to delineate population differences in the MAFs between different ethnicities, we compared MAFs obtained from different reference genomes. This analysis indicated that the Asian reference datasets had more rare and novel variants than global genome references. One of the possibilities could be the differences in sample sizes between reference genome datasets. Another possibility could be that we considered unreported or unavailable variants in the Indigenome dataset rare. Increased and continuous population sequencing efforts will aid in better determination of variant frequencies in future. Additionally, when we looked at individual SNPs, we identified variants that had opposite MAF values in GnomAD and Indigenome datasets and variants that are common across the global population are found to be rare for the Indian population. These differences in MAFs indicate genetic heterogeneity in different ethnicities.

Further, we explored the possible roles of other genes outside the SFARI database and found 153 genes to be associated with neurodevelopmental functions. Interestingly, six

(*ARCN1, DNM1, KIDINS220, NALCN, ZFHXA1* and *PDHA1)* genes were also reported to be mutated in ASD subjects, according to Fu et al., (2022). Furthermore, *de novo* pathogenic variants were found reported in *ARCN1, DNM1,* and *KIDINS220* genes in subjects exhibiting mild to moderate ASD with other associated disorders (Brereton et al., 2018; Izumi et al., 2016; Lazzara et al., 2018; Lam et al., 2020; Mei et al., 2023). Additionally, a study in a transgenic mouse line with reduced *KIDINS220* expression showed impaired response to both auditory and olfactory stimuli and also exhibited impaired startle response to high-intensity sounds and strongly reduced social odour recognition (Albini et al., 2024). Individuals with broader clinical phenotypes, including autistic features, were identified with homozygous variants in the *NALCN* gene (Bramswig et al., 2018). In ASD children, the accumulation of pyruvate and lactate was studied due to the genetic alteration in the X-linked gene coding for the PDHA1 subunit. This results in insufficient energy production and could lead to brain dysfunction in ASD (Lissens et al., 2000). Hence, these findings suggest that further studies on the genes mentioned above in ASD individuals are needed.

In addition, gene function enrichment analysis showed the involvement of genes in strong neuronal functions and synaptic functions. Remarkably, *CENPJ, PRSS12, NALCN* and *ST3GAL3* were enriched for abnormal social behaviour, which is one of the core behaviours of ASD. Complementing these findings, several studies have illustrated the association of these genes with social and behavioural abnormalities. McIntyre et al. (2012) reported that *CENPJ* knockout mice showed significant long-term memory impairment, indicating its role in social recognition. Similarly, *PRSS12* knockout mice took more time to investigate the familiar mouse in social novelty tests and affect social behaviours (Mitsui et al., 2009). Additionally, a sequencing study reported a variant in *NALCN* (a missense chr13 101756915G>A) in an individual with profoundly affected speech and social skills who was diagnosed with ASD and learning disabilities (Campbell et al., 2018). *ECM1, MANBA, UROC1, CEP152,* and *MAN1B1*

genes were enriched for aggressive behaviour. A WES study reported a novel pathogenic splice site variant in *ST3GAL3* (NM_174964.4: c.936+1delG) in a 12-year-old child who shows aggressive behaviour, has poor auditory and visual tracking and also exhibits autistic features, including poor eye contact, repetitive behaviour and difficulty in communication (Khamirani et al., 2021). Thus, these findings indicate that the neurodevelopmental spectrum of these genes may be more comprehensive and underreported, mostly among ASD individuals.

This study's sample size is small and does not include a trio model. However, we employed stringent parameters and multidimensional approaches for variant analysis and prioritization. The results are suggestive, and therefore, we plan to further validate our findings in a larger cohort in the near future.

5. Conclusion:

We report the primary study from India focused on identifying and characterizing deleterious variants in well-established ASD genes from the SFARI database in Indian ASD subjects. Our cohort showed pathogenic variants in 15% of SFARI genes, a relatively minor representation of SFARI genes. Our cohort showed a greater frequency of PTVs - adding an increased burden on Indian ASD 412subjects. Importantly, we found that common variants could be an additional risk factor for ASD etiology. Therefore, incorporating both rare and common high-impact pathogenic variants in well-established ASD genes would be beneficial to better understand the genetic underpinnings of ASD. We recommend using population-specific genome reference databases for meaningful genetic interpretation. Thus, studies in this direction would help better diagnose and treat ASD.

Author Contributions:

Meghana Kommerahalli Ramesh Raju: Data collection, study design, complete data analysis and data interpretation, design paper, complete manuscript preparation, review and

revision. Ashitha S Niranjana Murthy: Conceived and conceptualized the idea and study design, completed data interpretation, contributed to manuscript preparation, and critically reviewed the manuscript. Snijesh Valiya Parambath: Complete WES data analysis. Chaithra Shanthithadda: Data analysis. Durbagula Srividhya: Patient recruitment, sampling, and processing: Aruna Korlimarla: Supervision and mentoring. Naveen Kumar Gowda: Mentoring and supervision. Vijaya Raman: Clinical assessment and clinical inputs. Chetan Ghati Kasturirangan: Provided sequencing facility. Goutham Arunachal Udupi: Provided sequencing facility. Ramachandra Nallur Basappa: Design the concept and framework of the study, critically monitor and review the manuscript and provide final approval.

Acknowledgement:

We acknowledge the Department of Studies in Genetics and Genomics and the University of Mysore for providing us with laboratory facilities. We thank Dr Prakash P from the Department of Psychology, University of Mysore, Dr Shyamala from AIISH, Mysore and the late Dr Ashok M V from CARE-ADD, St John's Research Institute for providing the samples for this project. We thank Dr Ashok Manjula William James from CARE-ADD and Dr Swati Agarwala for their help in the initial stages of this project. We thank all databases and tools included in this study for making them readily available for research use. We thank the Indian Council of Medical Research (ICMR), New Delhi, India, for the senior-research fellowship (SRF) provided to Meghana K R and Chaitra S., DAE-RRC grant to Dr. N B Ramachandra and DST for providing the DST INSPIRE Faculty grant to Dr. Ashitha S.N.M.

Statements & Declarations

Ethics approval and consent to participate:

This study was performed in line with the principles of the Declaration of Karnataka. Approval was granted for 15 subjects by the Ethics Committee of University of Mysore (2016-17/135) and for 10 subjects by the Ethics Committee of St. John's Research Institute, Bangalore, Karnataka, India (2023/210)

Consent to participate

Written informed consent was obtained from the parents.

Consent to publish: Not Applicable

Funding:

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The datasets generated during and/or analysed during the current study are not available as we intend to maintain the confidentiality of participants, but are available from the corresponding author on reasonable request.

References:

Al-Mubarak B, Abouelhoda M, Omar A, et al (2017) Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from Saudi families. Scientific Reports 7(1), 5679. https://doi.org/10.1038/s41598-017-06033-1

Alarcón M, Abrahams BS, Stone JL, et al (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. American Journal of Human Genetics 82(1), 150–159. https://doi.org/10.1016/j.ajhg.2007.09.005

Albini M, Almacellas-Barbanoj A, Krawczun-Rygmaczewska A, et al (2024) Alterations in KIDINS220/ARMS Expression Impact Sensory Processing and Social Behavior in Adult Mice. International Journal of Molecular Sciences 25(4). https://doi.org/10.3390/ijms25042334

Arora NK, Nair MKC, Gulati S, et al (2018) Neurodevelopmental disorders in children aged 2-9 years: Population-based burden estimates across five regions in India. PLoS Medicine 15(7), e1002615. https://doi.org/10.1371/journal.pmed.1002615

Banerjee-Basu S, & Packer A, (2010) SFARI Gene: an evolving database for the autism research community. In Disease models & mechanisms Vol. 3, 133–135. https://doi.org/10.1242/dmm.005439

Baxter AJ, Brugha, TS, Erskine HE, et al (2015) The epidemiology and global burden of autism spectrum disorders. Psychological Medicine 45(3), 601–613. https://doi.org/10.1017/S003329171400172X

Bramswig NC, Bertoli-Avella AM, Albrecht B, et al (2018) Genetic variants in NALCN-UNC80-UNC79 ion channel complex components cause a broad clinical phenotype (NALCN channelopathies). Human Genetics 137(9), 753–768. https://doi.org/10.1007/s00439-018-1929-5

Brereton E, Fassi E, Araujo GC, et al (2018) Mutations in the PH Domain of DNM1 are associated with a nonepileptic phenotype characterized by developmental delay and neurobehavioral abnormalities. Molecular Genetics & Genomic Medicine 6(2), 294–300. https://doi.org/10.1002/mgg3.362

Campbell J, Fitz Patrick DR, Azam T, et al (2018) NALCN Dysfunction as a Cause of Disordered Respiratory Rhythm with Central Apnea. Pediatrics 141, S485–S490. https://doi.org/10.1542/peds.2017-0026

Cirnigliaro M, Chang TS, Arteaga SA, et al (2023) The contributions of rare inherited and polygenic risk to ASD in multiplex families. Proceedings of the National Academy of Sciences of the United States of America 120(31), e2215632120. https://doi.org/10.1073/pnas.2215632120

Fu JM, Satterstrom FK, Peng M, et al (2022) Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nature Genetics 54(9), 1320–1331. https://doi.org/10.1038/s41588-022-01104-0

Gaugler T, Klei L, Sanders SJ, et al (2014) Most genetic risk for autism resides with common variation. Nature Genetics 46(8), 881–885. https://doi.org/10.1038/ng.3039

Goddard KA, Hopkins PJ, Hall JM, et al (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. American Journal of Human Genetics 66(1), 216–234. https://doi.org/10.1086/302727

Grove J, Ripke S, Als TD, et al (2019) Identification of common genetic risk variants for autism spectrum disorder. Nature Genetics 51(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

He X, Sanders SJ, Liu L, et al (2013) An integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genetics 9(8), e1003671. https://doi.org/10.1371/journal.pgen.1003671

Iossifov I, O'Roak BJ, Sanders SJ, et al (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature 515(7526), 216–221. https://doi.org/10.1038/nature13908

Iossifov I, Ronemus M, Levy D, et al (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74(2), 285–299. https://doi.org/10.1016/j.neuron.2012.04.009

Izumi K, Brett M, Nishi E, et al (2016) ARCN1 Mutations Cause a Recognizable Craniofacial Syndrome Due to COPI-Mediated Transport Defects. American Journal of Human Genetics 99(2), 451–459. https://doi.org/10.1016/j.ajhg.2016.06.011

Khamirani HJ, Zoghi S, Faghihi F, et al (2021) Phenotype of ST3GAL3 deficient patients: A case and review of the literature. European Journal of Medical Genetics 64(8), 104250. https://doi.org/10.1016/j.ejmg.2021.104250

Khan H, Harripaul R, Mikhailov A, et al (2024) Biallelic variants identified in 36 Pakistani families and trios with autism spectrum disorder. Scientific Reports 14(1), 9230. https://doi.org/10.1038/s41598-024-57942-x

Klei L, McClain LL, Mahjani B, et al (2021) How rare and common risk variation jointly affect liability for autism spectrum disorder. Molecular Autism 12(1): 66. https://doi.org/10.1186/s13229-021-00466-2

Krumm N, Turner TN, Baker C, et al (2015) Excess of rare, inherited truncating mutations in autism. Nature Genetics 47(6): 582–588. https://doi.org/10.1038/ng.3303

Lam Z, Albaba S, Study D, et al (2020) Atypical, milder presentation in a child with CC2D2A and KIDINS220 variants. Clinical Dysmorphology 29(1): 10–16. https://doi.org/10.1097/MCD.0000000000000298

Lazzara A, Asghar S, Zacharia T, et al (2018) DNM1 Mutation in a child associated with progressive bilateral mesial temporal sclerosis. In Clinical case reports Vol. 6: 2037–2039. https://doi.org/10.1002/ccr3.1793

Leblond CS, Cliquet F, Carton C, et al (2019) Both rare and common genetic variants contribute to autism in the Faroe Islands. NPJ Genomic Medicine 4: 1. https://doi.org/10.1038/s41525-018-0075-2

Lissens W, De Meirleir L, Seneca S, et al (2000) Mutations in the X-linked pyruvate dehydrogenase (E1) alpha subunit gene (PDHA1) in patients with a pyruvate dehydrogenase complex deficiency. Human Mutation 15(3): 209–219. https://doi.org/10.1002/(SICI)1098-1004(200003)15:3<209:AID-HUMU1>3.0.CO;2-K

Mandamkulathil D, Pillai AA, Poovathinal SA, et al (2019) Lack of association of OXTR variants with autism spectrum disorders in a south Indian population. Journal of Indian Association for Child and Adolescent Mental Health 15(2): 102–123. https://doi.org/10.1177/0973134220190207

Matoba N, Liang D, Sun H, et al (2020) Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. Translational Psychiatry 10(1): 265. https://doi.org/10.1038/s41398-020-00953-9

McIntyre RE, Lakshminarasimhan Chavali P, Ismail O, et al (2012) Disruption of mouse Cenpj, a regulator of centriole biogenesis, phenocopies Seckel syndrome. PLoS Genetics 8(11): e1003022. https://doi.org/10.1371/journal.pgen.1003022

Mei D, Parrini E, Bianchini C, et al (2023) Autism and mild epilepsy associated with a de novo missense pathogenic variant in the GTPase effector domain of DNM1. In American journal of medical genetics Part C, Seminars in medical genetics. https://doi.org/10.1002/ajmg.c.32044

Mitsui S, Osako Y, Yokoi F, et al (2009) A mental retardation gene, motopsin/neurotrypsin/prss12, modulates hippocampal function and social interaction. The European Journal of Neuroscience 30(12): 2368–2378. https://doi.org/10.1111/j.1460-9568.2009.07029.x

O'Roak BJ, Vives L, Girirajan S, et al (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485(7397): 246–250. https://doi.org/10.1038/nature10989

Patra S, & Kar SK (2021) Autism spectrum disorder in India: a scoping review. International Review of Psychiatry (Abingdon, England) 33(1–2): 81–112. https://doi.org/10.1080/09540261.2020.1761136

Sanders SJ (2019) Next-Generation Sequencing in Autism Spectrum Disorder. Cold Spring Harbor Perspectives in Medicine 9(8). https://doi.org/10.1101/cshperspect.a026872

Sanders SJ, Murtha MT, Gupta AR, et al (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485(7397): 237–241. https://doi.org/10.1038/nature10945

Sandin S, Lichtenstein P, Kuja-Halkola R, et al (2017) The Heritability of Autism Spectrum Disorder. JAMA 318(12): 1182–1184. https://doi.org/10.1001/jama.2017.12141

Satterstrom FK, Kosmicki JA, Wang J, et al 2020 Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 180(3): 568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036

Sheth F, Shah J, Jain D, et al (2023) Comparative yield of molecular diagnostic algorithms for autism spectrum disorder diagnosis in India: evidence supporting whole exome sequencing as first tier test. BMC Neurology 23(1): 292. https://doi.org/10.1186/s12883-023-03341-0

Taioli E, Pedotti P, & Garte S (2004) Importance of allele frequency estimates in epidemiological studies. Mutation Research 567(1): 63–70. https://doi.org/10.1016/j.mrrev.2004.06.001

Tran KT, Le VS, Bui HTP, et al (2020) Genetic landscape of autism spectrum disorder in Vietnamese children. Scientific Reports 10(1): 5034. https://doi.org/10.1038/s41598-020-61695-8

Waltes R, Gfesser J, Haslinger D, et al (2014) Common EIF4E variants modulate risk for autism spectrum disorders in the high-functioning range. Journal of Neural Transmission (Vienna, Austria: 1996) 121(9): 1107–1116. https://doi.org/10.1007/s00702-014-1230-2

Weiner DJ, Wigdor EM, Ripke S, et al (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nature Genetics 49(7): 978–985. https://doi.org/10.1038/ng.3863

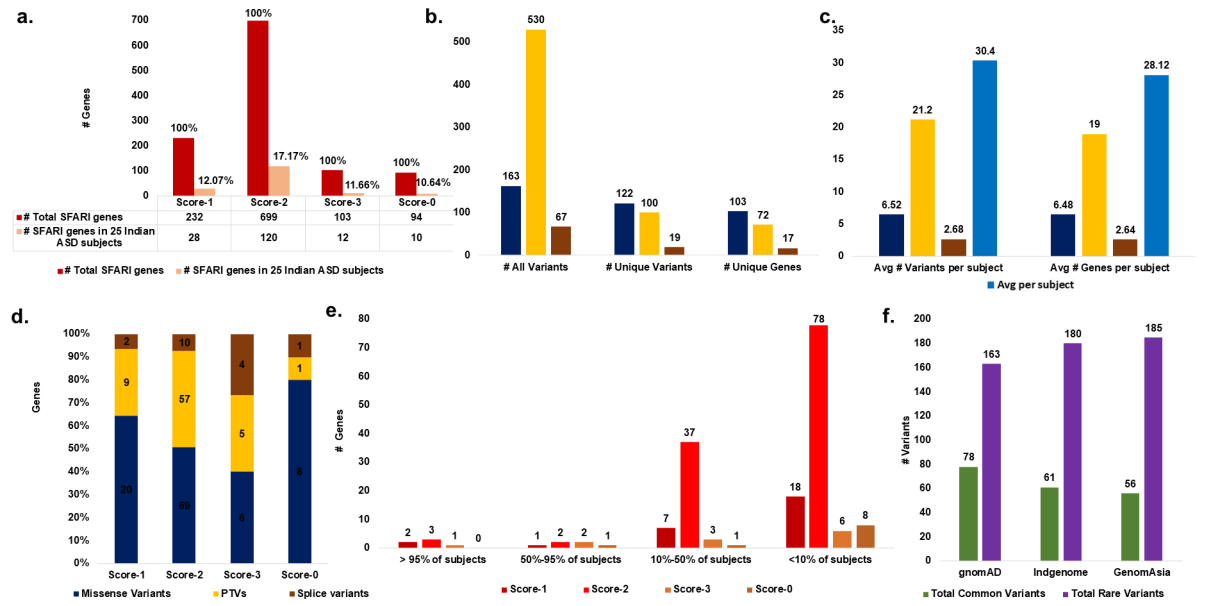
Figure Legends:

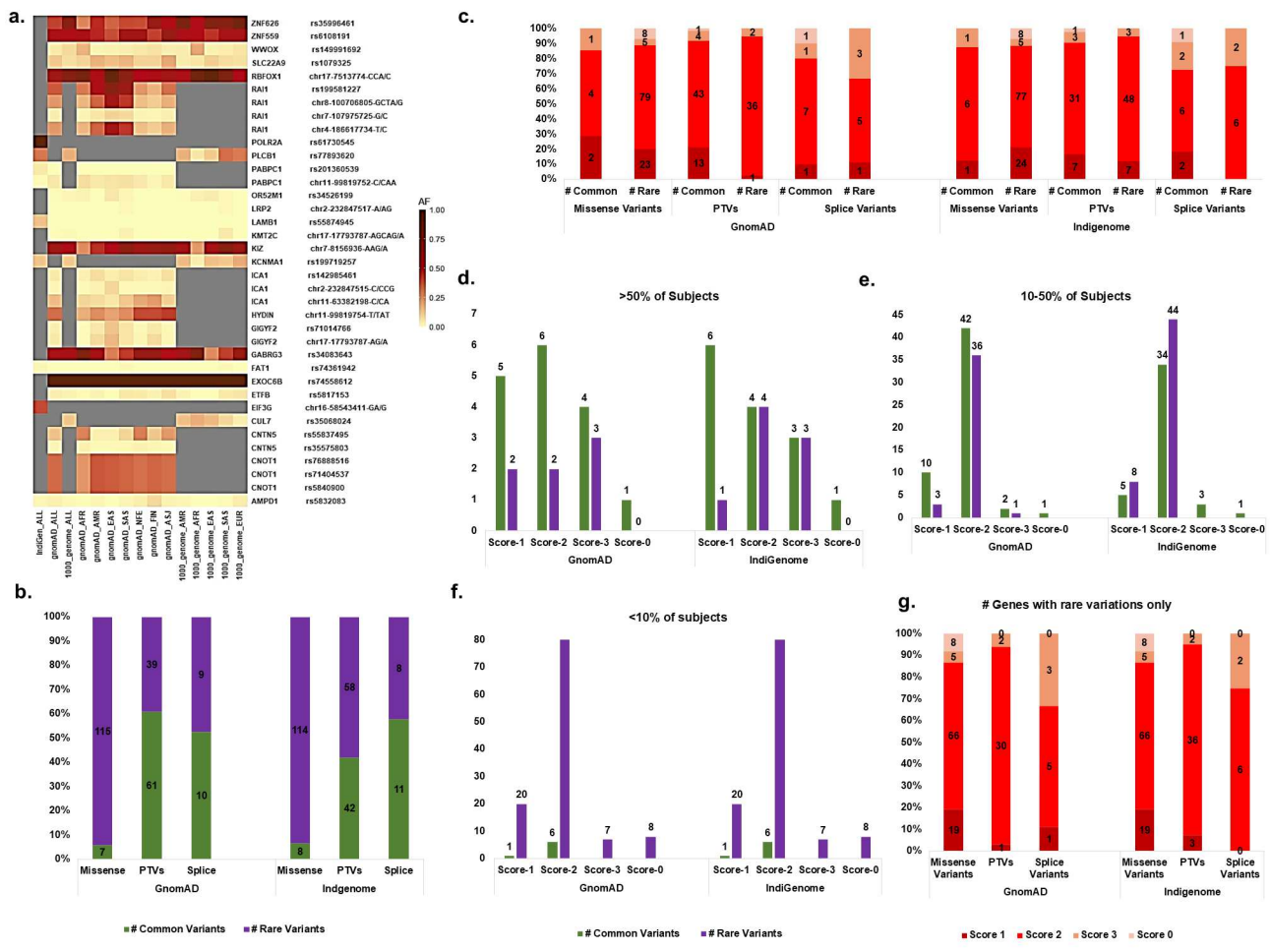
Fig. 1: Distribution of variants in 170 SFARI genes across four different SFARI scores in the 25 Indian ASD subjects (a) Represents the number of SFARI genes carrying variations in Indian ASD subjects with respect to the total number of SFARI genes according to gene scores. (b) Distribution of the total number of variants, unique variants and genes identified in all subjects. (c) Represents average number of variants and genes per ASD subject. (d) Distribution of different types of SNPs with respect to SFARI scores 1,2,3,0. A 100% stacked bar plot was created. (e) Represents the number of times the same variant was altered in different ASD subjects (represented as % of 25 subjects). (f) Differences in the distribution of common and rare variations with respect to GnomAD, Indigenome and GenomeAsia databases. Colour codes across the graphs remain constant, where missense variants, PTVs and splice variants are represented by navy blue, yellow and brown colours, respectively; different shades of red represent different SFARI scores; common and rare variants are shown using green and purple colours and blue bar indicates average variants/genes present in each ASD subject.

Fig. 2: Distribution patterns of common and rare variants in SFARI genes of Indian ASD subjects in GnomAD in comparison with Indigenome a) Heatmap of 37 variants whose MAF differred between GnomAD and Indigenome references. (b) Distribution of 241 common and rare variants with respect to their SNP types across GnomAD and Indigenome references. (c) Distribution of common and rare variants with respect to their SFARI scores, SNP types across GnomAD and Indigenome references. (d-f) Distribution pattern of common and rare variants with respect to the frequency of occurrence of these variations among 25 ASD subjects. (g) Distribution of the final list of SFARI genes carrying only rare variants according to Indigenome reference with respect to their SNP type and SFARI scores. Colour codes across the graphs remain constant, where different shades of red represent different SFARI scores and common and rare variants are shown using green and purple colours.

Fig. 3: Enrichment analysis of 101 nonSFARI genes identified in WES data of 25 Indian ASD subjects (a) Represents 23 nonSFARI and neurodevelopment (NDD) genes with neuronal functions showed the strongest evidence as candidate genes for ASD. (b) Enrichment analysis on 101 nonSFARI and NDD genes. Select enrichment terms with strongest neuronal roles are plotted against their combined score in this bar plot (c) PPI network for protein products of all

101 genes using all interaction resources on STRING.db identified 41 proteins that showed significant interaction (p value 3.57e-11) which clustered into 8 clusters represented using same colours.

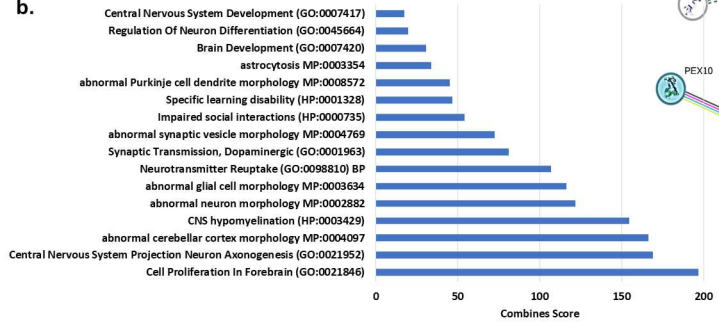

a.
Genes
100%
12.07%
100%
17.17%
100%
11.66%
100%
10.64%
Score-1
Score-2
Score-3
Score-0
Total SFARI genes
232
699
103
94
SFARI genes in 25 Indian ASD subjects
28
120
12
10
Total SFARI genes
SFARI genes in 25 Indian ASD subjects
b.
163
530
67
122
100
19
103
72
17
All Variants
Unique Variants
Unique Genes
c.
6.52
21.2
2.68
6.48
19
2.64
30.4
28.12
Avg # Variants per subject
Avg # Genes per subject
Avg per subject
d.
Genes
Score-1
Score-2
Score-3
Score-0
Missense Variants
PTVs
Splice variants
e.
Genes
> 95% of subjects
50%-95% of subjects
10%-50% of subjects
<10% of subjects
Score-1
Score-2
Score-3
Score-0
f.
Variants
gnomAD
Indgenome
GenomAsia
Total Common Variants
Total Rare Variants


a.
ZNF626 rs35996461
ZNF559 rs6108191
WWOX rs149991692
SLC22A9 rs1079325
RBFOX1 chr17-7513774-CCA/C
RAI1 rs199581227
RAI1 chr8-100706805-GCTA/G
RAI1 chr7-107975725-G/C
RAI1 chr4-186617734-T/C
POLR2A rs61730545
PLCB1 rs77893620
PABPC1 rs201360539
PABPC1 chr11-99819752-C/CAA
OR52M1 rs34526199
LRP2 chr2-232847517-A/AG
LAMB1 rs55874945
KMT2C chr17-17793787-AGCAG/A
KIZ chr7-8156936-AAG/A
KCNMA1 rs199719257
ICA1 rs142985461
ICA1 chr2-232847515-C/CCG
ICA1 chr11-63382198-C/CA
HYDIN chr11-99819754-T/TAT
GIGYF2 rs71014766
GIGYF2 chr17-17793787-AG/A
GABRG3 rs34083643
FAT1 rs74361942
EXOC5B rs74558612
ETFB rs5817153
EIF3G chr16-58543411-GA/G
CUL7 rs35068024
CNTN5 rs55837495
CNTN5 rs35575803
CNOT1 rs76888516
CNOT1 rs71404537
CNOT1 rs5840900
AMPD1 rs5832083
AF
1.00
0.75
0.50
0.25
0.00
IndiGen_ALL
gnomAD_ALL
1000_genome_ALL
gnomAD_AFR
gnomAD_AMR
gnomAD_EAS
gnomAD_SAS
gnomAD_NFE
gnomAD_FIN
gnomAD_ASJ
1000_genome_AMR
1000_genome_AFR
1000_genome_EAS
1000_genome_SAS
1000_genome_EUR
c.
Common
Rare
Missense Variants
PTVs
Splice Variants
GnomAD
Indigenome
d.
>50% of Subjects
Score-1
Score-2
Score-3
Score-0
GnomAD
IndiGenome
e.
10-50% of Subjects
f.
<10% of subjects
g.
Genes with rare variations only
Missense Variants
PTVs
Splice Variants
Common Variants
Rare Variants
Score 1
Score 2
Score 3
Score 0

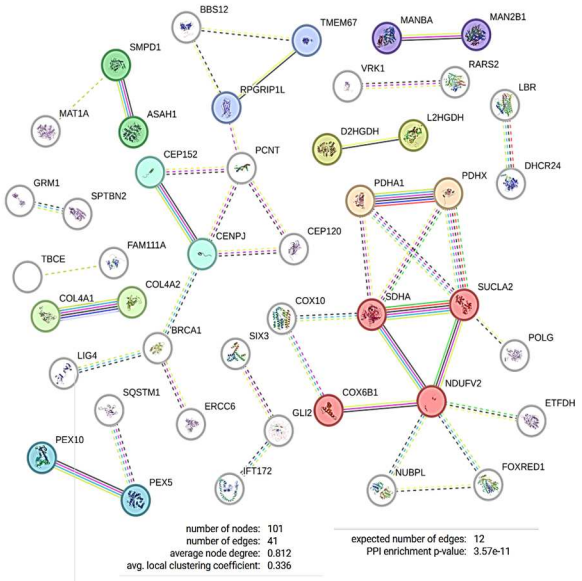
a.

Gene	SysNDD definitive	DBD Tier1 & AR	DDg2p definitive	Fu et al. 2022	Wang et al. 2022	post birth BrainSpan	pre birth BrainSpan	GTEx	SynGO v1.2
ARCN1	■			■		■		■	■
DNM1	■						■		
KIDINS220	■			■	■		■		
NALCN	■	■	■	■	■				
ZFHXA1	■			■					
PDHA1	■	■	■		■				
TUBA1A	■		■			■	■	■	■
COX6B1			■			■	■	■	■
APC2	■							■	■
GRM1	■							■	
SIX3	■		■					■	
SLC6A5	■		■					■	
TH	■		■					■	
FILIP1	■								■
MADD	■								■
PRSS12			■						■
ROGDI	■								■
SCN10A	■								■
SHROOM4	■								■
SPTBN2	■								■
SQSTM1		■							■
TIAM1	■								■
UBE3B	■	■	■						■

b.



c.



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.xlsx
- Tables.docx