Iran J Public Health, Vol. 54, No.5, May 2025, pp.1087-1097

Original Article

Genetic Investigation of Inherited Variants in a Multiplex Autism Spectrum Disorder (ASD) Family Using Whole-Genome Sequencing (WGS)

Mehdi Agha Gholizadeh [1], ***Farkhondeh Behjati*** [2], ****Masoud Garshasbi*** [1]

1. *Department of Medical Genetics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*
2. *Genetics Research Centre, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran*

***Corresponding Author:** Email: masoud.garshasbi@gmail.com

(Received 08 Apr 2024; accepted 15 Jun 2024)

Abstract

Background: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by early-onset challenges in social communication, repetitive behaviors, and clinical diversity. ASD is a highly heritable disorder, however, the exact mechanism by which inherited variants contribute to ASD in multiplex families, where more than one affected individual within a family is presented, remains unclear. We aimed to identify inherited genes in patients with ASD in a family with two affected siblings using Whole Genome Sequencing (WGS).

Methods: We performed WGS on two patients from a single family diagnosed with ASD. All of the patients were diagnosed with ASD using the gold-standard Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). We used various bioinformatics approaches to identify a list of prioritized candidate genes that may be associated with ASD or other neurodevelopmental disorders in this family.

Results: Our WGS analysis identified three potential candidate genes (*EVI5*:c.-82+866C>T, *RAPGEF1*:c.668C>T;p. Thr223Ile and *PDZD4*:c.-457G>A)) associated with ASD shared by the two patients. Additionally, utilizing various in-silico prediction tools and analysis of bioinformatics databases revealed that these rare variants are predicted to be deleterious and may contribute to ASDs. The identified variants are the first variants reported in ASD patients in the Iranian population that could be subjected to further validation studies.

Conclusion: These findings shed light on the genetic diversity of ASD within multiplex families and emphasize the complexity of genetic basis of ASD. Understanding the underlying genetic architecture of ASD is pivotal for advancing precise diagnostics and tailored therapeutic strategies.

Keywords: Autism spectrum disorder (ASD); Whole-genome sequencing (WGS); Multiplex families; Next generation sequence (NGS)

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by early-onset difficulties in social communication and interaction as well as restricted and repetitive

Copyright © 2025 Agha Gholizadeh et al. Published by Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.
(https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited
DOI: https://doi.org/10.18502/ijph.v54i5.18643

behaviors, interests, or activities (1). ASD is clinically diagnosed according to the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). The diagnosis of this condition is based on repeated and restricted behavior patterns as well as impairments in social communication and interaction (2). ASD is characterized by heterogeneity, encompassing a variety of genetic and environmental factors that contribute to its occurrence. As a result of its genetic component, ASD is considered to be a highly heritable condition, with estimates ranging from 64% to 91% (3). The etiology of ASD is therefore largely determined by genetic factors. To date, numerous genes and loci involving DNA copy number variations (CNVs) have been identified as contributors to susceptibility to ASD (4). Nonetheless, even the most common single-nucleotide variants (SNVs) such as *CHD8*, *ADNP*, *SCN2A*, *SHANK3*, and *NRXN1* or genomic copy number variations (CNVs), such as microdeletions or duplications, only account for a small fraction of ASD cases, approximately 3% (5). This observation highlights the highly diverse and heterogeneous nature of the genetic landscape underlying ASD. Recognizing individual genetic profiles can advance precision medicine and improve ASD understanding and management.

Whole-genome sequencing (WGS), as a state-of-the-art high-throughput technology, has significantly improved the accuracy of ASD diagnosis by approximately 20%. It has the potential to become the primary genetic testing method for neurodevelopmental disorders (6). WGS offers more consistent coverage within the genome coding regions compared to exome sequencing, enhancing ASD diagnosis accuracy by comprehensively analyzing the entire genome, making it a promising primary testing method for neurodevelopmental disorders like ASD (7). Nevertheless, the root cause of ASD within multiplex (MPX) families, where multiple family members are affected, remains a complex and inadequately understood aspect of this condition.

In this study, we performed WGS on two patients with ASD in a family. By leveraging an extensive analytical pipeline and employing bioinformatics analysis, which encompassed the detection of SNVs and CNVs across both coding and non-coding genomic regions, we identified potential three novel variants associated with ASD.

Materials and Methods

Ethical consideration

Informed consent was obtained from all subjects or their legal guardians. All clinical information and medical histories were collected at the Genetics Research Centre of the University of Social Welfare and Rehabilitation Sciences. The Ethics Committee of Tarbiat Modares University, Tehran, Iran (IR.MODARES.REC.1400.268) approved the study.

Subjects

We recruited four family members, consisting of two affected individuals and their parents, as illustrated in Fig. 1a. This consanguineous family originated from the Tabriz Province of Iran and was diagnosed with ASD. The patients were diagnosed with ASD according to the gold-standard Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Furthermore, an assessment using the autism behavior checklist was conducted, and the patient's score was calculated, exceeding the minimum cut-off of 68. The two patients underwent WGS and were assessed through comprehensive medical records, including general physical examinations, neuroimaging studies, karyotyping, and other conventional biochemical tests.

Whole-Genome Sequencing

DNA was extracted from the peripheral blood of all family members according to the manufacturer's instructions. We performed WGS to sequence the genomes of patients III-1 and III-2 using Illumina's Novaseq 6000 technology. To

outline the process briefly, DNA libraries were initially prepared by fragmenting the genomic DNA, followed by adapter ligation and subsequent PCR amplification. We used sequencing reads with an average length of 150 bp to ensure comprehensive coverage of the genome. The generated libraries were then subjected to paired-end sequencing, achieving an average read depth of 30X. The average depth of sequencing commonly falls within the range of 30-50X for WGS (8).

In-silico bioinformatics analysis

Variant Calling and Annotation

The raw sequencing reads underwent a thorough cleaning procedure, which entailed the removal of adapter sequences, as well as the elimination of reads containing low-quality bases covering more than 50% of their lengths or reads with unknown bases covering more than 10% of their lengths. Following this preprocessing, the cleaned reads were then aligned to the reference human genome (GRCh37/hg19) utilizing the Burrows-Wheeler Aligner (BWA) software (9). The alignment files were sorted for improved organization and efficiency. This sorting process was executed using SAMtools (10). For variant calling, we executed variant calling by employing the HaplotypeCaller (HC) tool using the Genome Analysis Toolkit (GATK) (11). This process encompassed the identification of single-nucleotide variants (SNVs) as well as insertions and deletions. To ensure the exclusion of low-quality variants, we implemented Variant Quality Score Recalibration (VQSR). After the variant filtering process, using the Variant Effect Predictor (VEP v107) (12) and Annovar (13), subsequent variants were annotated.

Variant Prioritization

To prioritize of variants, we focused on shared genotype between two patients, including homozygosity for recessive inheritance and heterozygosity for de novo dominant inheritance. Variants in coding exons with a minor allele frequency (MAF) lower than 1% in databases such as gnomAD (https://gnomad.broadinstitute.org), ExAC (Exome Aggregation Consortium) (http://exac.broadinstitute.org/), and the 1000 Genomes Project (http://www.1000genomes.org/) were included in our analysis. Moreover, the frequency of the exonic variants was checked in the Iranome as a local database (http://www.iranome.ir). Then, based on PHRED-like CADD score >10 (CADD tool v1.6) (14) encompassing the top 10% of potentially deleterious variants within the human genome, variants were prioritized. The selected threshold is based on established guidelines for identifying variants with higher pathogenicity (15). Conservation scores were then considered, with high evolutionary conservation indicating functional significance. Next, we evaluated the potential deleteriousness of missense variants through a comprehensive assessment using a variety of dbNSFP (16) including SIFT, PolyPhen V2-HDIV, PolyPhen V2-HVAR, LRT, MutationTaster, Mutation Assessor, FATHMM, MetaSVM, MetLR, PROVEAN, VEST3, and RI.

To detect sequence regulatory functions as well as potential enhancers and promoters, we used Sei (17). This deep-learning-based framework is proficient in predicting a compendium of 21,907 chromatin profiles spanning more than 1,300 cell lines and tissues. In addition, we conducted an assessment of epigenomic data and marks obtained from 127 cell lines courtesy of the NIH Roadmap Epigenomics Mapping Consortium, utilizing CADD v.1.6 (14). This utilizes epigenomic data from the NIH Roadmap Epigenomics Project, mapping the epigenome across tissues and cell types. The consortium provides comprehensive maps of active or silent genome regions, including data on DNA methylation, histone modification, chromatin accessibility, and RNA transcripts. Integrating this information helps assess the functional consequences of genetic variants, aiding in prioritizing variants with potential regulatory roles. Furthermore, we assessed the regions containing promoter and enhancer variants by chromHMM

(18) using the data from GM12878. Nine ENCODE cell lines that were analyzed by chromHMM did not encompass a neuronal cell line. However, most studies used the data from GM12878, as lymphoblastoid cell lines to evaluate gene expression profiling on mental disorders. Given that the nine we used data from GM12878 (Epstein-Barr Virus transformed lymphoblastoid cell line), as lymphoblastoid cell lines are used in many gene expression profiling studies in mental disorders.

All candidate variants were checked in HGMD (https://www.hgmd.cf.ac.uk/ac), SFARI (https://gene.sfari.org/), OMIM (https://omim.org), ClinGen (https://clinicalgenome.org/), Genereviews (https://www.ncbi.nlm.nih.gov/books/NBK1116/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and GeneCards (https://www.genecards.org/). This comprehensive assessment aimed to establish the relevance of these variants to autism, neurodevelopmental disorders, and the role of genes in neuronal activities.

Structural Variant Calling

Structural variants (SV) were generated using CNVnator (21) with 100 bin sizes. SV calls from two patients within each family were integrated using the R package intanSV (22), considering SVs calls that were shared between two patients with a reciprocal coordinate overlap of more than 10%. We exclude CNV calls that exhibit an overlap of more than 50% of their lengths with five or fewer CNV events in the DGV database considered common variants. Finally, we did not identify any CNVs that were associated with ASD and other neurodevelopmental disorders in the patients.

Results

WGS identified three variants in EVI5, RAPGEF1, and PDZD4

We performed WGS on two patients with an average depth of coverage of 30X. In total, we detected 4,919,776 and 5,017,460 variants in patients 1 (III-1) and 2 (III-2), respectively. Our objective was to identify a list of variants or genes that have a high likelihood of being responsible for the disease. To identify potentially deleterious mutations in both patients, we performed a filtering process as illustrated in Fig. 1b. Candidate variants were prioritized based on the American College of Medical Genetics and Genomics Association for Molecular Pathology (ACMG-AMP guidelines) (http://wintervar.wglab.org). Benign and likely benign variants were excluded from the study. Based on the family pedigree and consanguineous marriage, we implemented a recessive model (homozygous and compound heterozygous) to declare shared homozygous variants in patients. In addition, Considering the dominant model (heterozygous), no shared variants were identified in two patients. Consequently, we identified three homozygous variants, including an intron variant in *EVI5* (c.-82+866C>T), a missense variant in *RAPGEF1* (c.668C>T;p. Thr223Ile), and a single nucleotide variant (SNV) in *PDZD4* (c.-457G>A) (Table 1). Based on the ACMG guideline, these variants can be classified as variant of uncertain significance (VUS).

Table 1: Top variants prioritized following filtering process

Location	Gene	Type	Zy-gosity	CADD - PHRED score	ChromHMM*	Score	ACMG Classifica-tion
Chr1(hg19):g93249527 G>A NM_001350197.2:c.-82+866C>T	*EVI5*	In-tron	HOM	10.5	cHmmTssA	0.52	VUS
Chr9(hg19):g134505695 G>A NM_001377935.1:c.668C>T;p.(Thr223Ile)	*RAPGEF1*	Ex-on	HOM	24.7	NA	-	VUS
ChrX(hg19):g153096210 C>T NM_001303512.2:c.-457G>A	*PDZD4*	Pro-moter	HEM	12.79	cHmmReprPCWk	0.24	VUS

*Active transcription start sites (TssA), enhancers (Enh), bivalent TSS (TssBiv), bivalent enhancers (EnhBiv), genic enhancers (EnhG), flanking transcription states (TxFlnk), flanking bivalent TSS (TssBiv), active transcription flanking sites (TssAFlnk), transcription states (Tx) and weak transcription states (TxWk), repressed polycomb (ReprPC) and weak repressed polycomb regions (PeprPCWk), heterochromatin (Het) and quiescent regions (Quies). Scores closer to 1 indicate a higher proportion of cell types in the specified chromatin state. VUS; variant of uncertain significance

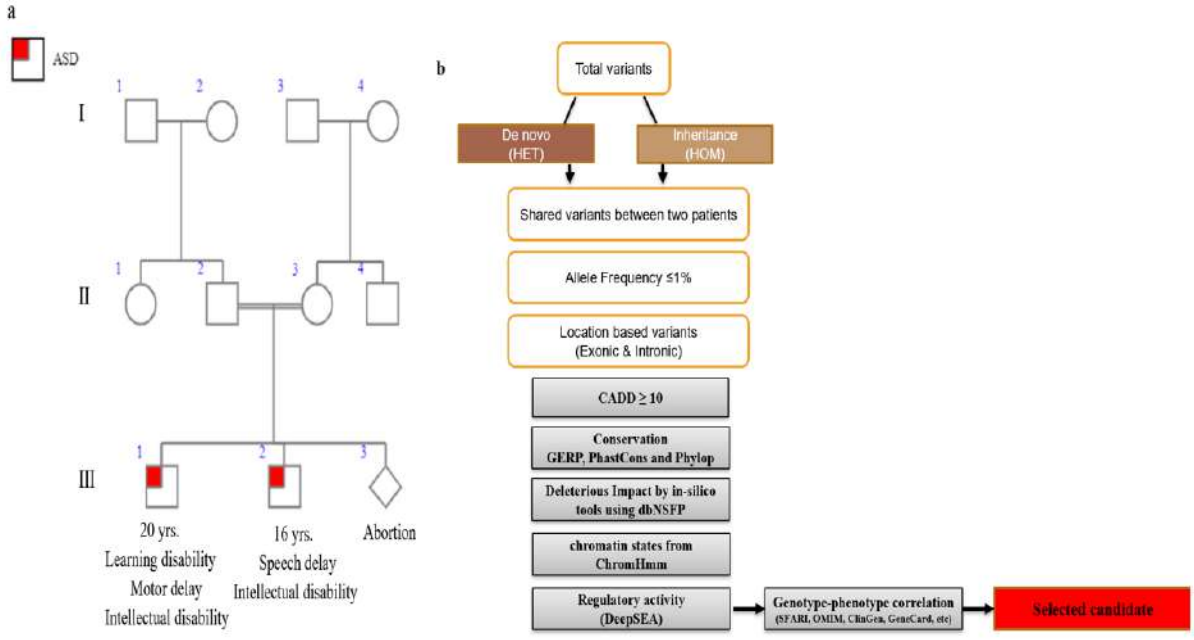

Fig. 1: **a** Pedigree of family study. **b** Schematic representation of filtering strategies exploited in this study.

Moreover, an assessment of conservation at both the nucleotide and protein levels through UCSC showed that the identified variants were located in a highly conserved region (Fig. 2a). Integrative Genomics Viewer (IGV) screenshot showed a homozygous state of the identified variants in two patients (Fig. 2b).

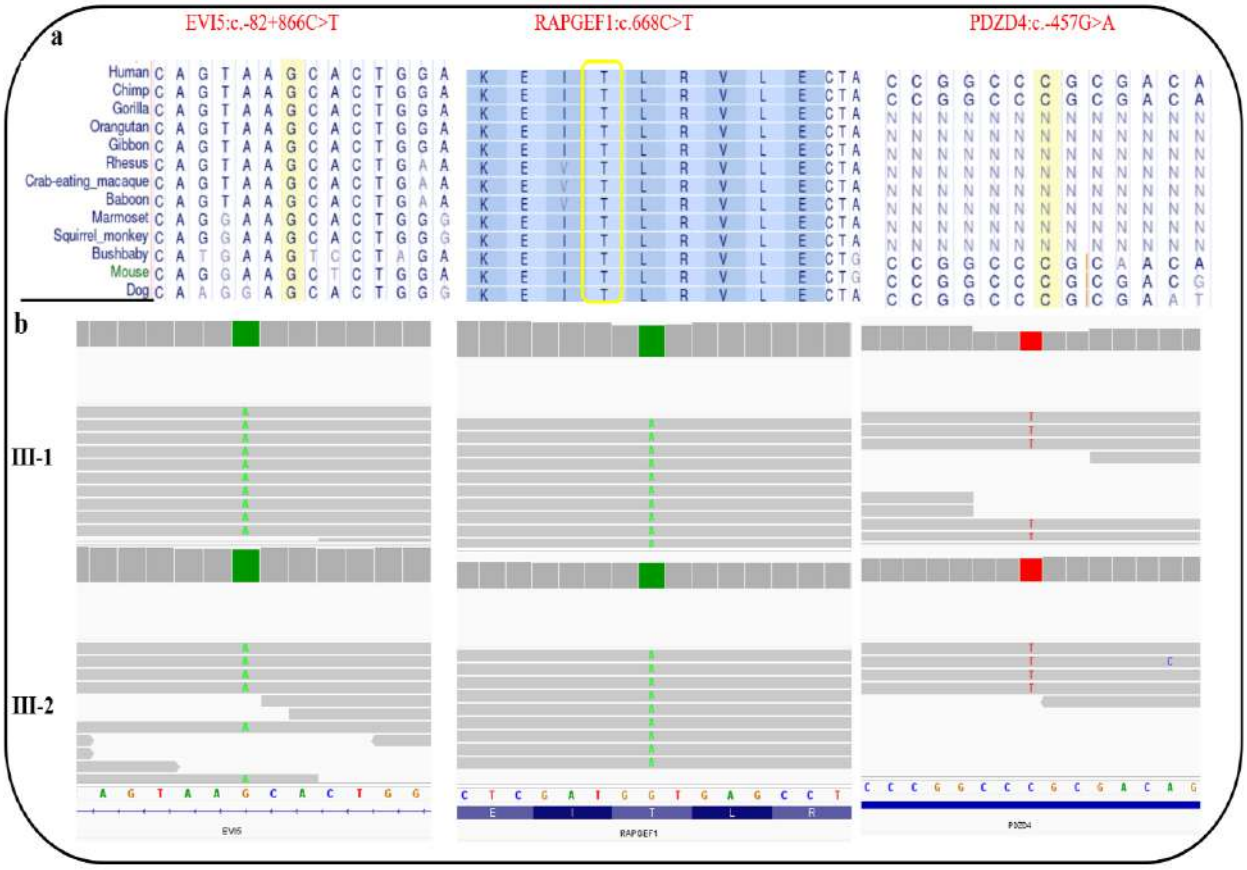

Fig. 2: **a** UCSC database demonstrates evolutionary conservation in nucleotide and protein levels of the variants site. **b** Integrative Genomics Viewer of the genome sequencing indicated a homozygous state of the variants in the patients (III-1, III-2)

In-silico analysis

The UCSC database showed that the intronic variant of *EVI5*:c.-82+866C>T was located within H3K4Me1 (enhancer/promoter-associated) and H3K4Me3 (promoter-associated) peaks and DNase I hypersensitivity site (Fig. 3a). However, upstream of the *PDZD4* variant (*PDZD4*:c.-457G>A) is located in heterochromatin and DNase I hypersensitivity site (Fig. 3b). The *EVI5*:c.-82+866C>T intronic variant is likely in an active regulatory genomic environment, indicated by its presence in H3K4Me1 and H3K4Me3 peaks and a DNase I hypersensitivity site. This suggests its potential impact on gene function and expression. Conversely, the *PDZD4*:c.-457G>A variant is located upstream in a region classified as heterochromatin, which is generally more condensed and less transcriptionally active. However, the presence of a DNase I hypersensitivity site near this variant suggests that the region, despite being heterochromatic, might still be capable of regulatory interaction.

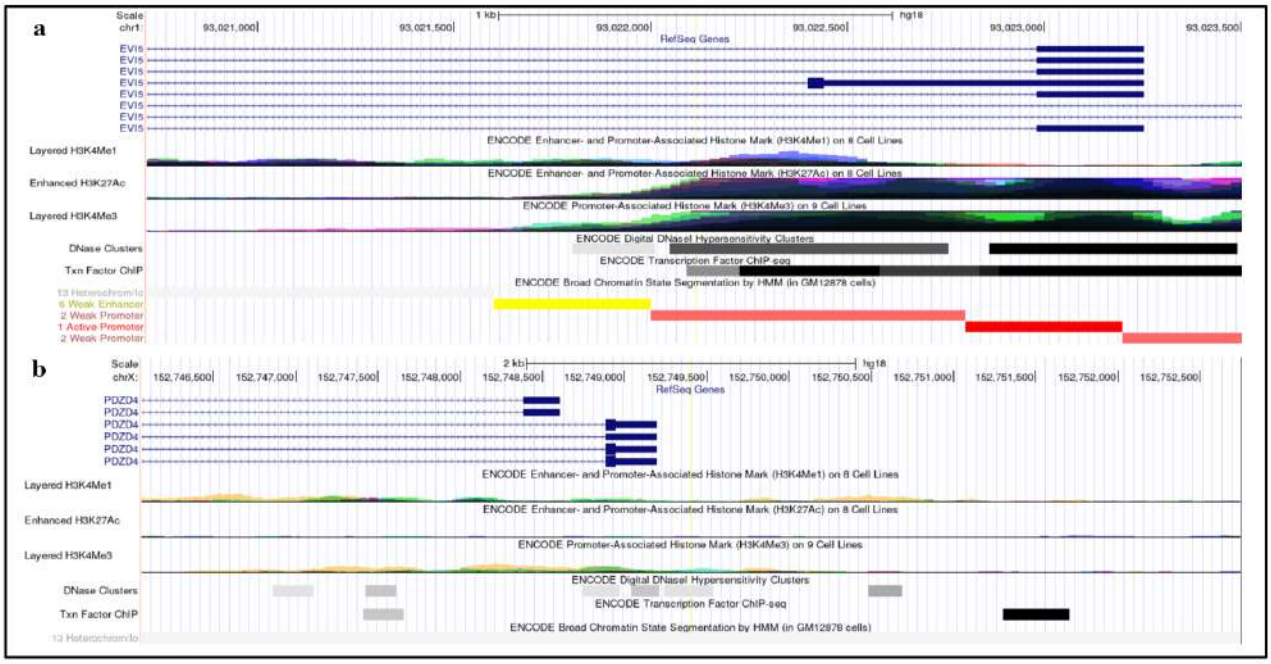

Fig. 3: UCSC genome browser predicted functionality of candidate non-coding variants in ENCODE lymphoblastoid cell lines. The yellow line shows the location of the non-coding variants shared by both patients. **a** *EVI5*:c.-82+866C>T variant predicted that located in the promoter. **b** *PDZD4*:c.-457G>A showed that located in heterochromatin

the protein structure and possible effects of the missense variant on *RAPGEF1* were analyzed by PyMol software (V.2.3.4), after building the PDB structure file based on the SWISS-PROT (20) (Fig. 4).

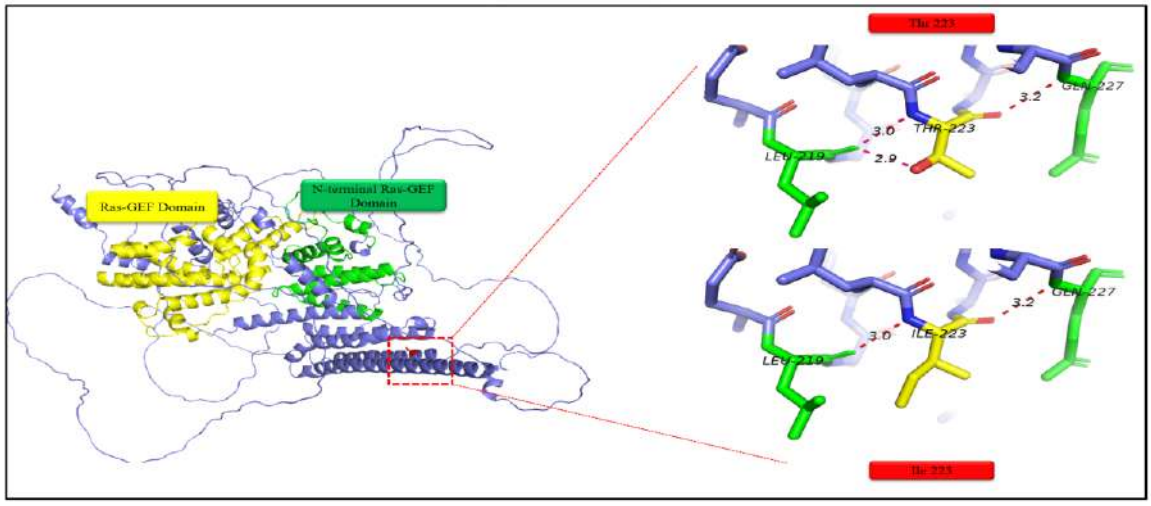

Fig. 4: The amino acid Threonine changes to Isoleucine at position 223. The picture was delineated by using PyMol (v:2.3.4). Substitution of Thr311 residue to Ile (yellow) reveals a decrease interaction between Ile 223 with Leu 223.

Discussion

In this study, we identified three candidate genes (*EVI5*, *RAPGEF1*, and *PDZD4*) in two patients in a multiplex ASD family using WGS. ASD is a complex neurodevelopmental condition characterized predominantly by restricted and repetitive behaviors, alongside impaired social interactions (23). Given the significant heritability of ASD, research on this disorder has placed a strong emphasis on uncovering the genetic factors contributing to its etiology (24). The significance of genetic factors in ASD is substantially supported by twin studies, particularly in identical twins (25). Despite this strong genetic component, common genetic variations, including single-nucleotide variants and CNVs from genes like *CHD8*, *ADNP*, *SCN2A*, *SHANK3*, and *NRXN1*, are responsible for only a small fraction of cases, underscoring the substantial heterogeneity and genetic diversity observed in ASD (5).

WGS has significantly improved the accuracy of ASD diagnosis by approximately 20%, positioning it as a transformative tool with the potential to become the primary genetic testing method for neurodevelopmental disorders (26). We employed a comprehensive WGS approach to identify specific genetic mutations that may be associated with susceptibility to ASDs in a single family. Following rigorous filtering, our investigation revealed three candidate genes (*EVI5*:c.-82+866C>T, *RAPGEF1*:c.668C>T; p.Thr223Ile, and *PDZD4*:c.-457G>A) that may increase susceptibility to ASDs in this family (Table 1).

The ecotropic viral integration site 5 (*EVI5*) gene regulates cell cycle progression and cytokines by stabilizing the F-Box gene product, and mutations in this gene are associated with multiple sclerosis (27). Two patients share a homozygous mutation (c.-82+866C>T) in intron 1 of *EVI5*. Our analysis confirmed that this variant is located in the promoter region (Fig. 3a) and a highly conserved area (Fig. 2a), suggesting its potential impact on gene function. This variant in the promoter region may influence gene expression by affecting transcription factor binding and, subsequently, the initiation of gene transcription. The location of this variant in a highly conserved region further suggests its functional significance, as evolutionary conservation often indicates critical roles in maintaining gene function. The studies have confirmed the association of *EVI5* with multiple sclerosis (28), and to date, the association of this gene with ASD or mental disorders has not been reported.

In addition, we identified a hemizygous variant located upstream of *PDZD4* (c.-457G>A) that was shared by both patients. This variant is located in the DNase I hypersensitivity site. DNase I hypersensitivity sites are regions of the genome that are more accessible to cleavage by DNase I enzyme, indicating an open chromatin structure and suggesting functional importance. The presence of the variant within a DNase I hypersensitivity site implies that this region may be involve in regulatory interactions and influence the binding of regulatory proteins, and thus impact gene expression. *PDZD4* encodes a protein that exhibits high expression levels in the brain and possesses ubiquitin-protein ligase activity. The association of this gene with ASD and neurodevelopmental disorders has been reported in a few studies. A rare variant of *PDZD4* was identified in patients with ASD (29). Furthermore, another study, by re-analyzing exome sequencing data, identified a variant in *PDZD4* in neurodevelopmental disease (30). Given the high expression of this protein in the brain, any disruption in its expression might impact brain function and contribute to neurodevelopmental disorders such as ASD. WGS of individuals with ASD in a Chinese population revealed an inherited hemizygous mutation in the *PDZD4* gene, which may be a predictor of ASD risk (31).

RAPGEF1 encodes a human guanine nucleotide exchange factor. It plays a role in signal trans-

duction by interacting with the SH3 domain of CRK and subsequently activating various members of the Ras family of GTPases. Our findings revealed a rare homozygous missense variant in exon 6 of *RAPGEF1* (NM_001377935.1:c.668C>T; p.(Thr223Ile)), shared by both patients with ASD. This variant has not been previously reported and was absent in the ExAC, gnomAD, and 1000G databases. Based on the ACMG guideline, this variant can be classified as a variant of uncertain significance (VUS). In silico prediction (Mutation Taster, SIFT, DANN, GenoCanyon, fitCons, and phred-scaled CADD>10) confirmed the deleterious effects of the identified variant in *RAPGEF1*. Loss-of-function intolerant (pLI) score (32) and intolerance to missense variation Z-score (33) in *RAPGEF1* gene are 6.28 and 3.13, respectively, indicating a high level of intolerance to loss-of-function variants.

Based on the tertiary structure analysis, the substitution of threonine with isoleucine at amino acid 223 of the *RAPGEF1* gene has the potential to decrease its interaction with Leu219 (Fig. 2). Threonine is a hydrophilic polar amino acid with a hydroxyl (-OH) group in its side chain, while isoleucine is a nonpolar amino acid with a hydrophobic aliphatic side chain. The drift in hydrophilic to hydrophobic can result in loss of hydrophilic interactions and negatively affect the interactions with associated molecules. Additionally, changes in the polarity of amino acids can disturbs correct folding and impair the protein's normal function.

RAPGEF1 is a gene prominently expressed in the brain. It plays a vital role in activating Rap1 downstream of the Reelin signaling pathway, which is crucial for neural development, particularly in processes like radial glial attachment and neuronal migration (34). Alterations in the dosage of *RAPGEF1* have been linked to cerebral palsy (35). Additionally, a missense variant in this gene was identified to be associated with neuropsychiatric symptoms in two siblings. Studies using a zebrafish model with a mutated rapgef1 gene have highlighted its role in brain and blood vessel development (36). Furthermore, mouse models have revealed that mice lacking *RAPGEF1* exhibit elevated levels of nuclear beta-catenin and increased proliferation of neuronal precursor cells in the cerebral cortex (37).

Overall, *EVI5* is involved in cytokinesis, cell cycle regulation, and immune response. *PDZD4* plays a role in neuronal development, synapse formation, and signaling. *RAPGEF1* regulates cell signaling and neuronal development. Existing literature suggests potential associations of these genes with ASD due to their roles in neurodevelopment, synaptic function, and signaling pathways implicated in ASD. Although our findings are limited to this specific family, the newly identified candidate genes can be important subjects for evaluation in future sequencing studies to determine their potential relationship with autism susceptibility. One limitation of our study is that we did not perform functional testing for the variants identified. The pathogenicity of the identified variants relies solely on predictions from in silico tools, although these tools are commonly employed in current sequencing analyses. Furthermore, further evidence is needed, especially from family-based WGS analyses involving more multiplex families with ASD, to strengthen our findings and expand our understanding of the genetic basis of ASD.

Conclusion

Using WGS, we identified novel variants of three potentially ASD-related genes (*EVI5*, *RAPGEF1*, and *PDZD4*) in a multiplex ASD family. These findings may be expanding our understanding of the clinical and genetic heterogeneity of genes contributing to ASD.

Journalism Ethics considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or

submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

The authors are grateful to patients and family members for collaboration in this study. We are also thankful to the staff of the DeNA laboratory, Tehran, Iran, for helping us in this research.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Association AP (2013). American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Arlington. https://repository.poltekkes-kaltim.ac.id/657/1/Diagnostic%20and%20statistical%20manual%20of%20mental%20disorders%20_%20DSM-5%20(%20PDFDrive.com%20).pdf
2. American Psychiatric Association, DSM-5 Task Force (2013). Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). American Psychiatric Publishing, Inc.
3. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry,* 57 (5):585-595.
4. Yuen RKC, Merico D, Bookman M, et al (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci,* 20 (4):602-611.
5. Ramaswami G, Geschwind DH (2018). Genetics of autism spectrum disorder. *Handb Clin Neurol,* 147:321-329.
6. Fernandez BA, Scherer SW (2017). Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. *Dialogues Clin Neurosci,* 19 (4):353-371.
7. Jiang YH, Yuen RK, Jin X, et al (2013). Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet,* 93 (2):249-263.
8. Sun Y, Liu F, Fan C, et al (2021). Characterizing sensitivity and coverage of clinical WGS as a diagnostic test for genetic disorders. *BMC Med Genomics,* 14:102.
9. Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics,* 25 (14):1754-1760.
10. Li H, Handsaker B, Wysoker A, et al (2009). The sequence alignment/map format and SAMtools. *Bioinformatics,* 25 (16):2078-2079.
11. McKenna A, Hanna M, Banks E, et al (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20 (9):1297-1303.
12. McLaren W, Pritchard B, Rios D, et al (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26 (16):2069-2070.
13. Wang K, Li M, Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38 (16):e164.
14. Kircher M, Witten DM, Jain P, et al (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet,* 46 (3):310-315.
15. Rentzsch P, Witten D, Cooper GM, et al (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res,* 47 (D1):D886-D894.
16. Liu X, Wu C, Li C, Boerwinkle E (2016). dbNSFP v3. 0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum Mutat,* 37 (3):235-241.
17. Chen KM, Wong AK, Troyanskaya OG, Zhou J (2022). A sequence-based global map of regulatory activity for deciphering human genetics. *Nat Genet,* 54 (7):940-949.
18. Ernst J, Kheradpour P, Mikkelsen TS, et al (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473 (7345):43-49.
19. Karolchik D, Baertsch R, Diekhans M, et al (2003). The UCSC genome browser database. *Nucleic Acids Res,* 31 (1):51-54.

20. Bairoch A, Boeckmann B (1992). The SWISS-PROT protein sequence data bank. *Nucleic Acids Res,* 20 Suppl(Suppl):2019-22.
21. Abyzov A, Urban AE, Snyder M, Gerstein M (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res,* 21 (6):974-984.
22. Jia L, Liu N, Huang F, et al (2020). intansv: an R package for integrative analysis of structural variations. *PeerJ,* 8:e8867.
23. Battle DE (2013) Diagnostic and Statistical Manual of Mental Disorders (DSM). *Codas*, 25(2):191-2.
24. Chaste P, Leboyer M (2012). Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues Clin Neurosci*, 14(3):281-92.
25. Hallmayer J, Cleveland S, Torres A, et al (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry,* 68 (11):1095-1102.
26. Trost B, Walker S, Haider SA, et al (2019). Impact of DNA source on genetic variant detection from human whole-genome sequencing data. *J Med Genet*, 56(12):809-817.
27. Mazdeh M, Rahimi M, Eftekharian MM, et al (2017). Ecotropic Viral Integration Site 5 (EVI5) expression analysis in multiple sclerosis patients. *Hum Antibodies,* 26 (3):113-119.
28. Razavian T, Shakib ME, Gharagozli K, et al (2020). Association of rs12487066, rs12044852, rs10735781, rs3135388, rs6897932, rs1321172, rs10492972, and rs9657904 polymorphisms with multiple sclerosis in Iranian population. *Oman Med J,* 35 (4):e150.
29. Piton A, Gauthier J, Hamdan F, et al (2011). Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry,* 16 (8):867-880.
30. Giovenino C, Trajkova S, Pavinato L, et al (2023). Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation For X-linked genes. *Eur J Hum Genet*, 31(11):1228-1236.
31. Wu J, Yu P, Jin X, et al (2018). Genomic landscapes of Chinese sporadic autism spectrum disorders revealed by whole-genome sequencing. *J Genet Genomics,* 45 (10):527-538.
32. Lek M, Karczewski KJ, Minikel EV, et al (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536 (7616):285-291.
33. Samocha KE, Robinson EB, Sanders SJ, et al (2014). A framework for the interpretation of de novo mutation in human disease. *Nat Genet,* 46 (9):944-950.
34. Voss AK, Britto JM, Dixon MP, et al (2008). C3G regulates cortical neuron migration, preplate splitting and radial glial cell attachment. *Development*, 135(12):2139-49.
35. Oskoui M, Gazzellone MJ, Thiruvahindrapuram B, et al (2015). Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun,* 6:7949.
36. Li N, Zhou P, Yang M, et al (2021). Zebrafish modeling mimics developmental phenotype of patients with RAPGEF1 mutation. *Clin Genet,* 100 (2):144-155.
37. Voss AK, Krebs DL, Thomas T (2006). C3G regulates the size of the cerebral cortex neural precursor population. *EMBO J,* 25(15):3652-3663.