scientific reports

Check for updates

OPEN

An integrative scoring approach for prioritization of rare autism spectrum disorder candidate variants from whole exome sequencing data

Apurba Shil[1,2,3], Noa Arava[2], Noam Levi[1,2,3], Liron Levine[4], Hava Golan[2,3,5], Gal Meiri[2,6], Analya Michaelovski[2,7], Yair Tsadaka[2,8], Adi Aran[9,10] & Idan Menashe[1,2,3]✉

Discerning clinically relevant autism spectrum disorder (ASD) candidate variants from whole-exome sequencing (WES) data is complex, time-consuming, and labor-intensive. To this end, we developed *AutScore*, an integrative prioritization algorithm of ASD candidate variants from WES data and assessed its performance to detect clinically relevant variants. We studied WES data from 581 ASD probands, and their parents registered in the Azrieli National Center database for Autism and Neurodevelopment Research. We focused on rare allele frequency (< 1%) and high-quality proband-specific variants affecting genes associated with ASD or other neurodevelopmental disorders (NDDs). We developed *AutScore* and *AutScore.r* and assigned each variant based on their pathogenicity, clinical relevance, gene-disease association, and inheritance patterns. Finally, we compared the performance of both *AutScore* versions with the rating of clinical experts and the NDD variant prioritization algorithm, *AutoCaSc*. Overall, 1161 rare variants distributed in 687 genes in 441 ASD probands were evaluated by *AutScore* with scores ranging from − 4 to 25, with a mean ± SD of 5.89 ± 4.18. *AutScore.r* cut-off of ≥ 0.335 performs better than *AutoCaSc* and *AutScore* in detecting clinically relevant ASD variants, with a detection accuracy rate of 85% and an overall diagnostic yield of 10.3%. Five variants with *AutScore.r* of ≥ 0.335 were distributed in five novel ASD candidate genes. *AutScore.r* is an effective automated ranking system for ASD candidate variants that could be implemented in ASD clinical genetics pipelines.

Keywords *AutScore*, *AutScore.r*, Candidate variants, ASD, WES, Prioritization algorithm

Abbreviations

ASD	Autism Spectrum Disorder
SNVs	Single Nucleotide Variants
indels	Insertions/Deletions
LGD	Likely Gene Disrupting
LP/P/VUS	Likely Pathogenic/Pathogenic/Variants of Uncertain Significance
LoF	Loss of Function
CNVs	Copy Number Variants
WES	Whole Exome Sequencing

[1]Department of Epidemiology, Biostatistics and Community Health Sciences, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel. [2]Azrieli National Centre for Autism and Neurodevelopment Research, Ben-Gurion University of the Negev, Beer Sheva, Israel. [3]Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer Sheva, Israel. [4]Bioinformatics Core Facility, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [5]Department of Physiology and Cell Biology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel. [6]Preschool Psychiatric Unit, Soroka University Medical Center, Beer Sheva, Israel. [7]Child Development Center, Soroka University Medical Center, Beer Sheva, Israel. [8]Child Development Center, Ministry of Health, Be'er Sheva 84100, Israel. [9]Neuropediatric Unit, Shaare Zedek Medical Center, Jerusalem, Israel. [10]Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. ✉email: idanmen@bgu.ac.il

WGS	Whole Genome Sequencing
ACMG/AMP	American College of Medical Genetics and Genomics/Association of Molecular Pathology
GATK	Genome Analysis Toolkit
IQR	Interquartile Range
NDDs	Neurodevelopmental Disorders
PPV	Positive Predictive Value
NPV	Negative Predictive Value
SFARI	Simons Foundation Autism Research Initiative
OMIM	Online Mendelian Inheritance in Man
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic

Recent advances in high-throughput sequencing technologies have revolutionized genetic studies of complex diseases[1–7]. The emergence of next-generation sequencing (NGS) platforms has enabled genomic analyses at an unprecedented scale and resolution. These technologies have facilitated whole-genome sequencing (WGS) and whole-exome sequencing (WES) of large cohorts, unveiling novel disease-associated loci and providing deeper insights into the genetic architecture of complex disorders[1–9].

Detecting disease-causing variants from WES/WGS data is a complex task. Today, most clinical genetics labs that analyze WES/WGS data follow the American College of Medical Genetics and Genomics (ACMG) guidelines for interpreting sequence variants[10]. This mainly includes detecting high-quality variants with lower allele frequency and damaging effects on the protein function. Other factors usually considered are the segregation of the variant with the phenotype and existing evidence for the variant or gene association with disease. To assist clinicians in this laborious process, several automated tools such as Exomiser[11], AMELIE[12], LIRICAL[13], AutoCaSc[14], etc., have been devised to prioritize disease-specific variants (mainly single nucleotide variants [SNVs] and insertions/deletions [indels]) from WES/WGS data.

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder (NDD) that has greatly benefited from the emergence of NGS technologies. Recent large-scale WES and WGS studies have identified thousands of ASD susceptibility genetic variants in hundreds of genes[5,15–20]. Nevertheless, despite these advances in ASD genetics, clinically meaningful genetic variants are identified only in 8–30% of affected probands[5,21,22]. Thus, there is a need for new approaches to facilitate the detection of ASD-specific variants from WES/WGS data.

Here, we present two scoring approaches called *AutScore* and *AutScore.r* that integrate variant and gene-level information such as pathogenicity, deleteriousness, clinical relevance, gene-disease association, and gene-variant inheritance pattern from a wide range of bioinformatics tools and databases to generate a single score for prioritizing clinically relevant ASD candidate variants from WES data for simplex and multiplex families. We applied both versions of *AutScore* to WES data from 581 Israeli ASD-affected probands and their parents. We assessed its performance by comparing the obtained results to a manual and blinded evaluation of the variants by clinicians and to *AutoCaSc*[14], an existing variant prioritization tool for NDDs.

Materials and methods

Study sample

Our sample included 581 children diagnosed with ASD, registered with the Azrieli National Centre for Autism and Neurodevelopment Research (ANCAN)[23,24]. Based on clinical records, none of the parents had registered themselves with ASD, intellectual disability, or other NDDs. Genomic DNA was extracted from saliva samples from children and their parents using Oragene®•DNA (OG-500/575) collection kits (DNA Genotek, Canada).

Whole exome sequencing (WES)

Whole Exome Sequencing (WES) analysis was conducted in two labs: (1) the Broad Institute as a part of the Autism Sequencing Consortium (ASC) project[17] and (2) the Clalit Health Services sequencing lab at Beilinson Hospital. WES was performed using Illumina HiSeq sequencers in both places, followed by the Illumina Nextera exome capture kit. The sequencing reads were aligned to human genome build 38 and aggregated into BAM/CRAM files. Then, the Genome Analysis Toolkit (GATK)[25] (Broad) or Illumina's DRAGEN pipeline[26] (Beilinson) was used for variant discovery and the generation of joint variant calling format (vcf) files.

Variant filtering and annotations

The multi-sample vcf files generated by the Genome Analysis Toolkit (GATK) and the DRAGEN platform were undertaken with identical procedures for variant filtering and annotation, as previously detailed[27]. Subsequently, we identified pathogenic (P), likely pathogenic (LP), or likely gene-disrupting (LGD) variants using the *InterVar*[28] tool in conjunction with our proprietary tool, *Psi-Variant*[27]. We kept only those LP/P/LGD variants that affected genes associated with ASD or other NDDs according to the SFARI gene[29] or the DisGeNET[30] databases for downstream analyses. Subsequently, 1161 candidate variants in 441 probands remained for further analysis (Fig. 1).

Prioritization of ASD candidate variants

We developed a metric called *AutScore* that integrates data from diverse bioinformatics tools to prioritize the detected list of ASD candidate variants as follows:

$$AutScore = \mathrm{I} + \mathrm{P} + \mathrm{D} + \mathrm{S} + \mathrm{G} + \mathrm{C} + \mathrm{H}$$

Where:

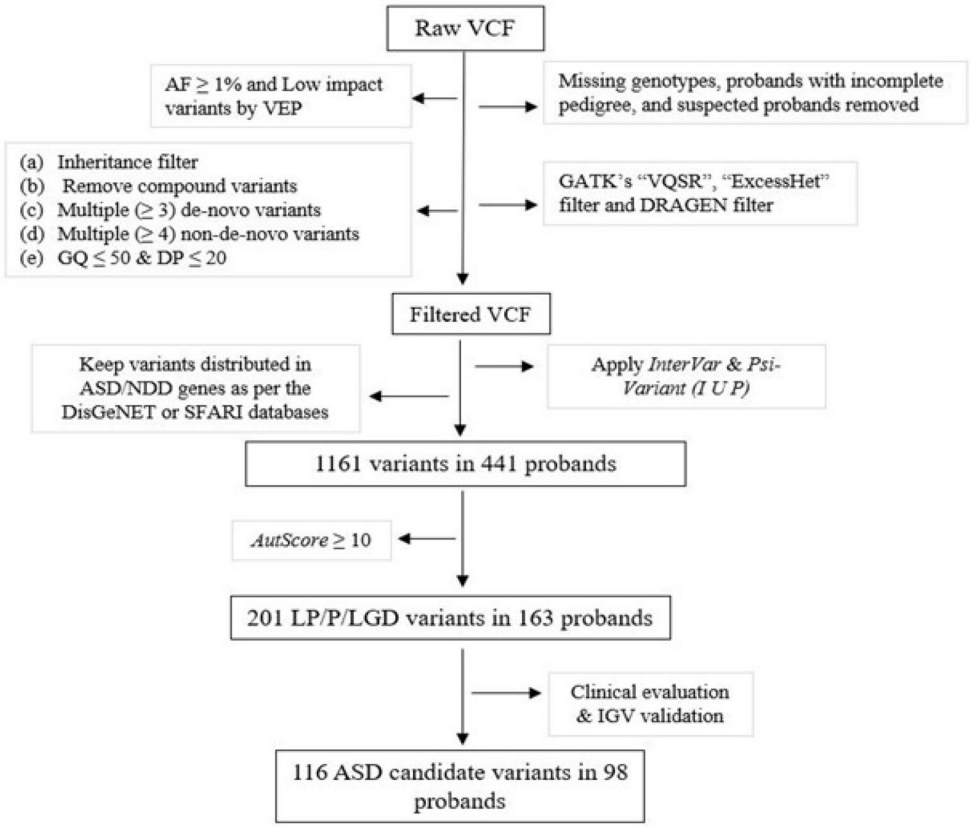

Fig. 1. Analysis workflow for detecting ASD candidate variants from the WES data. The final list of ASD candidate variants ($n = 116$) consists of 65 'likely' and 51 'possibly' ASD candidates.

- I – indicates the pathogenicity of a variant based on *InterVar*[28] classification as follows: 'benign' = -3; 'likely benign' = -1; 'variants of uncertain significance (VUS)' = 0; 'likely pathogenic' = 3, and 'pathogenic' = 6.
- P – cumulatively assess the deleteriousness of a variant based on the following six in-silico tools (SIFT[31] (< 0.05), PolyPhen-2[32] (≥ 0.15), CADD[33] (> 20), REVEL[34] (> 0.50), M_CAP[35] (> 0.025) and MPC[36] (≥ 2)). For each of these tools, a variant gets a score of 1 (deleterious) or 0 (benign), and these scores are aggregated to generate a single score ranging from 1 to 6.
- D – indicates the agreement of variant-phenotype segregation with the predicted segregation by the Domino tool[37] where agreement with Domino's 'very likely dominant/recessive' classes = 2; agreement with Domino's 'likely dominant/recessive' classes = 1; disagreement with Domino's 'very likely dominant/recessive' classes = -2; disagreement with Domino's 'likely dominant/recessive' classes = -1; and 0 were assigned for variants with Domino's 'either dominant or recessive' segregation.
- S – indicated the strength of association of the affected gene with ASD according to the SFARI gene database[29] where 'high confidence' = 3; 'strong candidate' = 2; 'suggestive evidence' = 1; and not in SFARI database = 0.
- G – indicated the strength of association of the affected gene with ASD according to the DisGeNET database[30] where weak/no association (GDA = 0 to 0.25) = 0: mild association (GDA = 0.25 to 0.50) = 1: moderate association (GDA = 0.50–0.75) = 2: strong association (GDA = 0.75 and above) = 3.
- C – pathogenicity of a variant based on ClinVar[38] where 'benign' = -3; 'likely benign' = -1; 'VUS' or not in ClinVar = 0; 'Likely pathogenic' = 1; 'Pathogenic' = 3.
- H – segregation of variants in the family weighted as (n^2)*-1* where n = number of probands in a family that carries the detected variants.

Clinical genetics validation

Variants with *AutScore* ≥ 10 (top quartile of candidate variants scores) were visually validated using the IGV software[39] and then manually examined by two clinical geneticists according to the standard ACMG/AMP guidelines[10]. The clinical genetic experts had ample experience in evaluating the clinical significance of different genetic tests, including WES of ASD triplets, in clinical genetics labs. They used their clinical experience to assess the clinical significance of the candidate variants regarding the ASD phenotype of the child and consequently assigned each variant one of the following rankings: 'Likely,' 'Possibly,' and 'Unlikely'. Importantly, each of the clinical experts independently assigned the ranking to the candidate variants without knowing *AutScore* values and the criteria used to determine these scores. Then, they compared their ranking and together reached a consensus ranking for each variant.

Refinement of *AutScore* weighting

To reduce the subjectivity of weights assigned to the different tools used by *AutScore*, we fitted a generalized linear model to these data with the different *AutScore* modules (i.e. I, P, D, S, G, C, H) being used as predictors and the clinical genetic ranking as the outcome (likely = 1 and possibly or unlikely = 0). For further clarity, we call this revised version of *AutScore* as "*AutScore.r*" throughout the manuscript. Details about the probabilistic weights (beta coefficients) and other statistics (standard error, p-value) that were used to develop *AutScore.r* can be found in Supplementary Table **S1**.

Statistical analysis

We used a Receiver Operating Characteristic (ROC) analysis to assess the performance of both *AutScore* versions in detecting ASD candidate variants using the clinical experts' rankings as the reference. We then used Yuden J's statistics[40] to identify the optimal cutoff in each *AutScore* version. Then, we used this cutoff to compute the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for the two *AutScore* versions in comparison to the clinical experts' ranking. The diagnostic yield (%) of each tool was also computed as the proportion of the number of ASD probands that have at least one ASD candidate variant out of the total affected ASD probands that completed their WES analysis. Finally, we compared the performance of the two *AutScore* versions in detecting ASD candidate variants as well as with the performance of *AutoCaSc*[14], an existing variant prioritization tool for NDDs.

Software

Data storage, management, and analyses were conducted in a high-performing Linux cluster using Python version 3.5 and R version 1.1.456. All statistical analyses and data visualization were performed and incorporated into R.

Results

A total of 1161 variants distributed in 687 genes in 441 ASD probands were evaluated by the *AutScore* algorithm. Variant's scores ranged from −4 to 25, with a mean ± SD of 5.89 ± 4.18 (Fig. 2). The clinical experts examined 201 (17.31%) variants with an *AutScore* of ≥ 10. Among these, 24 (11.9%) were found to be false positive indels in the visual assessment using the IGV software and thus removed from subsequent analyses. Of the remaining 177 variants, 65 (36.7%) were ranked as 'likely,' 51 (28.8%) as 'possibly,' and 61 (34.5%) as 'unlikely' ASD candidate variants (Supplementary Table **S2**).

Determining an optimal cutoff for *AutScore*

The ROC analyses of the two *AutScore* versions are depicted in Fig. 3. The revised version, *AutScore.r* had a slightly better performance in detecting "likely" ASD variants compared to the original version (AUC = 0.90, 95% CI = 0.84–0.95 vs. AUC = 0.84, 95% CI = 0.78–0.91 respectively). Yuden J's statistics suggested an optimal *AutScore* cutoff of ≥ 12 (Yuden J = 0.53) and a cutoff of ≥ 0.335 for *AutScore.r* (Yuden J = 0.69).

Comparing the performance between *AutScore*, *AutScore.r* and *AutoCasC* in detecting ASD candidate variants

Next, we used the suggested Yuden J's statistics cutoff of the two *AutScore* versions to compare their performance in detecting ASD candidate variants vis-à-vis the existing NDD prioritization tool, *AutoCaSc*, using its recommended cut-off of > 6[14]. The results of this comparison are presented in Fig. 4; Table 1. A moderate and statistically significant correlation was seen between the *AutScore* and *AutoCaSc* scores ($r = 0.58$ and $r = 0.55$ for *AutoScore and AutoScore.r* respectively, $p < 0.01$). All the three tools were highly sensitive in detecting ASD variants using their recommended cut-off (*AutScore* = 0.91, *AutScore.r* = 0.81 and *AutoCasC* = 0.92, respectively) and achieved a similar diagnostic yield of 9–10%. However, both *AutScore* and *AutScore.r* had better specificity, PPV and accuracy than *AutoCasC* with *AutScore.r* having a slightly better performance than *AutScore* in these parameters (Specificity: 0.87, 0.62 and 0.13; PPV: 0.79, 0.58 and 0.40; and Accuracy: 0.85, 0.72 and 0.43 for *AutScore.r*, *AutScore* and *AutoCasC* respectively, Table 1). The variant list ($n = 177$) with their respective scores from these three tools as well as their clinical assessment ranking is provided in Supplementary Table **S2**.

Characteristics of the variants detected by *AutScore.r*

Finally, we examined the characteristics of the 67 genetic variants detected by the best-performing tool – *AutScore.r using a cutoff of* ≥0.335 (Table 2). Of these variants, 53, 10, and 4 variants were ranked as 'likely', 'possibly', and 'unlikely' ASD candidate variants, respectively, by the clinical experts. Most of the detected variants (92.5%) were distributed in genes associated with ASD according to the SFARI Gene database[29] with most of them considered as high-confidence ASD genes (i.e., SFARI score of 1). Another five variants were distributed in 5 genes not listed in the SFARI database and thus could be considered as novel ASD candidate genes. Around 80% of the detected variants were classified as LP/P according to the ACMG/AMP variant interpretation criteria[10], and more than 73% were denovo variants.

Examination of *AutScore* performance in an ethnically diverse population

AutScore was developed based on WES data from an ASD cohort within the Israeli population. Thus, its performance parameters presented here could be specific to this population. To address this issue, we applied *AutScore.r* to WGS data from the BARAKA-Qatar study[41] of 372 individuals from 100 local families. Of the 10 dominant SNVs reported in Table 2 of their paper, 7 variants (70%) had an *AutScore.r* score of ≥ 0.335 (Supplementary Table **S3**). In addition, only one of the eight reported missense variants and none of the 5 x-linked variants had an *AutScore.r* score of ≥ 0.335. Of note, all the recessive and x-linked variants reported in the Qatar

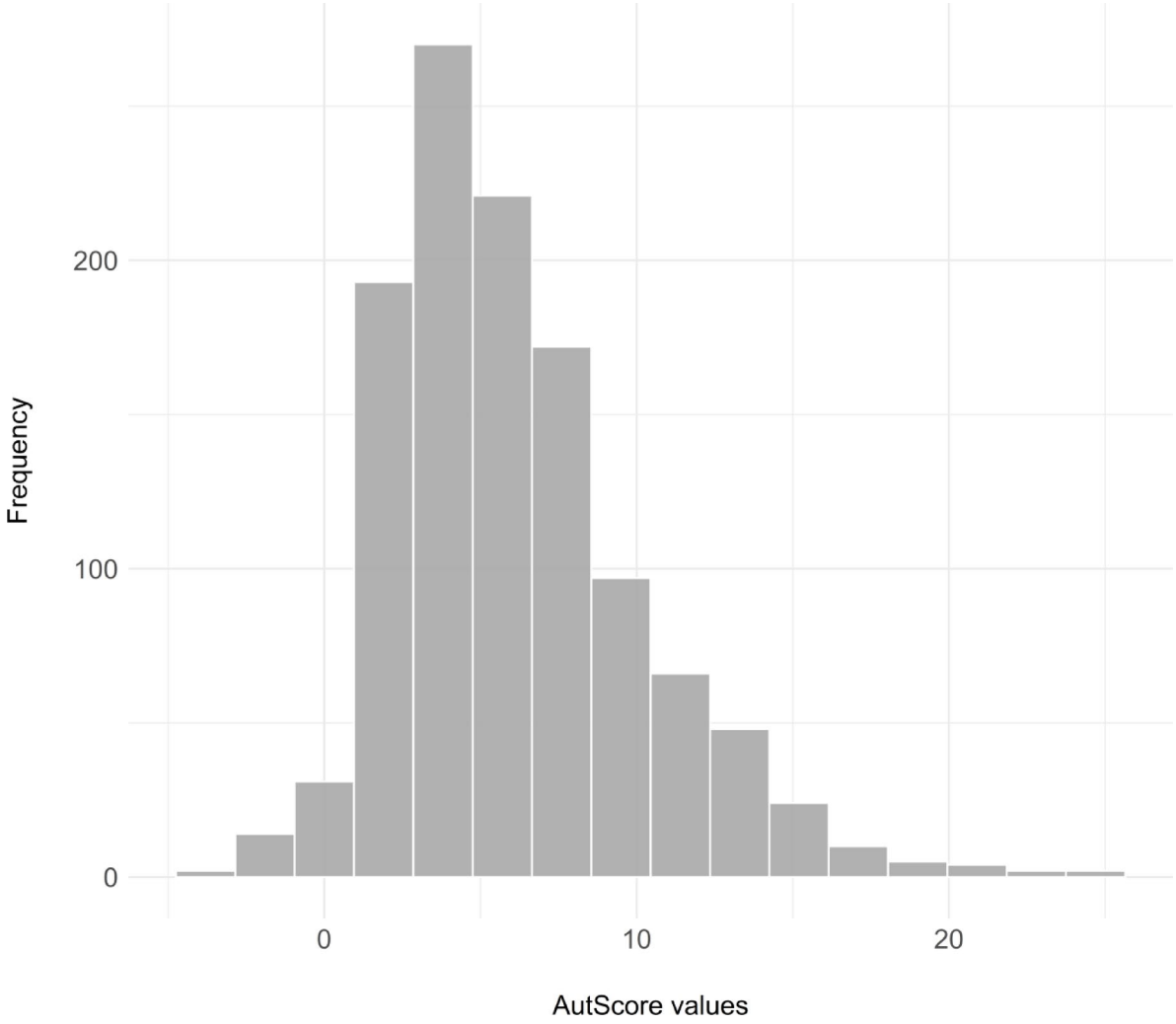

Fig. 2. Histogram depicting the distribution of *AutScore* ranking of the 1161 ASD candidate genetic variants.

study except the one detected by *AutScore.r*, were classified as VUS by the ACMG criteria (Supplementary Table **S3**).

Discussion

Discerning clinically relevant ASD candidate variants from many variants detected in WES analyses pose a formidable challenge for clinical experts, demanding considerable time and effort. Here, we present *AutScore*, a novel bioinformatics prioritization tool that integrates variant and gene-level information to prioritize ASD candidate variants derived from WES data. *AutScore* can be integrated into an existing bioinformatic pipeline for WES data analysis by pre-installing the ACMG/AMP[10] variant interpretation tool InterVar[14] and our in-house tool *Psi-Variant*[27]. Although *AutScore* was initially designed to assess the ASD clinical relevance of rare autosomal SNVs, it can be adapted for analyses of mitochondrial variants, and common heritable variants that are expected to enhance its applicability further. Finally, it can be continuously improved by adjusting the weights of its different modules using regression analyses as demonstrated in its revised *AutScore.r* version in this study.

Our results indicated that both the original and revised versions of *AutScore* are highly efficient in detecting clinically relevant ASD variants while achieving an overall diagnostic yield of ~ 10%, comparable to results from prior studies[5,21,22]. They also perform better than the existing NDD variant prioritization tool, *AutoCaSc*[14], in detecting clinically relevant ASD candidate variants. The higher accuracy of *AutScore* compared to *AutoCaSc* is likely because it was explicitly designed to detect ASD candidate variants. At the same time, *AutoCaSc* focuses on prioritizing candidate variants related to a broader range of NDDs.

The following limitations should be considered when using *AutScore*. First, the original *AutScore* metric was established using a trial-and-error approach, assigning certain weights and penalties to its different modules and that may lead to inherent subjectivity biases. This was overcome later by refining the weights of the different modules in the revised version of *AutScore.r*. Second, the accuracy of both *AutScore and AutScore.r* are not 100%, thus having some false positive and false negative findings. To address this limitation, one can use different thresholds of these tools to reduce the number of one type of these variants while increasing the other one. Third,

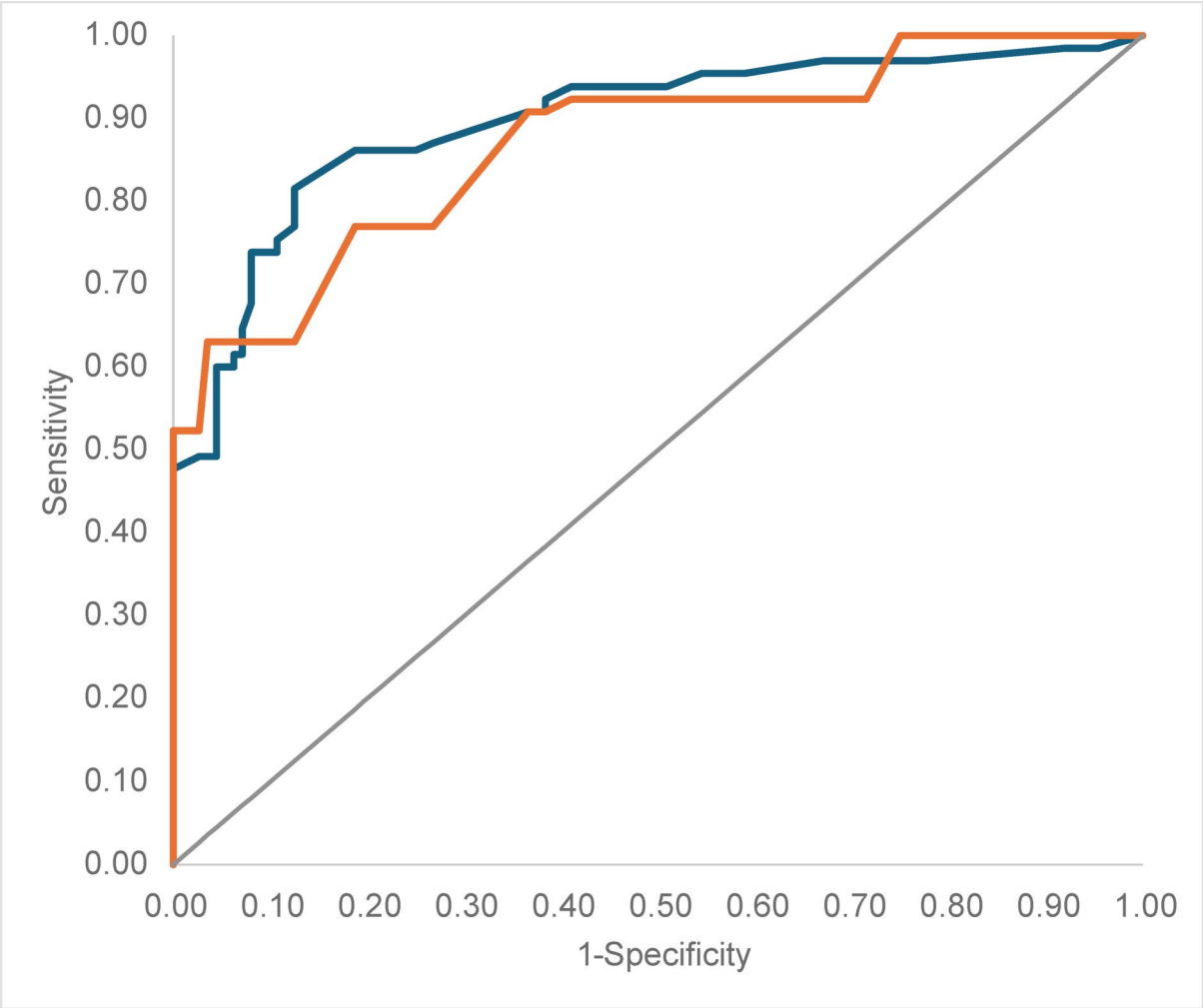

Fig. 3. ROC curves of *AutScore* (Orange line) and *AutScore.r* (Blue line) in detecting 'likely' ASD variants. Both approaches demonstrated good performance in detecting 'likely' ASD variants with *AutScore.r* performing slightly better than *AutScore* (AUC = 0.90, 95% CI = 0.84–0.95 vs. AUC = 0.84, 95%CI = 0.78–0.91) respectively).

in this study *AutScore* was constrained to specific genes from the DisGeNET[30] and SFARI Gene[29] databases and hence might have missed some potential candidate variants in genes not cataloged in these databases. Since both of these databases are constantly updating with new genes. future implication of *AutScore* may have even better performance. Forth, *AutScore.r* performance was based on variants from the top quartile of the original *AutScore*. Therefore, it is possible that our analysis underestimated the false-negative rates and, consequently also, the accuracy of the tool. Lastly, in its current form, *AutScore* is not designed to assess the effect of other types of genetic variants that may contribute to ASD such as CNVs[42] and compound heterozygotes[43]. Additionally, the metric may not function optimally in cases involving probands with incomplete pedigree information and unknown segregation patterns.

Conclusion

AutScore.r constitutes a highly effective automated ranking system designed to prioritize ASD candidate genetic variants in WES data. The utilization of *AutScore.r* holds the potential to significantly streamline the process of elucidating the specific genetic etiology of ASD within affected families. In doing so, it can contribute to expediting and enhancing the accuracy of clinical management and treatment strategies, ultimately leading to more effective interventions in the context of ASD.

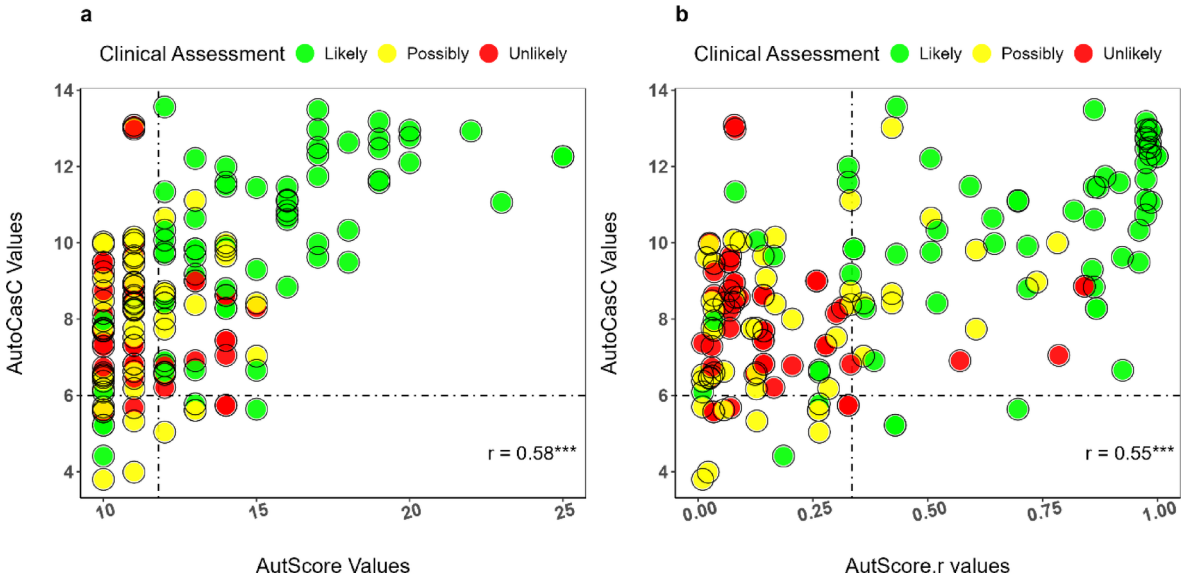

Fig. 4. Scatter plots of the relationships between (**A**) *AutScore* and *AutoCaSc* and (**B**) *AutScore.r* and *AutoCaSc* in detecting ASD candidate variants. Dashed horizontal and vertical lines depicting the cutoff used by the different tools in determining "positive findings". Colored dots depict the clinical genetic variant ranking ("likely" [green], "possibly" [yellow], and "unlikely" [red]).

Scoring Approaches	Sensitivity	Specificity	PPV	Accuracy (95% C.I.)	Yield (%) (Likely)
AutScore (≥ 12)	0.91	0.62	0.58	0.72 (0.65, 0.79)	9.81
AutScore.r (≥ 0.335)	0.81	0.87	0.79	0.85 (0.79, 0.90)	8.78
AutoCasC (> 6)	0.92	0.13	0.4	0.43 (0.36, 0.51)	9.98

Table 1. Comparing the performance between AutScore, AutScore.r and autocasc in detecting ASD candidate variants.

Functional Consequences	Frequency (%)
Frameshift Insertion/Deletion	17 (25.4%)
Splice Acceptor/Donor	5 (7.46%)
Stop Gained/Lost	13 (19.4%)
Missense	29 (43.3%)
Other	3 (4.48%)
Gene Type	
SFARI 1	51 (76.1%)
SFARI 2 and 3	11 (16.4%)
Any SFARI	62 (92.5%)
Novel Genes	5 (7.46%)
Inheritance Pattern	
Denovo	49 (73.1%)
X-linked	8 (11.9%)
Autosomal Recessive	10 (14.9%)
Variant Type	
Pathogenic (P)	35 (52.2%)
Likely Pathogenic (LP)	18 (26.9%)
VUS/LGD	14 (20.9%)
Clinical Assessment	
Likely ASD Candidates	53 (79.1%)
Possibly ASD Candidates	10 (14.9%)
Unlikely ASD Candidates	4 (5.97%)

Table 2. Characteristics of the variants detected with AutScore.$r \geq 0.335$ ($N = 67$). The *variant type* was determined using *InterVar* according to the ACMG/AMP criteria[10].

Data availability

WES data were generated as part of the ASC and are available in dbGaP with study accession: phs000298.v4.p3. More details about the input data and the implementation R script of AutScore.r can be found in the "Implementation Notes" section in the Supplement and at a publicly available GitHub repository: https://github.com/AppWick-hub/AutScore. Additional data will be available at reasonable requests to the corresponding author, Prof. Idan Menashe.

Received: 4 June 2024; Accepted: 25 March 2025
Published online: 15 April 2025

References

1. Rees, E. et al. Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. *Nat. Commun.* **12**, 1–9 (2021).
2. Zoghbi, A. W. et al. High-impact rare genetic variants in severe schizophrenia. *Proc. Natl. Acad. Sci. U S A.* **118**, 1–10 (2021).
3. An, J. Y. et al. Genome-wide de Novo risk score implicates promoter variation in autism spectrum disorder. Science 362, eaat6576 (2018) https://doi.org/10.1126/science.aat6576
4. Sanders, S. J. et al. Whole genome sequencing in psychiatric disorders: the WGSPD consortium. *Nature Neuroscience 2017 20:12* 20, 1661–1668 (2017).
5. Trost, B. et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **185**, 4409–4427e18 (2022).
6. Foo, J. N., Liu, J. J. & Tan, E. K. Whole-genome and whole-exome sequencing in neurological diseases. *Nat. Rev. Neurol.* **8**, 508–517 (2012).
7. Yuen, R. K. C. et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* **21**, 185–191 (2015).
8. Reuter, M. S. et al. Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA Psychiatry*. **74**, 293–299 (2017).
9. Forstner, A. J. et al. Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families. *Transl Psychiatry* **10**(1), 57, https://doi.org/10.1038/s41398-020-0732-y(2020).
10. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Sci.* **17**, 405–424 (2015).
11. Smedley, D. et al. Next-generation diagnostics and disease-gene discovery with the exomiser. *Nat. Protoc.* **10**, 2004–2015 (2015).
12. Birgmeier, J. et al. AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. *Sci. Transl Med.* **12**(544), eaau9113, https://doi.org/10.1126/scitranslmed.aau9113 (2020).
13. Robinson, P. N. et al. Interpretable clinical genomics with a likelihood ratio paradigm. *Am. J. Hum. Genet.* **107**, 403–417 (2020).
14. Popp, B., Lieberwirth, J., Benjamin, B., Kl, C. & Jamra, R. A. AutoCaSc : prioritizing candidate genes for neurodevelopmental disorders. **43**(12), 1795–1807 (2022).
15. Muers, M. Fruits of exome sequencing for autism. *Nature Reviews Genetics 2012 13:6* 13, 377–377 (2012).
16. Fu, J. M. et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**(9), 1320–1331. (2022).
17. Satterstrom, F. K. et al. Large-Scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584e23 (2020).

18. Yuen, R. K. C. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
19. Guo, H. et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genet. Sci.* **21**, 1611–1620 (2019).
20. Jiang, Y. H. et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **93**, 249–263 (2013).
21. Mahjani, B. et al. Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder. *Mol. Autism*. **12**, 1–12 (2021).
22. Tammimies, K. et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA - J. Am. Med. Association*. **314**, 595–903 (2015).
23. Dinstein, I. et al. The National autism database of Israel: a resource for studying autism risk factors, biomarkers, outcome measures, and treatment efficacy. *J. Mol. Neurosci.* **70**, 1303–1312 (2020).
24. Meiri, G. et al. Brief report: the Negev Hospital-University-Based (HUB) autism database. *J. Autism Dev. Disord*. **47**, 2918–2926 (2017).
25. McKenna, A. et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297 (2010).
26. Miller, N. A. et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* **7**, 1–16 (2015).
27. Shil, A. et al. Comparison of three bioinformatics tools in the detection of ASD candidate variants from whole exome sequencing data. *Sci. Rep.* | **13**, 18853 (123AD).
28. Li, Q., Wang, K. & InterVar Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am. J. Hum. Genet.* **100**, 267–280 (2017).
29. Abrahams, B. S. et al. SFARI gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism*. **4**, 2–4 (2013).
30. Piñero, J. et al. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).
31. Ng, P. C. & Henikoff, S. S. I. F. T. Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
32. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, **2**, https://doi.org/10.1002/0471142905.hg0720s76 (2013).
33. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
34. Ioannidis, N. M. et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
35. Jagadeesh, K. A. et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
36. Samocha, K. E. et al. Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* (2017). https://doi.org/10.1101/148353
37. Quinodoz, M. et al. Using machine learning to predict genes associated with dominant disorders. *Am. J. Hum. Genet.* **101**, 623–629 (2017).
38. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nature Biotechnology 2011 29:1* 29, 24–26 (2011).
40. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
41. Abdi, M. et al. Genomic architecture of autism spectrum disorder in Qatar: the BARAKA-Qatar study. *BMC Genome Med.* https://doi.org/10.1186/s13073-023-01228-w (2023).
42. Sebat, J. et al. Strong association of de Novo copy number mutations with autism. *Sci. (1979)*. **316**, 445–449 (2007).
43. Lin, B. D. et al. The role of rare compound heterozygous events in autism spectrum disorder. *Transl Psychiatry*. https://doi.org/10.1038/s41398-020-00866-7 (2020).

Acknowledgements

We thank the families who participated in this research; genetic studies would be impossible without their contributions. Additionally, we thank Prof. Norm O'Rourke, Department of Epidemiology, Biostatistics and Community Health Sciences, Ben Gurion University of the Negev, Israel for critically reviewing the first draft of the manuscript.

Author contributions

Conceptualization: A.S. and I.M.; methodology: A.S. and I.M.; software: A.S. and L.L.; validation: N.A. and N.L.; formal analysis: A.S.; resources: N.S., H.A.K, G.M., A.M., Y.T., A.A., H.G., and I.M.; data curation: A.S.; writing—original draft preparation: A.S. and I.M.; writing—review and editing: I.M., and A.S.; supervision: I.M.; project administration: I.M.; funding acquisition: I.M. All the authors have read and agreed to the published version of the manuscript.

Funding

This study was funded by the Israel Science Foundation (#1092/21).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Written consent was obtained from all parents of children involved in the study.

Consent for publication
All the data from the registered families presented here are deidentified.

Institutional review board statement
The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Soroka University Medical Center (SOR-076-15; 17 April 2016).

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96063-x.

Correspondence and requests for materials should be addressed to I.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.