

***AutScore* – An integrative scoring approach for prioritization of ultra-rare autism spectrum disorder candidate variants from whole exome sequencing data**

Apurba Shil^{1,2,3}, Noa Arava², Noam Levi^{1,2,3}, Liron Levine⁴, Hava Golan^{2,3,5}, Gal Meiri^{2,6}, Analya Michaelovski^{2,7}, Yair Tsadaka^{2,8}, Adi Aran^{9,10}, and Idan Menashe^{1,2,3*}

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

* Correspondence: Prof. Idan Menashe, Email: idanmen@bgu.ac.il; tel.: +972-8-6477456

Abstract

Background Discerning clinically relevant 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.

Methods 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%), high-quality proband-specific variants affecting genes associated with ASD or other neurodevelopmental disorders (NDDs). We assigned a score (i.e., *AutScore*) to each such variant based on their pathogenicity, clinical relevance, gene-disease association, and inheritance patterns. Finally, we compared the *AutScore* performance with the rating of clinical experts and the NDD variants prioritization algorithm, *AutoCasC*.

Results Overall, 1161 ultra-rare variants distributed in 687 genes in 441 ASD probands were evaluated by *AutScore* with scores ranging from -4 to 25, with a mean \pm SD of 5.89 ± 4.18 . *AutScore* cut-off of ≥ 12 outperforms *AutoCasC* in detecting clinically relevant ASD variants, with a detection accuracy rate of 72.3% and an overall diagnostic yield of 11.9%. Sixteen variants with *AutScore* of ≥ 12 were distributed in fifteen novel ASD genes.

Conclusion *AutScore* is an effective automated ranking system for ASD candidate variants that could be implemented in ASD clinical genetics pipelines.

Keywords *AutScore*, candidate variants, ASD, WES, prioritization algorithm.

1 **Introduction**

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

8 Detecting disease-causing variants from WES/WGS data is a complex task. Today, most
9 clinical genetics labs that analyze WES/WGS data follow the American College of Medical
10 Genetics and Genomics (ACMG) guidelines for interpreting sequence variants [10]. This
11 mainly includes detecting high-quality variants with lower allele frequency and damaging
12 effects on the protein function. Other factors usually considered are the segregation of the
13 variant with the phenotype and existing evidence for the variant or gene association with the
14 disease. To assist clinicians in this laborious process, several automated tools such as Exomiser
15 [11], AMELIE [12], LIRICAL [13], AutoCasC [14], etc., have been devised to prioritize
16 disease-specific variants (mainly single nucleotide variants [SNVs] and insertions/deletions
17 [indels]) from WES/WGS data.

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

24 Here, we present an automated scoring approach called *AutScore* that integrates variant and
25 gene-level information such as pathogenicity, deleteriousness, clinical relevance, gene-disease

26 association, and gene-variant inheritance pattern from a wide range of bioinformatics tools and
27 databases to generate a single score for prioritizing clinically relevant ASD candidate variants
28 from WES data for simplex and multiplex families. We applied the *AutScore* to WES data from
29 581 Israeli ASD-affected probands and their parents. We assessed its performance by
30 comparing the obtained results to a manual and blinded evaluation of the variants by clinicians
31 and to *AutoCasC* [14], an existing variant prioritization tool for neurodevelopmental disorders
32 (NDDs).

33

34 **Materials and Methods**

35 Study Sample

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

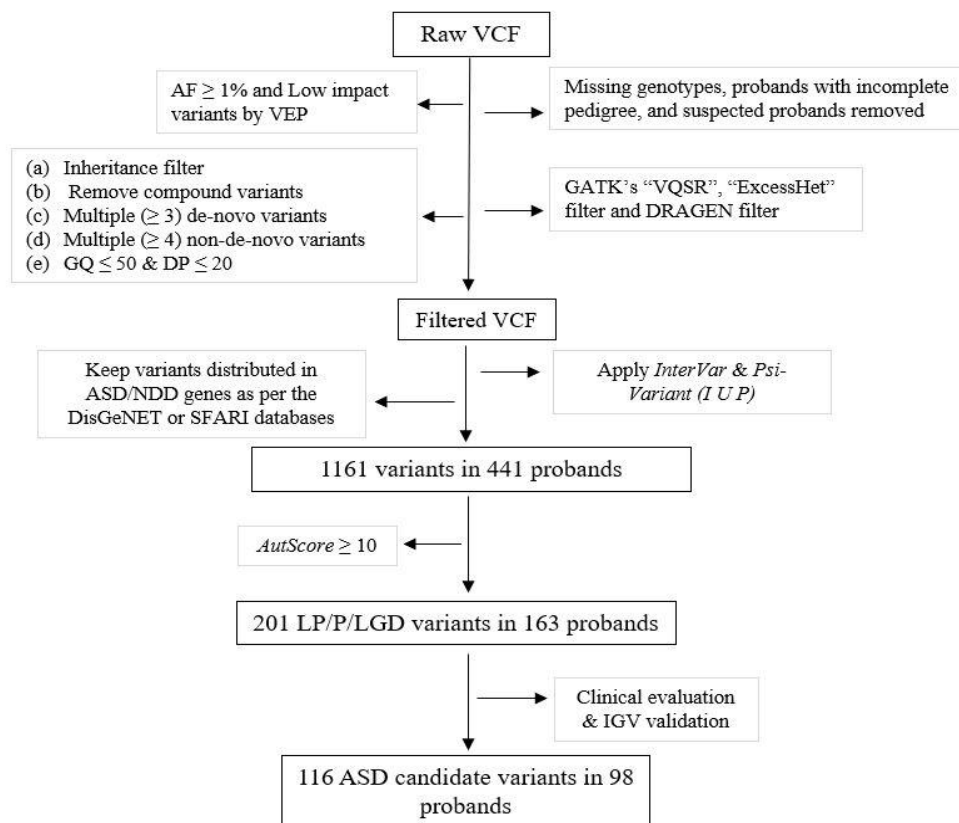
42 Whole Exome Sequencing (WES)

43 Whole Exome Sequencing (WES) analysis was conducted in two labs: (1) the Broad Institute
44 as a part of the Autism Sequencing Consortium (ASC) project [25] and (2) the Clalit Health
45 Services sequencing lab at Beilinson Hospital. WES was performed using Illumina HiSeq
46 sequencers in both places, followed by the Illumina Nextera exome capture kit. The sequencing
47 reads were aligned to human genome build 38 and aggregated into BAM/CRAM files. Then,
48 the Genome Analysis Toolkit (GATK) [26] (Broad) or Illumina's DRAGEN pipeline [27]
49 (Beilinson) was used for variant discovery and the generation of joint variant calling format

50 (vcf) files.

51 Variant filtering and annotations

52 The multi-sample vcf files generated by the Genome Analysis Toolkit (GATK) and the
53 DRAGEN platform were undertaken with identical procedures for variant filtering and
54 annotation, as previously detailed [28]. Subsequently, we identified pathogenic (P), likely
55 pathogenic (LP), or likely gene-disrupting (LGD) variants using the *InterVar* [29] tool in
56 conjunction with our proprietary tool, *Psi-Variant* [28]. We kept only those LP/P/LGD variants
57 that affected genes associated with ASD or other neurodevelopment disorders (NDDs)
58 according to the SFARI gene [30] or the DisGeNET [31] databases for downstream analyses.
59 Subsequently, 1161 candidate variants in 441 probands remained for further analysis (Fig. 1).



60

61 **Fig. 1** Analysis workflow for detecting ASD candidate variants from the WES data.

62 Prioritization of ASD candidate variant

63 We developed a metric called *AutScore* to prioritize the detected list of ASD candidate variants
64 as follows:

$$65 \quad \textit{AutScore} = I + P + D + S + G + C + H$$

66 Where:

67 • I – indicates the pathogenicity of a variant based on *InterVar* [29] classification as
68 follows: ‘benign’ = -3; ‘likely benign’ = -1; ‘variants of uncertain significance (VUS)’ = 0;
69 ‘likely pathogenic’ = 3, and ‘pathogenic’ = 6.

70

71 • P – cumulatively assess the deleteriousness of a variant based on the following six in-
72 silico tools (SIFT [32] (< 0.05), PolyPhen-2 [33] (≥ 0.15), CADD [34] (> 20), REVEL [35]
73 (> 0.50), M_CAP [36] (> 0.025) and MPC [37] (≥ 2)). For each of these tools, a variant gets
74 a score of 1 (deleterious) or 0 (benign), and these scores are aggregated to generate a single
75 score ranging from 1 to 6.

76

77 • D – indicates the agreement of variant-phenotype segregation with the predicted
78 segregation by the Domino tool [38] where agreement with Domino’s ‘very likely
79 dominant/recessive’ classes = 2; agreement with Domino’s ‘likely dominant/recessive’
80 classes = 1; disagreement with Domino’s ‘very likely dominant/recessive’ classes = -2;
81 disagreement with Domino’s ‘likely dominant/recessive’ classes = -1; and 0 were assigned
82 for variants with Domino’s ‘either dominant or recessive’ segregation.

83

84 • S – indicated the strength of association of the affected gene with ASD according to
85 the SFARI gene database [30] where ‘high confidence’ = 3; ‘strong candidate’ = 2;
86 ‘suggestive evidence’ = 1; and not in SFARI database = 0.

87

- 88 • G – indicated the strength of association of the affected gene with ASD according to
89 the DisGeNET database [31] where weak/no association (GDA=0 to 0.25) = 0: mild
90 association (GDA=0.25 to 0.50) = 1: moderate association (GDA=0.50-0.75) = 2: strong
91 association (GDA=0.75 and above) = 3.
- 92
- 93 • C – pathogenicity of a variant based on ClinVar [39] where ‘benign’ = -3; ‘likely
94 benign’ = -1; ‘VUS’ = 0; ‘Likely pathogenic’ = 1; ‘Pathogenic’ = 3.
- 95
- 96 • H – segregation of variants in the family weighted as $(n^2)-1$ where n=number of
97 probands in a family that carries the detected variants.

98 Clinical genetics validation

99 Variants with *AutScore* ≥ 10 (top quartile of candidate variants scores) were visually validated
100 using the IGV software [40] and then manually examined by clinical geneticists according to
101 the standard ACMG/AMP guidelines [10]. The clinical experts assessed the likelihood of the
102 variants contributing to the ASD phenotype of the child and assigned each variant one of the
103 following rankings: ‘Likely,’ ‘Possibly,’ and ‘Unlikely’.

104 Statistical Analysis

105 We used a Receiver Operating Characteristic (ROC) analysis to assess the performance of
106 *AutScore* in detecting ASD candidate variants using the clinical experts' rankings as the
107 reference. We also accordingly compared the sensitivity, specificity, positive predictive value
108 (PPV), negative predictive value (NPV), and accuracy. In addition, diagnostic yield (%) was
109 computed as the proportion of the number of ASD probands that have at least one ASD
110 candidate variant out of the total affected ASD probands that completed their WES analysis.
111 We compared the performance of *AutScore* in detecting ASD candidate variants with the

112 performance of *AutoCasC* [14], an existing variant prioritization tool for NDDs. The agreement
113 between *AutScore* and *AutoCasC* scores, and between these scores and the clinical assessment
114 ranking, were assessed using Pearson’s correlation and Cohen’s Kappa statistic, respectively.

115 Software

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

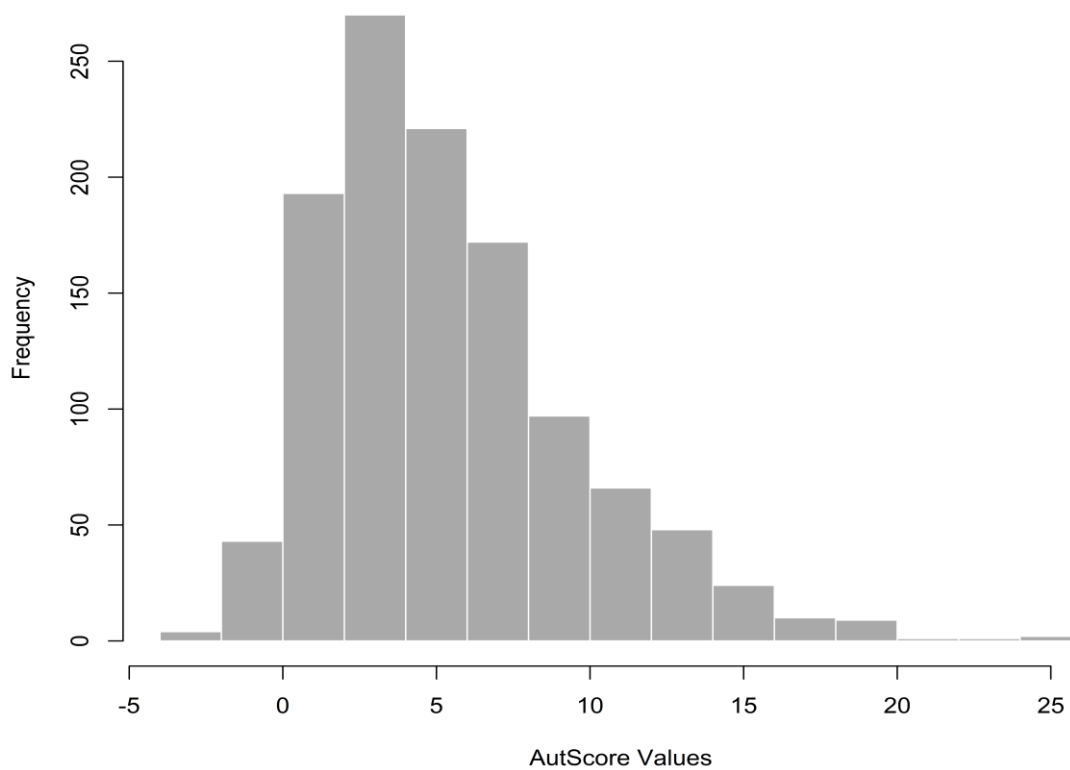
119 **Results**

120 A total of 1161 variants distributed in 687 genes in 441 ASD probands were evaluated by the
121 *AutScore* algorithm. Variant’s scores ranged from -4 to 25, with a mean \pm SD of 5.89 ± 4.18
122 (Fig. 2). The clinical experts examined 201 (17.31%) variants with an *AutScore* of ≥ 10 . Among
123 these, 24 (11.9%) were suspected as false positive indels during the visual assessment using
124 the IGV software and thus removed from subsequent analyses. Of the remaining 177 variants,
125 65 (36.7%) were ranked as ‘likely,’ 51 (28.8%) as ‘possibly,’ and 61 (34.5%) as ‘unlikely’
126 ASD candidate variants (Supplementary Table S1).

127 Identifying an optimum *AutScore* cut-off

128 Two analyses were carried out to identify the optimal *AutScore* cut-off (Fig. 3). First, an ROC
129 analysis using the clinical experts’ ranking: “likely” as the true set of ASD candidate variants
130 indicated that *AutScore* is an effective tool for detecting ASD clinically meaningful variants
131 (AUC=0.843, 95% CI= 0.779-0.907) (Fig. 3A). Applying Yuden J’s analysis to these data
132 suggested that an *AutScore* of ≥ 12 would be the most effective cut-off (Yuden J=0.52). The
133 same cut-off was also indicated by integrating detection accuracy and diagnostic yield (Max of
134 Yield + Accuracy /10=17.04) (Fig. 3B).

135



136

137 **Fig. 2** Histogram depicting the distribution of total LP/P/LGD variants assessed by *AutScore*

138 (N=1161)

139

140

141

142

143

144

145

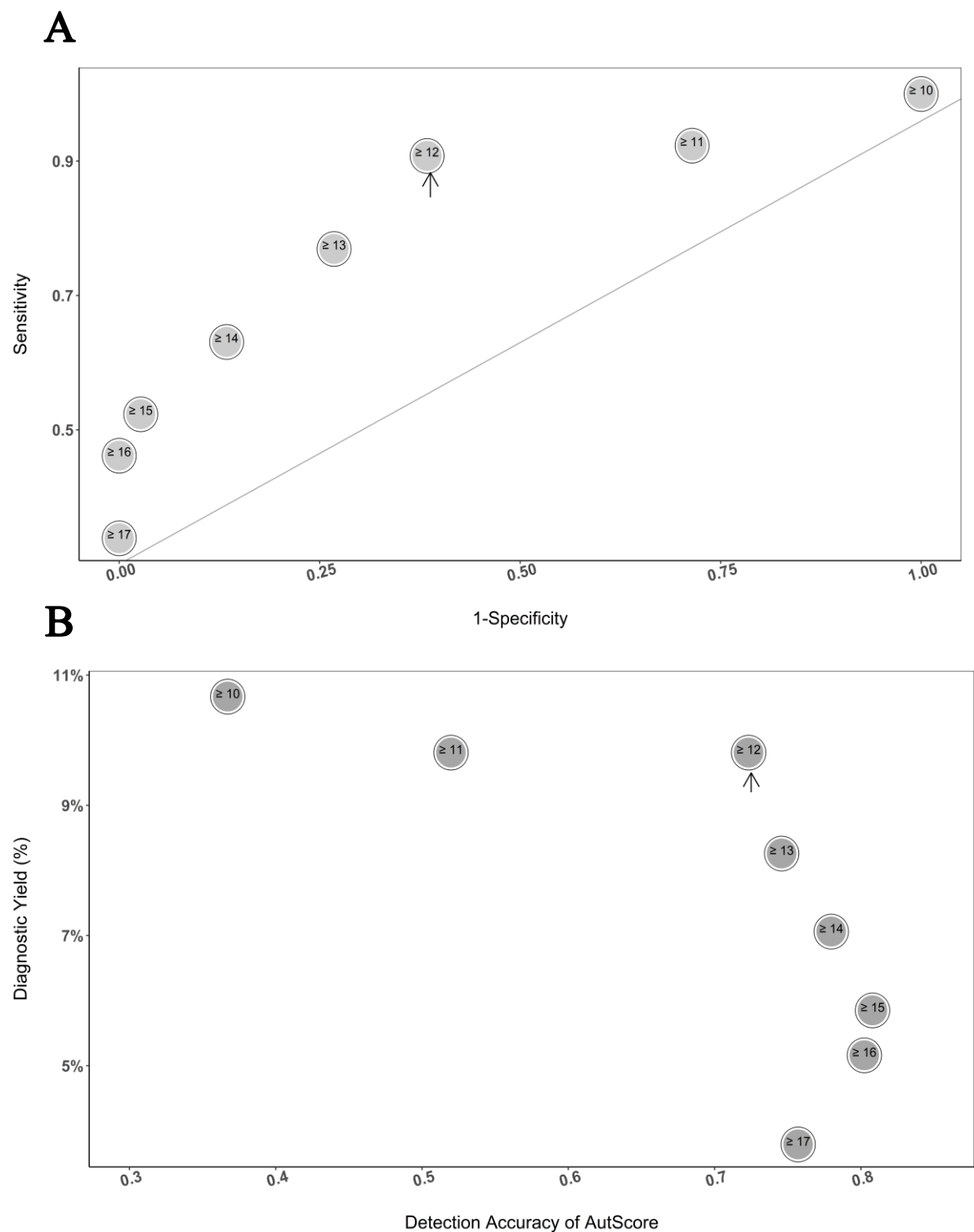
146

147

148

149

150



151 **Fig. 3** Assessing *AutScore*'s optimal cut-off for detection of ASD susceptibility variants. **A** A

152 receiver operating characteristics (ROC) analysis for different *AutScore* cut-offs. An arrow

153 indicates the best cut-off based on Yuden *J*'s statistics. **B** Scatterplot of the detection accuracy

154 (x-axis) and the resulting diagnostic yield (Y-axis) for different *AutScore* cut-offs. An arrow
155 indicates the best cut-off based on both values' aggregated maximum.

156 Comparing *AutScore* with *AutoCasC*

157 Next, we compared the performance of *AutScore* (using the selected cut-off ≥ 12) vis-à-vis the
158 existing NDD prioritization tool, *AutoCasC*, using its recommended cut-off of >6 [14], in
159 detecting ASD candidate variants (i.e., likely, possibly) (Fig. 4). A moderate, but statistically
160 significant correlation ($r=0.58$; $p<0.05$) was observed between *AutScore* and *AutoCasC*. Both
161 tools had high sensitivity in detecting ASD variants using their recommended cut-off (0.91 and
162 0.92, respectively; Table 1). Yet, *AutScore* outperformed in all other diagnostic characteristics
163 except in its diagnostic yield (Specificity: 0.616, PPV: 0.578 and Accuracy: 72.3%; 95% C.I:
164 65.1%-78.8% vs. Specificity: 0.133, and PPV: 0.397 Accuracy: 43.5%; 95% C.I: 35.9%-51.3%
165 respectively) (Table 1). In addition, *AutScore* results had a better agreement with the clinical
166 expert rankings than those of the *AutoCasC* (percentage agreement =72.3% and Cohen's
167 Kappa= 0.468 vs. percentage agreement=43.5% and Cohen's Kappa= 0.04 respectively; Table
168 2). The variant list ($n=177$) with *AutScore*, clinical assessments, and *AutoCasC* values is
169 provided in Supplementary Table S1.

Table 1: Comparing the performance between *AutScore* (≥ 12) and *AutoCasC* (> 6) in detecting ASD candidate variants

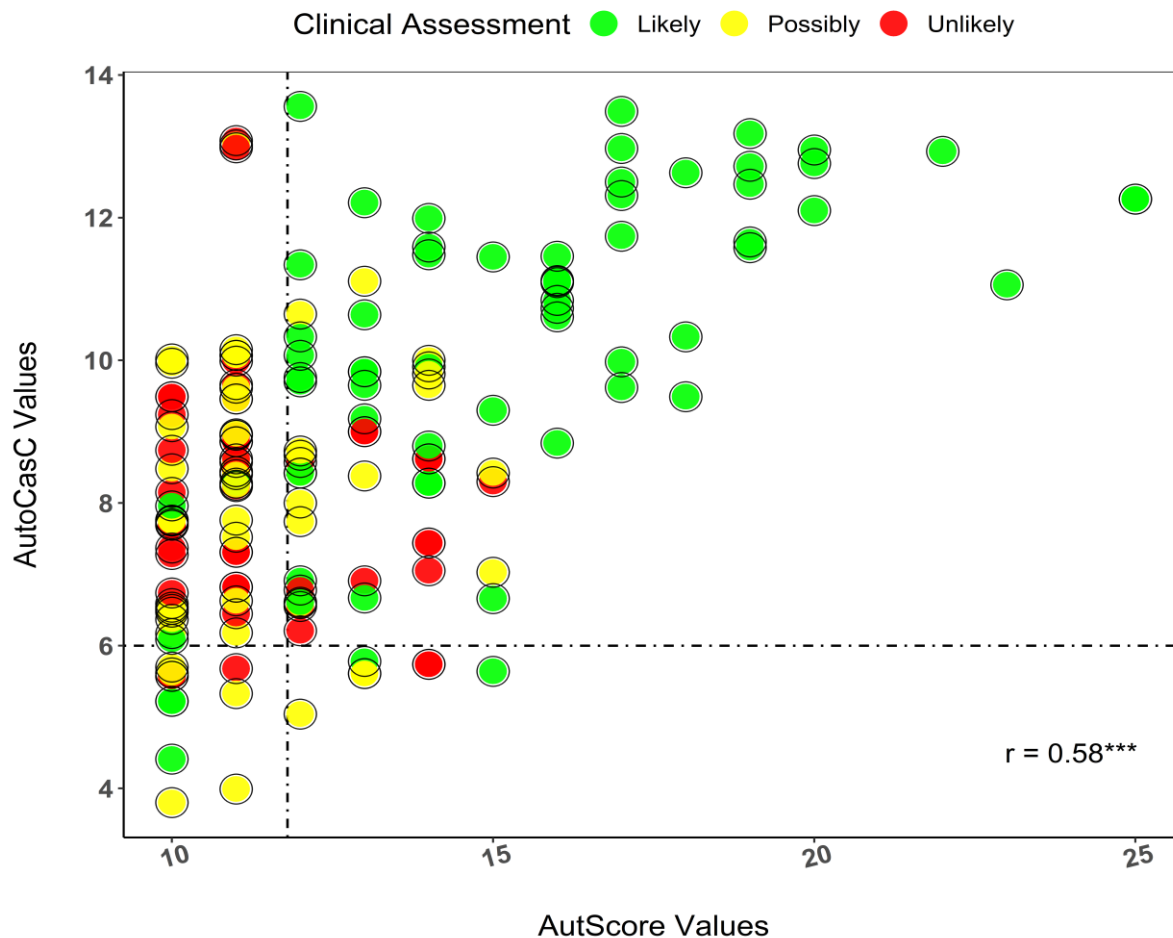
Scoring Approaches	Sensitivity	Specificity	PPV	Accuracy (95% C.I.)	Diagnostic Yield (Likely)	Diagnostic Yield (Likely+Possibly)	Yuden J
<i>AutScore</i> (≥ 12)	0.91	0.62	0.58	0.72 (0.65, 0.79)	9.81	11.9	0.52
<i>AutoCasC</i> (> 6)	0.92	0.13	0.40	0.43 (0.36, 0.51)	9.98	15.5	0.06

170

171

Table 2: Concordance between *AutScore* (≥ 12), *AutoCasC* (> 6), and Clinical Expert Rankings in detecting ASD candidate variants (N=177 variants)

Combinations	Percentage Agreement	Cohen's Kappa (P-Value)
<i>AutScore</i> (≥ 12) Vs <i>AutoCasC</i> (> 6)	57.6	0.07 (0.20)
<i>AutScore</i> (≥ 12) Vs Clinical Assessment	72.3	0.45 (0.00)
Clinical Assessment Vs. <i>AutoCasC</i> (> 6)	43.5	0.04 (0.26)



172 **Fig. 4** A clustered scatter diagram comparing the performance of *AutScore* (≥ 12), *AutoCasC*
173 (> 6), and the clinical assessments (e.g., likely (green), possibly (yellow), and unlikely (red)).

174 Characteristics of the LP/P/LGD variants detected by *AutScore*

175 Overall, 102 variants had an *AutScore* ≥ 12 . Of these, 59, 18, and 25 variants were ranked as
176 ‘likely’, ‘possibly’, and ‘unlikely’ ASD candidate variants, respectively, by the clinical experts
177 (Table 3). Most of the detected variants (45.1%) were distributed in high-confidence ASD
178 genes according to the SFARI Gene database [30] (i.e., SFARI score of 1). Another 29 (28.4%)
179 variants were detected in 23 genes (29.9%) not listed in the SFARI database and thus may be
180 considered novel ASD genes. More than 90% of the detected variants were classified as LP/P

181 according to the ACMG/AMP variant interpretation criteria [10], and more than 62% were
182 *denovo* variants.

Table 3: Characteristics of the detected variants with *AutScore* ≥ 12 (N=102)

Functional Consequences	Freq (%)
Frameshift Insertion/Deletion	25 (24.5%)
Splice Acceptor/Donor	6 (5.9%)
Stop Gained/Lost	22 (21.6%)
Missense	37 (36.3%)
Other	12 (11.8%)
Gene Type	
SFARI 1	46 (45.1%)
SFARI 2 and 3	27 (26.5%)
Any SFARI	73 (71.6%)
Novel Genes	29 (28.4%)
Inheritance Pattern	
<i>Denovo</i>	64 (62.7%)
X-linked	5 (4.9%)
Autosomal Recessive	33 (32.4%)
Variant Type	
Pathogenic (P)	56 (54.9%)
Likely Pathogenic (LP)	38 (37.3%)
VUS/LGD	8 (7.8%)
Clinical Assessment	
Likely	59 (57.8%)
Possibly	18 (17.6%)
Unlikely	25 (24.5%)

183 Discussion

184 Discerning clinically relevant ASD candidate variants from many variants poses a formidable
185 challenge for clinical experts, demanding considerable time and effort. Here, we present
186 *AutScore*, a novel bioinformatics prioritization tool that integrates variant and gene-level
187 information to prioritize ASD candidate variants derived from WES data. *AutScore* can be
188 integrated into an existing bioinformatic pipeline for WES data analysis by pre-installing the
189 ACMG/AMP [10] variant interpretation tool InterVar [14] and our in-house tool *Psi-Variant*
190 [28]. Although *AutScore* was initially designed to assess the ASD clinical relevance of rare
191 autosomal SNVs, it can be adapted for analyses of copy number variants (CNVs),

192 mitochondrial variants, and common heritable variants that are expected to enhance its
193 applicability further.

194 Our results indicated that *AutScore* is highly efficient in detecting clinically relevant ASD
195 variants. Using its most effective cut-off (i.e., ≥ 12), it achieves an overall diagnostic yield of
196 11.9%, comparable to results from prior studies [5,21,22]. We showed that *AutScore*
197 outperforms the existing NDD variant prioritization tool, *AutoCasC* [14], in detecting clinically
198 relevant ASD candidate variants. The higher accuracy of *AutScore* compared to *AutoCasC* is
199 likely because it was explicitly designed to detect ASD candidate variants. At the same time,
200 *AutoCasC* focuses on prioritizing candidate variants related to a broader range of NDDs.

201 The following limitations should be considered when using *AutScore*. First, the *AutScore*
202 metric was established using a trial-and-error approach, assigning certain weights and penalties
203 to its different elements. It is possible to mitigate this inherent subjectivity using a machine
204 learning model-based prioritization score. Since such models require larger datasets of true
205 ASD variants, we plan to upgrade to *AutScore* when such datasets are available. Second,
206 *AutScore* is constrained to specific genes from the DisGeNET [31] and SFARI Gene [30]
207 databases. Consequently, it might have missed some potential candidate variants in genes not
208 cataloged in these databases. Third, the performance of *AutScore* data has not been assessed in
209 WGS data. Hence, caution should be taken when applying this ranking tool to prioritize ASD
210 candidate variants derived from WGS data. Fourth, the estimates derived from *AutScore*,
211 including accuracy, PPV, and yield, were computed based on WES data from an ASD cohort
212 within the Israeli population. Thus, these estimates could vary in other populations. Lastly,
213 *AutScore* may not function optimally in cases involving probands with incomplete pedigree
214 information and unknown segregation patterns.

215 **Conclusion**

216 *AutScore* constitutes a highly effective automated ranking system designed to prioritize ASD

217 candidate genetic variants in WES data. The utilization of *AutScore* holds the potential to
218 significantly streamline the process of elucidating the specific genetic etiology of ASD within
219 affected families. In doing so, it can contribute to expediting and enhancing the accuracy of
220 clinical management and treatment strategies, ultimately leading to more effective
221 interventions in the context of ASD.

List of 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

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

Declarations

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).

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 de-identified.

Availability of data and materials

WES data were generated as part of the ASC and are available in dbGaP with study accession: phs000298.v4.p3. All the codes will be available upon reasonable request to the corresponding author.

Competing interests

The authors declare no competing interests.

Funding

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

Authors' 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.

Acknowledgments

We thank the families who participated in this research; genetic studies would be impossible without their contributions.

Authors' information (optional)

Reference

- [1] E. Rees *et al.*, ‘Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations’, *Nat Commun*, vol. 12, no. 1, pp. 1–9, 2021, doi: 10.1038/s41467-021-25532-4.
- [2] A. W. Zoghbi *et al.*, ‘High-impact rare genetic variants in severe schizophrenia,’ *Proc Natl Acad Sci U S A*, vol. 118, no. 51, pp. 1–10, 2021, doi: 10.1073/pnas.2112560118.
- [3] J.-Y. An *et al.*, ‘Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder’, 2018, doi: 10.1126/science.aat6576.
- [4] S. J. Sanders *et al.*, ‘Whole genome sequencing in psychiatric disorders: the WGSPD consortium’, *Nature Neuroscience 2017 20:12*, vol. 20, no. 12, pp. 1661–1668, Nov. 2017, doi: 10.1038/s41593-017-0017-9.
- [5] B. Trost *et al.*, ‘Genomic architecture of autism from comprehensive whole-genome sequence annotation’, *Cell*, vol. 185, no. 23, pp. 4409–4427.e18, 2022, doi: 10.1016/j.cell.2022.10.009.
- [6] J. N. Foo, J. J. Liu, and E. K. Tan, ‘Whole-genome and whole-exome sequencing in neurological diseases’, *Nat Rev Neurol*, vol. 8, no. 9, pp. 508–517, 2012, doi: 10.1038/nrneurol.2012.148.
- [7] R. K. C. Yuen *et al.*, ‘Whole-genome sequencing of quartet families with autism spectrum disorder’, *Nat Med*, vol. 21, no. 2, pp. 185–191, 2015, doi: 10.1038/nm.3792.
- [8] M. S. Reuter *et al.*, ‘Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders’, *JAMA Psychiatry*, vol. 74, no. 3, pp. 293–299, 2017, doi: 10.1001/jamapsychiatry.2016.3798.
- [9] A. J. Forstner *et al.*, ‘Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families’, *Transl Psychiatry*, vol. 10, no. 1, 2020, doi: 10.1038/s41398-020-0732-y.
- [10] S. Richards *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’, *Genetics in Medicine*, vol. 17, no. 5, pp. 405–424, 2015, doi: 10.1038/gim.2015.30.
- [11] D. Smedley *et al.*, ‘Next-generation diagnostics and disease-gene discovery with the Exomiser’, *Nat Protoc*, vol. 10, no. 12, pp. 2004–2015, 2015, doi: 10.1038/nprot.2015.124.
- [12] J. Birgmeier *et al.*, ‘AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature’, *Sci Transl Med*, vol. 12, no. 544, 2020, doi: 10.1126/scitranslmed.aau9113.
- [13] P. N. Robinson *et al.*, ‘Interpretable Clinical Genomics with a Likelihood Ratio Paradigm’, *Am J Hum Genet*, vol. 107, no. 3, pp. 403–417, 2020, doi: 10.1016/j.ajhg.2020.06.021.
- [14] B. Popp, J. Lieberwirth, B. Benjamin, C. Kl, and R. A. Jamra, ‘AutoCaSc : Prioritizing candidate genes for neurodevelopmental disorders’, pp. 1–14, 2022.
- [15] M. Muers, ‘Fruits of exome sequencing for autism’, *Nature Reviews Genetics 2012 13:6*, vol. 13, no. 6, pp. 377–377, May 2012, doi: 10.1038/nrg3248.
- [16] J. M. Fu *et al.*, ‘Rare coding variation provides insight into the genetic architecture and phenotypic context of autism’, *Nat Genet*, vol. 54, no. September, 2022, doi: 10.1038/s41588-022-01104-0.
- [17] F. K. Satterstrom *et al.*, ‘Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism’, *Cell*, vol. 180, no. 3, pp. 568–584.e23, 2020, doi: 10.1016/j.cell.2019.12.036.
- [18] R. K. C. Yuen *et al.*, ‘Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder’, *Nat Neurosci*, vol. 20, no. 4, pp. 602–611, 2017, doi: 10.1038/nn.4524.
- [19] H. Guo *et al.*, ‘Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes’, *Genetics in Medicine*, vol. 21, no. 7, pp. 1611–1620, 2019, doi: 10.1038/s41436-018-0380-2.
- [20] Y. H. Jiang *et al.*, ‘Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing’, *Am J Hum Genet*, vol. 93, no. 2, pp. 249–263, 2013, doi: 10.1016/j.ajhg.2013.06.012.

- [21] B. Mahjani *et al.*, ‘Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder’, *Mol Autism*, vol. 12, no. 1, pp. 1–12, 2021, doi: 10.1186/s13229-021-00465-3.
- [22] K. Tammimies *et al.*, ‘Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder’, *JAMA - Journal of the American Medical Association*, vol. 314, no. 9, pp. 595–903, 2015, doi: 10.1001/jama.2015.10078.
- [23] I. Dinstein *et al.*, ‘The National Autism Database of Israel: a Resource for Studying Autism Risk Factors, Biomarkers, Outcome Measures, and Treatment Efficacy’, *Journal of Molecular Neuroscience*, vol. 70, no. 9, pp. 1303–1312, 2020, doi: 10.1007/s12031-020-01671-z.
- [24] G. Meiri *et al.*, ‘Brief Report: The Negev Hospital-University-Based (HUB) Autism Database’, *J Autism Dev Disord*, vol. 47, no. 9, pp. 2918–2926, 2017, doi: 10.1007/s10803-017-3207-0.
- [25] F. K. Satterstrom, J. A. Kosmicki, J. Wang, K. Roeder, M. J. Daly, and J. D. Buxbaum, ‘Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism Article Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism’, *Cell*, pp. 1–17, 2020, doi: 10.1016/j.cell.2019.12.036.
- [26] A. McKenna *et al.*, ‘The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data’, *Genome Res*, vol. 20, no. 9, p. 1297, Sep. 2010, doi: 10.1101/GR.107524.110.
- [27] N. A. Miller *et al.*, ‘A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases’, *Genome Med*, vol. 7, no. 1, pp. 1–16, Sep. 2015, doi: 10.1186/S13073-015-0221-8/FIGURES/4.
- [28] A. Shil *et al.*, ‘Comparison of three bioinformatics tools in the detection of ASD candidate variants from whole exome sequencing data’, *Scientific Reports* /, vol. 13, p. 18853, 123AD, doi: 10.1038/s41598-023-46258-x.
- [29] Q. Li and K. Wang, ‘InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines’, *Am J Hum Genet*, vol. 100, no. 2, pp. 267–280, 2017, doi: 10.1016/j.ajhg.2017.01.004.
- [30] B. S. Abrahams *et al.*, ‘SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs)’, *Mol Autism*, vol. 4, no. 1, pp. 2–4, 2013, doi: 10.1186/2040-2392-4-36.
- [31] J. Piñero *et al.*, ‘DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants’, *Nucleic Acids Res*, vol. 45, no. D1, pp. D833–D839, 2017, doi: 10.1093/nar/gkw943.
- [32] P. C. Ng and S. Henikoff, ‘SIFT: Predicting amino acid changes that affect protein function’, *Nucleic Acids Res*, vol. 31, no. 13, pp. 3812–3814, 2003, doi: 10.1093/nar/gkg509.
- [33] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2*, vol. 2, no. SUPPL.76. 2013. doi: 10.1002/0471142905.hg0720s76.
- [34] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher, ‘CADD: Predicting the deleteriousness of variants throughout the human genome’, *Nucleic Acids Res*, vol. 47, no. D1, pp. D886–D894, 2019, doi: 10.1093/nar/gky1016.
- [35] N. M. Ioannidis *et al.*, ‘REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants’, *Am J Hum Genet*, vol. 99, no. 4, pp. 877–885, 2016, doi: 10.1016/j.ajhg.2016.08.016.
- [36] K. A. Jagadeesh *et al.*, ‘M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity’, *Nat Genet*, vol. 48, no. 12, pp. 1581–1586, 2016, doi: 10.1038/ng.3703.
- [37] K. E. Samocha *et al.*, ‘Regional missense constraint improves variant deleteriousness prediction’, *bioRxiv*, 2017, doi: 10.1101/148353.
- [38] M. Quinodoz, B. Royer-Bertrand, K. Cisarova, S. A. Di Gioia, A. Superti-Furga, and C. Rivolta, ‘DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders’, *Am J Hum Genet*, vol. 101, no. 4, pp. 623–629, 2017, doi:

- 10.1016/j.ajhg.2017.09.001.
- [39] M. J. Landrum *et al.*, ‘ClinVar: improving access to variant interpretations and supporting evidence’, *Nucleic Acids Res*, vol. 46, no. D1, pp. D1062–D1067, Jan. 2018, doi: 10.1093/NAR/GKX1153.
- [40] J. T. Robinson *et al.*, ‘Integrative genomics viewer’, *Nature Biotechnology 2011 29:1*, vol. 29, no. 1, pp. 24–26, Jan. 2011, doi: 10.1038/nbt.1754.