AutScore – an integrative scoring approach for prioritization of ultra-rare candidate ASD genetic variants from whole exome sequencing (WES) data

Apurba Shil [1,2], Noa Arava [2], Noa Sadigurschi [2,3,4], Hadeel Abu-Kaf [2,4], Gal Meiri [2,4], Analya Michaelovski [2,5], Yair Tsadaka [2,6], Adi Aran [7], Ilan Dinstein [2,8], Hava Golan [2,3], and Idan Menashe [1,2]

[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] Department of Physiology and Cell Biology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel; [4] Preschool Psychiatric Unit, Soroka University Medical Center, Beer-Sheva, Israel; [5] Child Development Center, Soroka University Medical Center, Beer-Sheva, Israel; [6] Child Development Center, Ministry of Health, Beer-Sheva, Israel; [7] Psychology Neuropediatric Unit, Shaare Zedek Medical Center, Jerusalem, Israel; [8] Psychology Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Background

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental disorder with a high heritability estimate.

Whole-exome sequencing (WES) is an effective approach for detecting single-nucleotide variants (SNVs) associated with different human diseases. However, existing tools to identify clinically relevant SNVs in WES data are not specific to ASD.

Here we present '***AutScore***' an integrative scoring approach to prioritize rare ASD candidate SNVs from WES data.

Methods

We analyzed WES data from 438 children with ASD and their parents from the Azrieli National Centre for Autism and Neurodevelopment Research as follows:

Data cleaning

First, we removed SNVs with poor sequence quality (GQ ≤ 50 and DP ≤ 20), missing genotypes, minor allele frequency >1%, and false positive SNVs. Then, we used the pedigree structure of the families to detect proband-specific genotypes ((a) Denovo; (b) X-linked; (c) Autosomal Recessive (d) Dominant). Then, we used *InterVar* to detect Pathogenic (P)/Likely Pathogenic (LP) SNVs according to the ACMG/AMP guidelines and integration of six in-silico prediction tools (SIFT, PolyPhen-2, CADD, REVEL, M_CAP, and MPC (Psi-Variant (P))) to detect missense likely gene disrupting (LGD) SNVs. Finally, we analyzed only SNVs that affected genes associated with neurodevelopmental disorders (NDDs) according to the disGeNET database and/or with ASD according to the SFARI database.

SNV prioritization

We applied ***AutScore***, an in-house integrative algorithm that scores SNVs as follows:

$$AutScore = I + P + D + S + G + C + H$$

- ***I*** indicates the pathogenicity of an SNV based on InterVar (-3: benign; -1: likely benign; 0: VUS; 3: Likely pathogenic; 6: Pathogenic).
- ***P*** indicates whether an SNV is annotated as deleterious according to the default cut-off of six in-silico tools (MPC, CADD, SIFT, M-CAP, REVEL, and PolyPhen-2).
- ***D*** indicates the agreement of SNV segregation with the predicted inheritance pattern of the gene according to Domino. (-2: strong disagreement; -1: moderate disagreement, 0: inconclusive, 1: moderate agreement, and 2: strong agreement).
- ***S*** indicates the strength of association of the affected gene with ASD according to the SFARI gene database (0: no association; 1: weak association; 2, moderate association; 3 strong association).
- ***G*** indicates the strength of association of the affected gene with ASD according to the DisGeNET database (0: no association; 1: weak association; 2: moderate association; 3: strong association).
- ***C*** indicates the pathogenicity of an SNV based on ClinVar (-3: benign; -1: likely benign; 0: VUS; 3: Likely pathogenic; 6: Pathogenic).
- ***H*** indicates the segregation of SNVs in the family as follows: $H=(n^2)-1$, where n=number of probands with the detected SNVs.

AutScore validation

- The top quartile of candidate SNVs, according to *AutScore*, were examined by clinical experts according to the standard medical guidelines/criteria). They classified these SNVs into the following classes "Very Likely," "Maybe," and "Unlikely" ASD candidate SNVs.
- In addition, *AutScore* results were compared to the results of AutoCasC an existing SNV prioritization tool for neurodevelopmental disorders (NDDs) to assess the performance in detected ASD candidate SNVs.

Results

- Overall, 823 ultra-rare SNVs distributed in 552 genes in 327 ASD probands were evaluated by AutScore (median=5; IQR=5; range= -4 to 23) **(Figure 1)**.
- The clinical experts ranked 44 SNVs as "Very Likely," and 19 SNVs as "Maybe," resulting in an overall detection yield of 13.6%.
- Eleven SNVs were detected in eleven novel ASD genes.

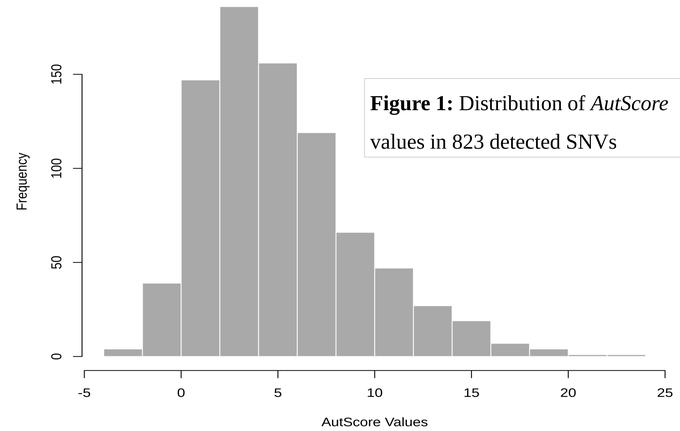

Figure 1: Distribution of *AutScore* values in 823 detected SNVs

- The optimal cut-off of AutScore was ≥ 12, resulting in a detection accuracy of 85.4% and a positive predictive value (PPV) of 0.687 (**Figure 2**).
- The concordance between AutScore, AutoCaSc, and Clinical assessments are as follows:
  (a) AutScore Vs AutoCasC: 49.3% (Cohen's Kappa: 0.06)
  (b) AutScore Vs Clinical Assessments: 85.4% (Cohen's Kappa: 0.76)
  (c) AutoCasC Vs Clinical Assessments: 38.1% (Cohen's Kappa=0.04)
- *AutScore* performs better in detecting ASD candidate SNVs than *AutoCasC*, an NDD SNV prioritization tool.

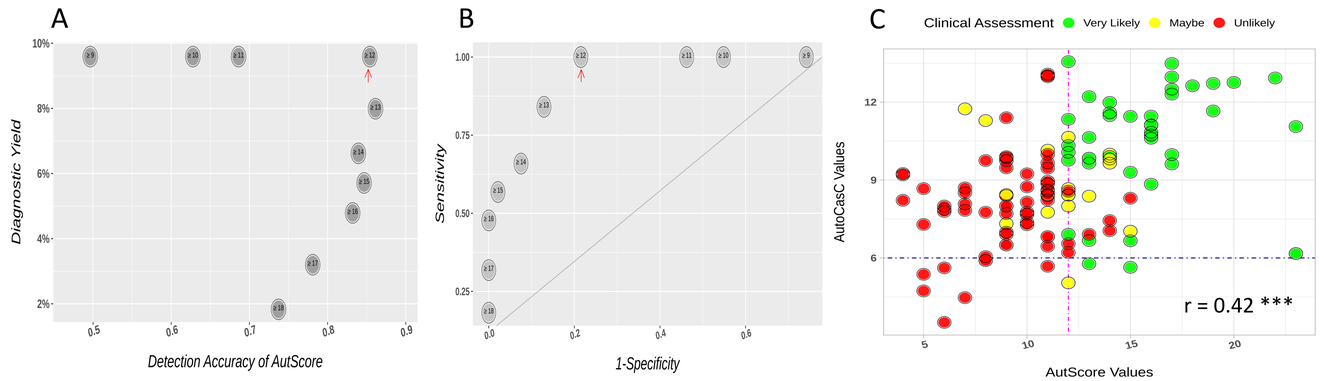

Figure 2: AutScore validation. Assessing the best cut-off of *Autscore* according to the resulting diagnostic yield and detection accuracy (**A**) and according to a ROC curve (**B**) suggests scores of ≥12 as the most effective cut-off. **(C)** Correlation between AutScore and AutoCasC scores and their agreement with the clinical classification.

Conclusions

- AutScore is an efficient automatic tool for prioritizing ASD candidate SNVs in WES data.
- Important extensions of this tool should include analyses of copy-number variants, compound heterozygote SNVs, and partially penetrant SNVs.

Contacts: idanmen@bgu.ac.il ; apurba.shil316@gmail.com