Effectiveness of three bioinformatics tools in the detection of ASD candidate variants from whole exome sequencing data

Apurba Shil [1,2,3], Liron Levin[4], Hava Golan[2,3,5], Gal Meiri[2,6], Analya Michaelovski[2,7], Yair Tsadaka[2,8], Adi Aran[9], Ilan Dinstein[2,3,10] and Idan Menashe[1,2,3*]

[1]Department of Epidemiology, Biostatistics, and Health Community 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]The School of Brain Sciences and Cognition, 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, Beer-Sheva, Israel.

[9]Psychology Neuropediatric Unit, Shaare Zedek Medical Center, Jerusalem, Israel.

[10]Psychology Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel

***** Correspondence: idanmen@bgu.ac.il; tel.: +972-8-6477456

Keywords: Whole-exome sequencing, Autism Spectrum disorder, Bioinformatics, Genetics, Genomics

Abstract

Background

Autism spectrum disorder (ASD) is a heterogenous multifactorial neurodevelopmental condition with a significant genetic susceptibility component. Thus, identifying genetic variations associated with ASD is a complex task. Whole-exome sequencing (WES) is an effective approach for detecting extremely rare protein-coding single-nucleotide variants (SNVs). However, interpreting the functional and clinical consequences of these variants requires integrating multifaceted genomic information.

Methods

We compared the effectiveness of three bioinformatics tools in detecting ASD candidate SNVs from WES data of 250 ASD family trios registered in the National Autism Database of Israel. We studied only rare (<1% population frequency), proband-specific SNVs. The pathogenicity of SNVs, according to the American College of Medical Genetics (ACMG) guidelines, was evaluated by the *InterVar* and *TAPES* tools. In addition, likely gene-disrupting (LGD) SNVs were detected based on an in-house bioinformatics tool, designated *Psi-Variant*, that integrates results from seven in-silico prediction tools. We compared the effectiveness of these three approaches – and their combinations – in detecting SNVs in high-confidence ASD genes.

Results

Overall, 605 SNVs in 499 genes distributed in 193 probands were detected by these tools. The overlap between the tools was 64.1%, 17.0%, and 21.6% for *InterVar–TAPES*, *InterVar–Psi-Variant*, and *TAPES–Psi-Variant*, respectively. The intersection between *InterVar* and *Psi-Variant* ($\underline{I} \cap \underline{P}$) was the most effective approach in detecting ASD genes (OR = 5.38, 95% C.I. = 3.25–8.53). This combination detected 102 SNVs in 99 genes among 80 probands (approximate 36% diagnostic yield).

Conclusions

Our results suggest that integration of different variant interpretation approaches in detecting ASD candidate SNVs from WES data is superior to each approach alone. Inclusion of additional criteria could further improve the detection of ASD candidate variants.

Background

Autism spectrum disorder (ASD) comprises a collection of heterogeneous neurodevelopmental disorders that share two behavioral characteristics—difficulties in social communication, and restricted, repetitive behaviors and interests [1, 2]. The etiology of ASD has a significant genetic component, as is evident from multiple twin and family studies [3–6]. Yet, over the years, very few genetic causes of ASD have been discovered; thus, today, despite extensive research, an understanding of the overall genetic architecture of ASD remains obscure [6, 7].

The emergence of next-generation sequencing (NGS) approaches in the past decade has transformed the genetic research of complex traits [8]. These NGS technologies have facilitated high-throughput DNA sequencing for large cohorts of patients, allowing the comparison of multiple single-nucleotide variants (SNVs) between large groups of patients [9–12]. In this realm, whole-exome sequencing (WES) is particularly suitable for studying the genetics of heterogenous traits such as ASD, as it focuses on a relatively limited number of protein-coding SNVs [9–11,13–17].

Understanding the functional consequences of coding SNVs is, however, not a trivial task. In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published standards and guidelines to generalize sequence variant interpretation and to address the issue of inconsistent interpretation across laboratories [8] . The resulting system for classifying SNVs recommends 28 criteria (16 for pathogenic and 12 for benign SNVs) and provides a set of scoring rules based on variant population allele frequency, variant functional annotation, variant familial segregation, etc.

[8,18]; SNVs are classified as pathogenic (P), likely pathogenic (LP), variants of uncertain significance (VUS), likely benign (LB) or benign (B). Subsequently, multiple in-silico tools were developed to implement these ACMG/AMP criteria for annotating the prospective pathogenicity of SNVs detected in WES studies.

While the ACMG/AMP scoring approach is highly effective for detecting de-novo highly penetrant mutations for rare Mendelian disorders, it is less suitable for detecting inherited partially penetrant SNVs [19]. Such SNVs, which are usually annotated as VUS in terms of the ACMP/AMP criteria, are expected to contribute significantly to the risk of developing neurodevelopmental conditions, including ASD [9,16,17,20,21]. Thus, relying solely on the ACMG/AMP criteria for variant annotation in WES studies of ASD may result in under-representation of susceptibility SNVs, which will lead to a lower diagnostic yield for ASD. To overcome this potential limitation, we have developed *"Psi-Variant*," as a pipeline to detect different types of likely gene-disrupting (LGD) SNVs, including protein truncating and deleterious missense SNVs. We applied *Psi-Variant* – in comparison with *InterVar* and *TAPES*, two SNV interpretation tools that use the ACMG/AMP criteria – to a large WES dataset of ASD to evaluate the concordance between these tools to detect SNVs and to assess their effectiveness in detecting ASD susceptibility SNVs.

Methods

Study sample

The study sample comprised 250 children diagnosed with ASD who are registered in the National Autism Database of Israel (NADI) [22,23] and whose parents gave consent for participation in this study. Based on our clinical records, none of the parents in the study has been diagnosed with ASD, intellectual disability, or any other type of neurodevelopmental disorder. Genomic DNA was extracted from saliva samples that were collected from

participating children and their parents using Oragene®•DNA (OG-500/575) collection kits (DNA Genotek, Canada).

Whole exome sequencing

WES analysis was performed on the above-mentioned samples with Illumina HiSeq sequencers, followed by the Illumina Nextera exome capture kit, at the Broad Institute as part of the Autism Sequencing Consortium, as described previously [11]. Sequencing reads aligned to Genome Reference Consortium Human Build 38 and aggregated into BAM/CRAM files were analyzed using the Genome Analysis Toolkit (GATK) [24] to generate a joint variant calling format (vcf) file for all subjects in the study. We excluded data for 30 probands from the raw vcf file due to incomplete pedigree information or to low-quality WES data. Thus, WES data for 220 ASD trios were analyzed in this study (Fig. 1).

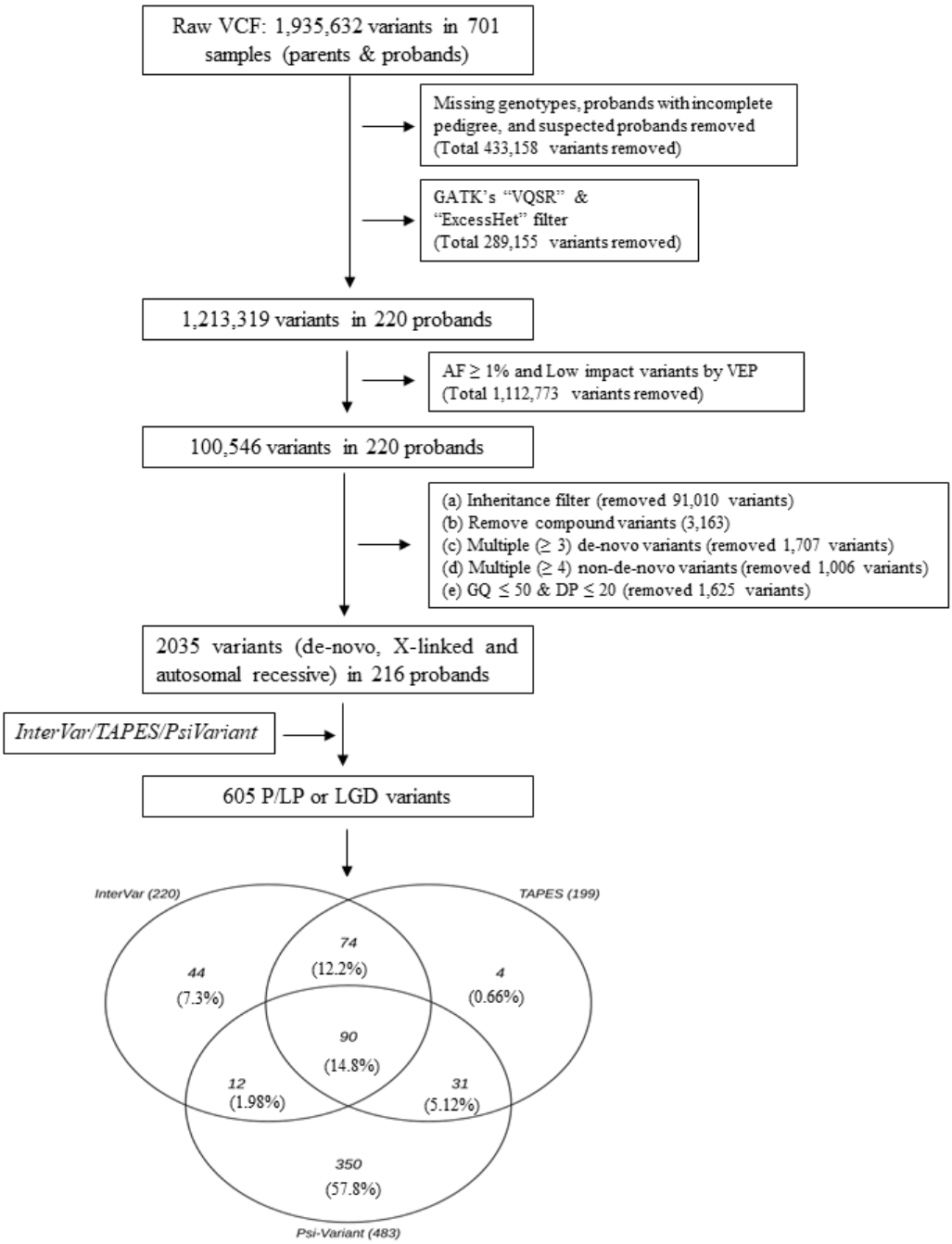

Fig. 1 Analysis workflow for detecting LP/P/LGD SNVs from the WES data. LP/P SNVs were detected by *InterVar* and *TAPES* by implementing ACMG/AMP criteria. LGD SNVs were detected by *Psi-Variant* by utilizing in-house criteria.

Data analysis

The SNV detection process in this study is outlined in Fig. 1. and explained below.

Data cleaning

The raw vcf file contained 1,935,632 SNVs. From this file, we removed SNVs with missing genotypes and/or SNVs in regions with low read coverage (≤ 20 reads), and/or with low genotype quality (GQ ≤ 50). In addition, we removed all common SNVs (i.e., those with a population minor allele frequency >1%) [25] as well as those that did not pass the GATK's "VQSR" and "ExcessHet" filters. Thereafter, we used an in-house machine learning (ML) algorithm to remove other potentially false-positive SNVs. The details of this ML algorithm and its efficiency in classifying between true positive and false positive SNVs are summarized in the supplementary file. Finally, we used the pedigree structure of the families to identify proband-specific genotypes, including de-novo SNVs, recessively inherited SNVs, and X-linked SNVs (in males). From these genotypes, we removed multiallelic SNVs and those that were classified as "de-novo" and appeared in more than two individuals in the sample.

Identifying ASD candidate SNVs

We searched for candidate ASD SNVs using three complementary approaches. First, we applied *InterVar* [18] and *TAPES* [26], two commonly used publicly available command-line tools that use ACMG/AMP criteria [8], to detect LP/P SNVs. In addition, we assigned the ACMG/AMP PS2 criterion to all the de-novo SNVs to detect additional LP/P SNVs from the list of VUS. Since both *InterVar* and *TAPES* are less sensitive tools for detecting recessive possible gene disrupting (LGD) SNVs [19], we developed an integrated in-house tool, designated *Psi-Variant*, to detect LGD SNVs. The *Psi-Variant* workflow starts by using Ensembl's Variant Effect Predictor (VEP) [25] to annotate the functional consequences for each

variant in a multi- sample vcf file. Then, all frameshift indels, nonsense and splice acceptor/donor SNVs are further analyzed by the LoFtool [27] with scores of < 0.25 are annotated as intolerant SNVs. In addition, it applies six different in-silico tools to all missense substitutions and annotates them as "deleterious/damaging" if at least three (≥ 50%) of them were exceeded the following cutoffs: SIFT [28] (< 0.05), PolyPhen-2 [29] (≥ 0.15), CADD [30] (> 20), REVEL [31] (> 0.50), M_CAP [32] (> 0.025) and MPC [33] (≥ 2). These scores were extracted by utilizing the dbNSFP database [34].

Comparison between InterVar, TAPES and Psi-Variant

We compared the number of SNVs detected by each of the three tools as well as percentages of SNVs that were detected by different combinations of these tools. Thereafter, we used the list of ASD genes (n = 1031) from the SFARI Gene database [35] (accessed on 11 January 2022) as the gold standard to test the effectiveness of these combinations. We computed the odds ratio (OR) and positive predictive value (PPV) for detecting candidate ASD SNVs in SFARI genes.

Software

Data storage, management, and analysis were conducted on a high-performing computer cluster in a Linux environment using Python version 3.5 and R Studio version 1.1.456. All the statistical analyses and data visualizations were incorporated in R Studio.

Results

Detection of candidate SNVs by the different tools

A total of 605 SNVs in 193 probands were detected by *InterVar* (n = 220), *TAPES* (n = 199) or *Psi-Variant* (n = 483) from a dataset of 2,035 high-quality, ultra-rare SNVs with proband-specific genotypes (Fig. 1). Of these, 90 SNVs (14.9%) were detected by all three tools. The

highest concordance in detected SNVs was observed between *InterVar* and *TAPES* (64.3%), followed by *TAPES* and *Psi-Variant* (21.6%) and *InterVar* and *Psi-Variant* (17.0%).

The characteristics of the detected SNVs are shown in Table 1. Significantly higher rates of LoF and missense SNVs were detected by all three tools compared to the rates of these SNVs in the clean vcf file ($P < 0.001$). As expected, missense variants comprised the majority of detected SNVs, with 81.6%, 58.8% and 51.4% of the SNVs detected by *Psi-Variant, TAPES,* and *InterVar*, respectively. Notably, a higher number of frameshift SNVs were detected by *Psi-Variant* than by *InterVar* and *TAPES* (58 vs. 39 and 22, respectively), but the percentages of these SNVs out of the total number of detected SNVs was lower due to the markedly higher number of SNVs detected by *Psi-Variant*.

Notably, almost all (≥ 95%) SNVs detected by either *InterVar* or *TAPES* were de-novo SNVs, while de-novo SNVs comprised only 36.2% of the SNVs detected by *Psi-Variant*, which also detected a high portion of X-linked and autosomal recessive SNVs (37.1% and 26.7%, respectively). Examination of the distribution of the detected SNVs in genes associated with ASD according to SFARI Gene database [35] revealed a two-fold enrichment of SNVs distributed in ASD genes (for all detection tools) compared to their portion in the clean vcf file, and even a higher enrichment of SNVs distributed in high-confidence ASD genes ($P < 0.001$).

Table 1 Characteristics of the detected SNVs by *InterVar*, *TAPES*, and *Psi-Variant* from the WES data

Characteristics	**Preliminary output (n = 1213319)**	***InterVar* (n = 220)**	***TAPES* (n = 199)**	***Psi-Variant* (n = 483)**
Functional consequence				
Frameshift (insertions/deletions)	4232 (0.349%)	39 (17.7%) *	22 (11.1%) *	58 (12.01%) *
Missense	95919 (7.91%)	113 (51.4%) *	117 (58.8%) *	394 (81.6%) *
Stop Gain/Loss/retain, Start Gain/Loss	2105 (0.17%)	16 (7.27%) *	13 (6.53%) *	16 (3.31%) *
Non-frameshift/in-frame	4062 (0.33%)	42 (19.1%) *	43 (21.61%) *	--
Splice acceptor/donor/region	18817 (1.55%)	4 (1.82%)	4 (2.01%)	12 (2.48%)
Synonymous, downstream/upstream gene, intron variant	871205 (71.8%)	6 (2.73%)	0 (0%)	3 (0.62%)
Other	216979 (17.9%)	--	--	--
Inheritance pattern wise				

De-novo	43052 (3.55%)	209 (95%) *	193 (97%) *	175 (36.2%) *
Autosomal recessive	70948 (5.85%)	9 (4.09%)	5 (2.51%)	179 (37.1%) *
X-linked	9103 (0.75%)	2 (0.91%) *	1 (0.5%) *	129 (26.7%) *
Other	1090216 (89.8%)	--	--	--
Gene type wise				
SFARI genes with score 1	19236 (1.58%)	15 (6.82%) *	12 (6.03%) *	21 (4.35%) *
All SFARI genes (with scores 1-3)	93681 (7.72%)	32 (14.5%) *	24 (12.1%) *	75 (15.5%) *
Other genes	1119638 (92.28%)	188 (85.4%) *	175 (87.9%) *	408 (84.5%) *

* <0.05 level of significance; two-sided two proportions Z test

Effectiveness of ASD candidate SNVs detection

To assess the effectiveness of the different tools in detecting ASD candidate SNVs, we calculated the PPV and the OR for detecting ASD genes (i.e., those listed in the SFARI Gene database [35]) for different combinations of utilization of the three tools. The results of these analyses are shown in Fig. 2. Utilization of any of the three tools resulted in a significant enrichment of ASD genes, with the highest enrichment being observed in SNVs detected by *InterVar* (PPV = 0.178; OR = 4.10, 95% confidence interval (C.I.) = 2.77–5.90) followed by *TAPES* (PPV = 0.158; OR = 3.53, 95% C.I. = 2.28–5.27) and *Psi-Variant* (PPV = 0.143; OR = 3.21, 95% C.I. = 2.39–4.22). Notably, better performance in detecting ASD candidate SNVs was obtained at the intersection of the detected SNVs between *InterVar* and *Psi-Variant* ($\underline{I} \cap \underline{P}$) (PPV = 0.222; OR = 5.38, 95% CI = 3.25–8.53). The $\underline{I} \cap \underline{P}$ combination was also the most effective in detecting SNVs in high-confidence ASD genes (i.e., those with score of 1 in the SFARI Gene database [35] (Fig. 2). Overall, the $\underline{I} \cap \underline{P}$ combination detected 102 SNVs distributed in 99 genes (22 SFARI genes) in 80 probands (a detection yield of 36%).

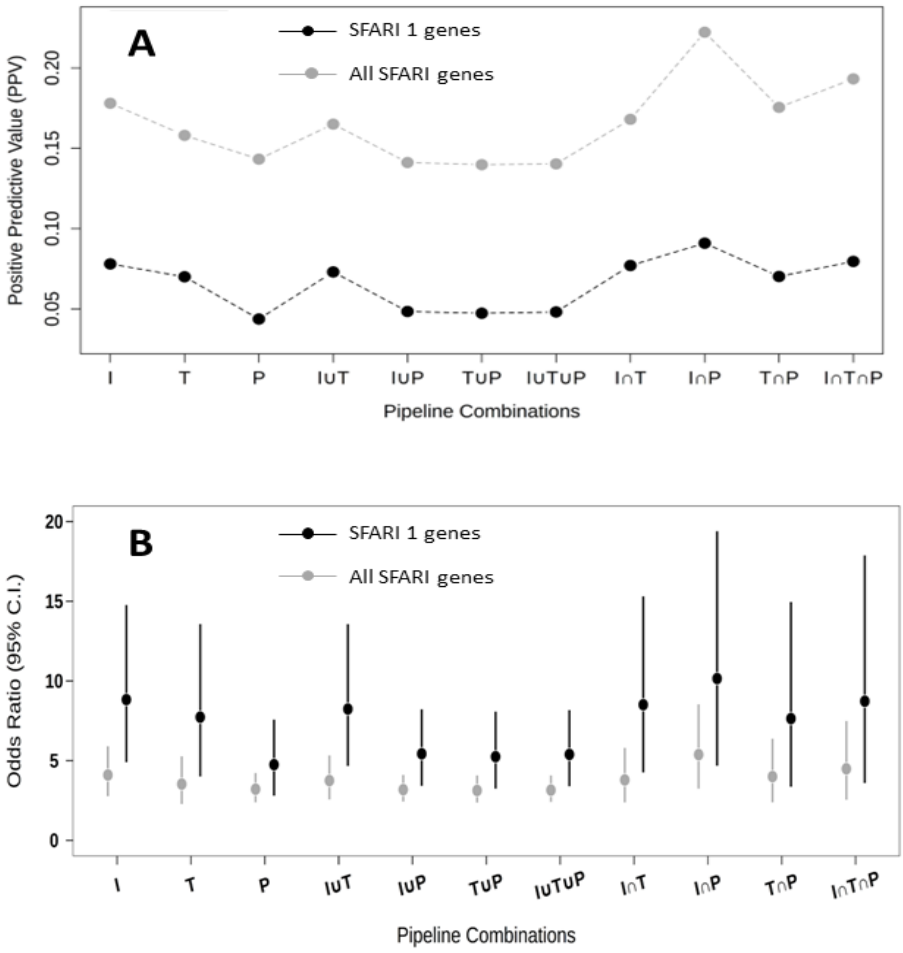

Fig. 2 Effectiveness of *InterVar (I)*, *TAPES (T)*, *Psi-Variant (P)*, and their combinations in detecting candidate variants in ASD genes. **A** Positive predictive value (PPV) of detecting candidate variants in SFARI 1 and all SFARI genes. **B** Odds Ratios (ORs) of detecting candidate variants in SFARI 1 and all SFARI genes.

Discussion

In this study, we assessed the concordance and effectiveness of three bioinformatics tools in the interpretation of SNVs detected in WES of children with ASD. There was better agreement

in variant detection between *InterVar* and *TAPES* than between *Psi-Variant* and each of these two tools, probably because both *InterVar* and *TAPES* are based on the ACMG/AMP guidelines [8], while *Psi-Variant* uses the interpretation of six in-silico tools in assessing the functional consequence of LGD SNVs. In addition, most (94%) of the SNVs detected by either *InterVar* or *TAPES* were de-novo SNVs, compared to only 36% of the SNVs detected by *Psi-Variant*. This difference may be attributed to the fact that ACMG/AMP guidelines are particularly designed to detect de-novo highly penetrant SNVs, while inherited SNVs (autosomal recessive and X-linked) are usually classified as VUS [19]. Importantly, such rare inherited SNVs have been found to be associated with a variety neurodevelopmental conditions, including ASD [9,16,17,20,21]. Another major difference between these tools lies in the detection of in-frame insertions/deletions that comprised ~20% of the SNVs detected by either *InterVar* or *TAPES*, while such SNVs were discarded by *Psi-Variant*. We decided to exclude these SNVs from *Psi-Variant* because their clinical relevance has been demonstrated in several genetic disorders [36,37] but not in ASD [38–40].

Another important factor that could affect the concordance between the three tools is the annotation tools they use. Specifically, both *InterVar* and *TAPES* use AnnoVar [41] for their variant annotation, while *Psi-Variant* uses Ensembl's VEP [25]. It has already been shown that AnnoVar and VEP have a low concordance in the classification of LoF SNVs [42]. In addition, each tool, *InterVar*, *TAPES*, and *Psi-Variant*, utilizes a different set of in-silico tools for the classification of missense SNVs, with SIFT [28] alone being shared by all three tools. These differences are probably the reason for the large differences in the detection of missense SNVs between the three tools (Table 1).

Today, there are no accepted guidelines for the detection of ASD susceptibility SNVs from WES data. Therefore, many genetic labs use the ACMG/AMP guidelines [8] for this purpose. Our findings suggest that combining these guidelines with other variant classification criteria

improves detection of ASD susceptibility SNVs. Thus, this approach could be applied in the future for drawing up specific guidelines for the detection of ASD susceptibility SNVs.

Of note, 77.5% of the SNVs detected by the most effective integrative pipeline ($\underline{I} \cap \underline{P}$) affect genes with no known association to ASD, according to the SFARI Gene database [35]. This finding highlights the capability of the integrative pipeline ($\underline{I} \cap \underline{P}$) to detect of novel ASD genes. Obviously, the association of these genes and SNVs with ASD susceptibility needs to be validated in additional studies.

The results of this study should be considered under the following limitations. First, the effectiveness assessments of the different tools and their combinations were based on ASD genes from the SFARI Gene database [35]. While this is the most commonly used database for ASD genes and is continuously updated, it is based on data curated from the literature and may thus include genes that were falsely associated with ASD. Second, the variant detection analyses were performed on WES data of a cohort from the Israeli population, which may not necessarily be representative of the genetic architecture of ASD. Third, the tools used in this study were designed to detect only extremely rare SNVs with relatively large functional effects. Thus, a more effective approach for detection of ASD susceptibility variants should also include interpretation of copy-number variants [43–46] and other variants with milder functional effects [16,47,48] Finally, it should be noted that there are many other approaches for variant interpretation from WES data. Thus, it is possible that combinations of other approaches would be more effective in the detection of ASD susceptibility variants from WES data than the approaches investigated in this study.

Conclusions

Our findings suggest that an integration of different variant interpretation approaches is more effective in the detection of ASD candidate SNVs from WES data than each of the examined

approaches alone. Inclusion of additional criteria to this integrative approach may further improve its effectiveness in the detection of ASD candidate variants.

List of abbreviations

ACMG/AMP: American College of Medical Genetics and Genomics/Association of Molecular Pathology; ASD: autism spectrum disorder; C.I.: confidence interval; GATK: Genome Analysis Toolkit; LGD: likely gene disrupting; LoF: loss of function; LP: likely pathogenic; ML: machine learning; NADI: National Autism Database in Israel; NGS: next-generation sequencing; OR: odds ratio; P: pathogenic; PPV: positive predictive value; SNV: single nucleotide variants; VEP: Variant Effect Predictor; vcf: variant calling format; VUS: variants of uncertain significance; WES: whole exome sequencing.

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

Informed consent was obtained from all the families 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. The generated results and codes are available in a GitHub public repository: https://github.com/AppWick-hub/Psi-Variant or available upon reasonable request to the corresponding author Prof. Idan Menashe (idanmen@bgu.ac.il).

Competing interests

The authors declare no competing interests.

Funding

This study was supported by a grant from 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*: A.S. and I.M.; *formal analysis*: A.S.; *resources*: H.G., G.M., A.M., Y.T., A.A. and I.D.; *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.

Acknowledgements

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

Reference

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edn. Am Psychiatr Publ. 2013;
2. Meng-Chuan Lai, Michael V Lombardo SB-C. Autism. Lancet. 2014;
3. Yoo H. Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications. Exp Neurobiol. 2015;24(4):257.
4. Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Lancet. 2018;392(10146):508–20.
5. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156(3):255–74.
6. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. Arch Gen Psychiatry. 2011;68(11):1095–102.
7. Devlin B, Boone BE, Levy SE, Lihm J, Buxbaum JD, Wu Y, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature [Internet]. 2012;485(7397):242–6. Available from: http://dx.doi.org/10.1038/nature11011
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Med. 2015;17(5):405–24.
9. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci [Internet]. 2019;22(12):1961–5. Available from: http://dx.doi.org/10.1038/s41593-019-0527-8
10. Fu JM, Satterstrom FK, Peng M, Brand H, Collins RL, Dong S, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat Genet. 2022;54(September).
11. Satterstrom FK, Kosmicki JA, Wang J, Roeder K, Daly MJ, Buxbaum JD. 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. 2020;1–17.
12. Wu D, Dou J, Chai X, Bellis C, Wilm A, Shih CC, et al. Large-Scale Whole-Genome Sequencing of Three Diverse Asian Populations in Singapore. Cell. 2019;179(3):736-749.e15.
13. Feliciano P, Zhou X, Astrovskaya I, Turner TN, Wang T, Brueggeman L, et al. Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. npj Genomic Med. 2019;4(1).
14. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature [Internet]. 2014;515(7526):209–15. Available from: http://dx.doi.org/10.1038/nature13772
15. Ishay RT, Shil A, Solomon S, Sadigurschi N, Abu-kaf H, Meiri G, et al. Diagnostic Yield and Economic Implications of Whole-Exome Sequencing for ASD Diagnosis in Israel. 2022;
16. Wang T, Zhao PA, Eichler EE. Rare variants and the oligogenic architecture of autism. Trends Genet [Internet]. 2022;1–9. Available from: https://doi.org/10.1016/j.tig.2022.03.009
17. Doan RN, Lim ET, De Rubeis S, Betancur C, Cutler DJ, Chiocchetti AG, et al. Recessive gene disruptions in autism spectrum disorder. Nat Genet [Internet]. 2019;51(July). Available from: http://dx.doi.org/10.1038/s41588-019-0433-8
18. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet [Internet]. 2017;100(2):267–80. Available from: http://dx.doi.org/10.1016/j.ajhg.2017.01.004
19. Houge G, Laner A, Cirak S, de Leeuw N, Scheffer H, den Dunnen JT. Stepwise ABC system for classification of any type of genetic variant. Eur J Hum Genet. 2022;30(2):150–9.
20. Wilfert AB, Turner TN, Murali SC, Hsieh PH, Sulovari A, Wang T, et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. Nat Genet. 2021;53(8):1125–34.
21. Halvorsen M, Samuels J, Wang Y, Greenberg BD, Fyer AJ, McCracken JT, et al. Exome sequencing in obsessive–compulsive disorder reveals a burden of rare damaging coding variants. Nat Neurosci [Internet]. 2021;24(8):1071–6. Available from: http://dx.doi.org/10.1038/s41593-021-00876-8
22. Dinstein I, Arazi A, Golan HM, Koller J, Elliott E, Gozes 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. 2020;70(9):1303–12.
23. Meiri G, Dinstein I, Michaelowski A, Flusser H, Ilan M, Faroy M, et al. Brief Report: The Negev Hospital-University-Based (HUB) Autism Database. J Autism Dev Disord. 2017;47(9):2918–26.

24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297.
25. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol [Internet]. 2016;17(1):1–14. Available from: http://dx.doi.org/10.1186/s13059-016-0974-4
26. Xavier A, Scott RJ, Talseth-Palmer BA. TAPES: A tool for assessment and prioritisation in exome studies. PLoS Comput Biol [Internet]. 2019;15(10):1–9. Available from: http://dx.doi.org/10.1371/journal.pcbi.1007453
27. Fadista J, Oskolkov N, Hansson O, Groop L. LoFtool: A gene intolerance score based on loss-of-function variants in 60 706 individuals. Bioinformatics. 2017;33(4):471–4.
28. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812–4.
29. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Vol. 2, Current Protocols in Human Genetics. 2013.
30. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886–94.
31. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet [Internet]. 2016;99(4):877–85. Available from: http://dx.doi.org/10.1016/j.ajhg.2016.08.016
32. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48(12):1581–6.
33. Samocha KE, Kosmicki JA, Karczewski KJ, O'Donnell-Luria AH, Pierce-Hoffman E, MacArthur DG, et al. Regional missense constraint improves variant deleteriousness prediction. bioRxiv. 2017;
34. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med. 2020;12(1):1–8.
35. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. 2013;4(1):2–4.
36. Sergouniotis PI, Barton SJ, Waller S, Perveen R, Ellingford JM, Campbell C, et al. The role of small in-frame insertions/deletions in inherited eye disorders and how structural modelling can help estimate their pathogenicity. Orphanet J Rare Dis [Internet]. 2016;11(1):1–8. Available from: http://dx.doi.org/10.1186/s13023-016-0505-0
37. Sallah SR, Sergouniotis PI, Hardcastle C, Ramsden S, Lotery AJ, Lench N, et al. Assessing the Pathogenicity of In-Frame CACNA1F Indel Variants Using Structural Modeling. J Mol Diagnostics [Internet]. 2022;24(12):1232–9. Available from: https://doi.org/10.1016/j.jmoldx.2022.09.005
38. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron [Internet]. 2012;74(2):285–99. Available from: http://dx.doi.org/10.1016/j.neuron.2012.04.009
39. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, et al. De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. Cell Rep [Internet]. 2014;9(1):16–23. Available from: http://dx.doi.org/10.1016/j.celrep.2014.08.068
40. Kopp N, Amarillo I, Martinez-Agosto J, Quintero-Rivera F. Pathogenic paternally inherited NLGN4X deletion in a female with autism spectrum disorder: Clinical, cytogenetic, and molecular characterization. Am J Med Genet Part A. 2021;185(3):894–900.
41. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):1–7.
42. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of transcripts and software has a large effect on variant annotation. Genome Med. 2014;6(3).
43. Husson T, Lecoquierre F, Cassinari K, Charbonnier C, Quenez O, Goldenberg A, et al. Rare genetic susceptibility variants assessment in autism spectrum disorder: detection rate and practical use. Transl Psychiatry [Internet]. 2020;10(1). Available from: http://dx.doi.org/10.1038/s41398-020-0760-7
44. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, et al. Genomic Patterns of De Novo Mutation in Simplex Autism. Cell [Internet]. 2017;171(3):710-722.e12. Available from: https://doi.org/10.1016/j.cell.2017.08.047
45. Leppa VMM, Kravitz SNN, Martin CLL, Andrieux J, Le Caignec C, Martin-Coignard D, et al. Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. Am J Hum Genet [Internet]. 2016;99(3):540–54. Available from: http://dx.doi.org/10.1016/j.ajhg.2016.06.036

46. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47(6):582–8.
47. Du Y, Li Z, Liu Z, Zhang N, Wang R, Li F, et al. Nonrandom occurrence of multiple de novo coding variants in a proband indicates the existence of an oligogenic model in autism. Genet Med. 2020;22(1):170–80.
48. Guo H, Duyzend MH, Coe BP, Baker C, Hoekzema K, Gerdts J, et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. Genet Med [Internet]. 2019;21(7):1611–20. Available from: http://dx.doi.org/10.1038/s41436-018-0380-2