Using Artificial Intelligence for Implementing New Recommendations of the PVS1 Loss of Function Criterion

Yaron Einhorn, MSc;[1] Adaia Kamshov, MD, MSc;[1] Moshe Einhorn, MSc;[1] Nurit Paz-Yaacov, PhD;[1] Oron Lev, MSc;[1] Yuval Yaron, MD;[1,2] Susan Gross, MD, FACMG;[1,3] Elaine Lyon, PhD, FACMG[1,4]

[1]Genoox, Tel Aviv, Israel; [2]Tel Aviv Sorasky Medical Center, Tel Aviv, Israel; [3]Icahn School of Medicine at Mount Sinai, New York, NY, USA; [4]Clinical Services Laboratory, HudsonAlpha Institute of Biotechnology, Huntsville, AL, USA

ABSTRACT

Background: The 2015 ACMG/AMP sequence variant interpretation guidelines provided a framework for classifying variants based on several benign (B) and pathogenic (P) evidence criteria, including a pathogenic criterion (PVS1) for predicted loss of function (LOF) variants. Recently, the ClinGen Sequence Variant Interpretation (SVI) Workgroup published new recommendations for more accurate interpretations of null variants. These recommendations take into consideration a variety of evidences, including but not limited to, the functional role of the affected region, the size of the region, the type of variant, and the variant location on the transcript, to develop a decision tree for assigning a final weight of evidence for a P classification based on the PVS1 rule.

Aim: We hypothesized than an artificial intelligence (AI)-based algorithm can be an efficient tool for implementing the ClinGen SVI Workgroup recommendations and sought to benchmark a novel AI-based Variant Classification Engine (aiVCE), using variants assessed by the ClinGen SVI Workgroup and null variants in the ClinVar database.

Methods: We benchmarked the aiVCE against the same 56 variants employed by the ClinGen SVI Workgroup (Hum Mutat 2018;39:1517-1524. doi: 10.1002/humu.23626)). The aiVCE was further examined using 16,333 null variants from the ClinVar database (version 01-11-18).

Results: The aiVCE demonstrated a high level of concordance with the ClinGen SVI Workgroup, meeting the PSV1 rule for 56/56 (100%) variants and for 51/56 (91%) variants with the same strength of classification (Table 1). All 5 variants discordant as to classification strength owed the disagreement as to whether the truncated/altered region identified is critical to the protein function. In all cases, the regions were classified by the ClinGen SVI Workgroup as those with an unknown function, which resulted in a 'moderate' strength, while the aiVCE suggested the region is critical to the protein function, yielding a 'strong' rule strength. The aiVCE algorithm for defining these regions as critical to protein function were consistent with the ClinGen SVI Workgroup's recommendations due to the number of non-truncating P variants within the region and number of P variants within and/or downstream to the exon. In the ClinVar null variant experiment (Table 2), while most of the ClinVar P variants were designated as having 'very strong' evidence for being P according to the PSV1 rule, most of the ClinVar B and uncertain variants did not meet the PVS1 rule. The 229 ClinVar uncertain variants were designated by the aiVCE as having 'very strong' evidence for P according to the revised ClinGen SVI Workgroup's PSV1 rule recommendations.

Conclusions: The aiVCE algorithm demonstrated excellent concordance with the ClinGen SVI Workgroup results. Further, the 5 variants demonstrating discordant classification strength illustrate an advantage of using an AI-based tool, as the identification of these variants as P were in agreement with the new ClinGen SVI Workgroup recommendations. Implementation of new recommendations for more accurate interpretation of null variants involves demanding and complex bioinformatics work, and AI-based solutions can enhance current interpretation guidelines.

Introduction

American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) Standards and Guidelines – PVS1 Criteria[1]

- The 2015 ACMG/AMP sequence variant interpretation guidelines provided a framework for classifying variants based on several benign (B) and pathogenic (P) evidence criteria
  - Includes a pathogenic criterion for predicted loss of function (LOF) variants (PVS1)
  - No criterion-specific guidance for implementation was provided
- Recently, the ClinGen Sequence Variant Interpretation (SVI) Workgroup published new recommendations for more detailed interpretations of null variants[2]
  - Recommendations take into consideration a variety of evidences (e.g., functional role of affected region, size of region, type of variant, variant location on transcript)
  - According to ClinGen SVI Workgroup recommendations,[2] nonsense or frameshift variants are evaluated for:
    - potential to undergo nonsense-mediated decay
    - presence/absence from biologically-relevant transcript(s)
    - critical nature to protein function/amount of protein removed
    - frequency within the general population

    to assign a modified final weight of evidence for a P classification based on the PVS1 rule as:
    - pathogenic very strong (PVS)
    - pathogenic strong (PS)
    - pathogenic moderate (PM)
    - pathogenic supporting (PP)

Clinical Genome Resource (ClinGen) Expert Gene/Disease Panels (EPs)[2]

- EPs tasked with defining application of the ACMG/AMP guidelines for sequence variant interpretation in specific genes or diseases
- Recently, FDA recognized the genetic variant information in the Clinical Genome Resource (ClinGen) consortium's ClinGen Expert Curated Human Genetic Data as a source of valid scientific evidence that can be used to support clinical validity[3]
- The unique database was employed to assess the artificial intelligence-based Variant Classification Engine (aiVCE) at the rule level and compare the aiVCE performance to EP decisions

Artificial Intelligence-Based Variant Classification Engine (aiVCE)

- Data-driven; based on the ACMG/AMP sequence variant classification guidelines
- Automates majority of ACMG/AMP classification rules
- Sequence variant classification is accomplished by building prediction models at the gene and rule levels, based on various data sources (e.g., ClinVar, ClinGen, Uniprot, gnomAD, ExAC, Orphanet, etc.)
- Classification takes into account the gene and diseases associated with the variant
- Professional expertise can be applied to algorithm to determine thresholds specific to the gene being interrogated

Aims

- Evaluate the aiVCE's ability to determine whether PSV1 should be applied to null variants, and its strength, via comparisons against two datasets:
  1) Variants assessed by the ClinGen SVI Workgroup
  2) 16,333 ClinVar null variants with known classification

Methods

- Recent recommendations of the SVI workgroup were incorporated as part of the aiVCE methods for determining whether PSV1 criterion is met and its strength
- aiVCE was benchmarked using two separate datasets
  - Same 56 variants employed by the ClinGen SVI Workgroup[4]
  - All 16,333 null variants in ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/); version 01-11-18, applying the ClinGen SVI Workgroup recommendations for PSV1

Results

ClinGen SVI Workgroup Variants (Table 1)

- The aiVCE demonstrated a high level of concordance with the ClinGen SVI Workgroup recommendations, meeting the PSV1 rule for:
  - 56/56 (100%) all variants (overall classification)
  - 51/56 (91%) variants with the same strength of classification
- 5 variants with higher classification strength:
  - Disagreement as to whether the truncated/altered region identified is critical to protein function
  - For all, regions were classified by the ClinGen SVI Workgroup as those with an unknown function, resulting in a 'moderate' rule strength, while the aiVCE suggested regions were critical to the protein function, yielding a 'strong' rule strength
  - The aiVCE algorithm defined these regions as critical to protein function based on the number of non-truncating P variants within the region and number of P variants within and/or downstream to the exon

Table 1. Benchmarking the aiVCE vs. ClinGen SVI Workgroup variants: PVS1 rule strength

ClinGen SVI / *aiVCE*	PP	PM	PS	PVS
PP	0	0	0	0
PM	0	**9**	0	0
PS	0	**5**	**7**	0
PVS	0	0	0	**35**

Application of ClinGen SVI Workgroup Recommendations for PSV1 to ClinVar Null Variants using aiVCE (Figure 1)

- Most (>95%) of the ClinVar P variants met the PVS1 criteria with:
  - >90% having 'very strong' or 'strong' strength
  - Only 5.7% of P variants did not meet the rule at all
- Most (75.2%) of the ClinVar B variants did not meet the PVS1 rule
- The distribution between met (64.8%) and unmet (35.2%) was less consistent for VUS
  - B vs. P results: $p < 0.0001$ (Chi-square test)
- 229 of the 1,586 (14.4%) ClinVar VUS were designated by the aiVCE as having 'very strong' evidence for P according to the revised ClinGen SVI Workgroup's PSV1 rule recommendations

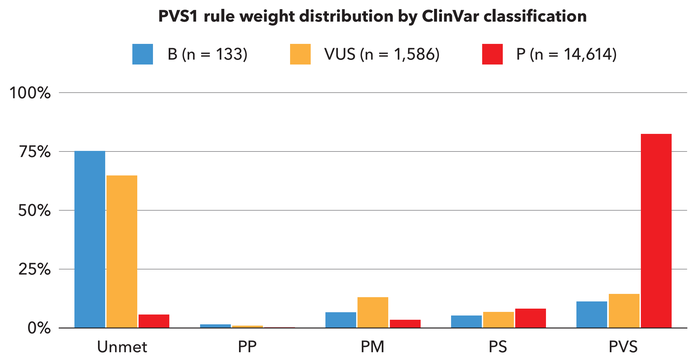

Figure 1. ClinGen SVI Workgroup recommendations for PSV1: application to ClinVar null variants (N=16,333) using the aiVCE

CONCLUSIONS

- Benchmarking against ClinVar data demonstrates the robustness of the algorithm ($p < 0.0001$) across thousands of variants
- The aiVCE algorithm demonstrated excellent concordance with the ClinGen SVI Workgroup results
- The 5 variants with higher classification strength illustrate an advantage of using an automated tool that identifies critical regions based on locations of other known P variants in agreement with the new ClinGen SVI Workgroup recommendations
- Implementation of new recommendations for more accurate interpretation of null variants involves demanding and complex bioinformatics work; AI-based solutions can enhance current interpretation guidelines

References

1. 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 Med* 2015;17:405-24.
2. Rivera-Muñoz EA, e al. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat* 2018;39:1614-22.
3. United States Food and Drug Administration. FDA takes new action to advance the development of reliable and beneficial genetic tests that can improve patient care. December 4, 2018 https://www.fda.gov/NewsEvents/Newsroom/ PressAnnouncements/ucm627555.htm
4. Abou Tayoun AN, et al. Recommendations for interpreting loss of function PSV1 ACMG/AMP variant criterion. *Hum Mutat* 2018;39:1517-24.