Genoox

Reinterpretation of Sequence Variants Using Artificial Intelligence – Results of 2 Benchmarking Experiments

Yaron Einhorn, MSc;[1] Adaia Kamshov, MD, MSc;[1] Nurit Paz-Yaacov, PhD;[1] Moshe Einhorn, MSc;[1] Steven Harrison, PhD;[2] Yuval Yaron, MD[1,3]

[1]Genoox, Tel Aviv, Israel; [2]Broad Institute of MIT/Harvard, Cambridge, MA, USA; [3]Tel Aviv Sorasky Medical Center, Tel Aviv, Israel

ABSTRACT

Introduction: Reinterpretation of sequence variants is an important task of diagnostic laboratories, because variant classification may change over time as new evidence is accrued. Recent publications estimate the time required for reinterpretation at 1-2 hours per variant, thus creating a daunting bottleneck (Harrison S. et al., *Human Mutat* 2018;39:1641-9). Some variants, are exceptionally challenging to classify, owing to contradicting evidence of their pathogenicity. To address challenges in variant interpretation, a novel Artificial Intelligence (AI) Variant Classification Engine (aiVCE) was developed, based on the American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines for the interpretation of sequence variants. The aiVCE is motivated by the work of recent ClinGen expert groups, which specify assertion criteria based on the unique features of particular genes or domains of interest for each variant. In addition, the aiVCE is able to resolve contradicting classifications submitted to ClinVar.

Methods: The utility of this aiVCE was evaluated against two recently published datasets of variants submitted to ClinVar (Harrison S. et al., *Human Mutat* 2018;39:1641-9). The first dataset included 21,199 variants with concordant classification among at least 2 of 41 selected laboratories (Concordant Dataset). Because the aiVCE relies on ClinVar consensus, PP5/BP6 rules ("reputable source recently reports variant as pathogenic/benign") were not taken into account. The second dataset included 127 variants with outlier classifications of medically significant differences (MSDs), which were ultimately resolved following reassessment prompted by that study (Discordant Dataset). In both assessments, variants were grouped as "actionable" (pathogenic+likely pathogenic [P+LP]) vs "non-actionable" (variant of uncertain significance+likely benign+benign [VUS+LB+B]).

Results: In the Concordant Dataset, 19,816/21,199 (93.5%) variants were classified by the aiVCE in agreement with ClinVar when comparing "actionable" vs "non-actionable" classifications (Table 1). Of the variants that were not in agreement, no critical discrepancies in classification were detected, e.g., no P or LP variants were classified as B or LB by the aiVCE. In the Discordant Dataset, when assessing the 127 variants with MSDs that were resolved after reassessment, 94 (77.2%) were classified by the aiVCE in concordance with the published post-reassessment classification (Table 2). All 33 remaining variants were classified by the aiVCE in concordance with the outlier classification. Of these, two variants (CHEK2_NM_007194.3:c.470T>C, HFE_NM_000410.3:c.845G>A) were detected by the aiVCE as "Risk susceptibility variants" and therefore, despite of their high frequency, all B frequency rules (BS1, BS2, BA1) were disabled automatically. The majority of variants classified as P/LP in ClinVar and VUS by the aiVCE were due to ACMG criteria that cannot be automatically applied, such as most of the criteria related to case-level data.

Conclusions: These benchmarking results demonstrate the aiVCE's ability to classify variants at a high rate of concordance with expert curation. The high level of agreement with challenging variants, such as those in the Discordant Dataset, makes the aiVCE ideal to expedite reinterpretation of variants accumulated over time. The aiVCE can be designed to monitor and alert the lab of variants with significant classification changes at predetermined time intervals.

Introduction

Reinterpretation of Sequence Variants

- Important task of diagnostic laboratories, given that variant classification may change over time as new variant evidence is accrued
- Estimated time required for reinterpretation –
  - 1–2 hours per variant
  - Some variants exceptionally challenging to classify, owing to contradicting evidence of their pathogenicity
  - Reinterpretation creates significant backlog in diagnostic laboratories[1]
- Recently, Harrison, et al.[2] facilitated resolution of classification differences on a large scale by encouraging clinical laboratories to reassess outlier classifications of variants with medically significant differences (MSDs)
  - Several datasets were evaluated, yielding 127 variants with discordant classification among ≥3 laboratories that were resolved due to reassessment prompted by study

Artificial Intelligence-Based Variant Classification Engine (aiVCE)

- Data-driven; based on the American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) Standards and Guidelines for sequence variant classification
- Automates majority of ACMG/AMP classification rules
- Variant classification 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

- Benchmark the utility of the aiVCE as a reinterpretation tool by performing a retrospective study using challenging variants for which there were originally contradicting interpretations
- Assess aiVCE robustness by benchmarking against large number of variants for which there is consensus between all submitters in ClinVar

Methods

Benchmarking Experiments

- Employed two recently published datasets of variants submitted to ClinVar[2]
- Consistent with Harrison, et al., 2019,[2] our analyses used the ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) April 2017 database to identify all variants in ClinVar with submissions from ≥2 of the selected 41 clinical laboratories
- Dataset #1 – 'Concordant' = 21,199 variants with concordant classification among ≥2 of the 41 laboratories
  - PP5/BP6 rules ("reputable source recently reports variant as pathogenic/benign") not taken into account
- Dataset #2 – 'Discordant' = 127 variants with outlier classifications of medically significant differences (MSDs) where ≥3 laboratories classified the variant
  - Considered challenging variants, as there were MSDs in their interpretation
  - Ultimately resolved following laboratory reassessment
- Three major classification levels (P/LP, VUS, and LB/B) were employed; classifications for each variant were compared to determine concordance or discordance
- Variants were grouped as "actionable" (P+LP) vs. "non-actionable" (VUS+LB+B)

Results

Concordant Dataset (Tables 1 and 2)

- aiVCE classified 19,816/21,199 (93.5%) variants in agreement with ClinVar when comparing "actionable" vs "non-actionable" classifications
- Of 1,383/21,199 (6.5%) variants not in agreement, no critical discrepancies (e.g., no P or LP variants classified as B or LB by the aiVCE or vice-versa) were identified

Table 1. Benchmarking aiVCE using the ClinVar 'Concordant' dataset

ClinVar / aiVCE	Actionable	Non-Actionable
Actionable	3,542 (16.7%)	1,272 (6.0%)
Non-Actionable	111 (0.5%)	16,274 (76.8%)

Table 2. Benchmarking aiVCE using the ClinVar 'Concordant' dataset – classification distribution

ClinVar / aiVCE	P	VUS	B
P	3542	1272	0
VUS	111	2735	1336
B	0	2481	9722

Discordant Dataset (Tables 3 and 4)

- aiVCE classified 94/127 (77.2%) variants with MSDs, that were resolved after reassessment, in agreement with the published post-reassessment classification
- aiVCE classified remaining 33/127 in agreement with the outlier classification
  - Of these, two variants (CHEK2_NM_007194.3:c.470T>C, HFE_NM_000410.3:c.845G>A) were detected by the aiVCE as "Risk susceptibility variants;" therefore, despite of their high frequency, all B frequency rules (BS1, BS2, BA1) were disabled automatically
  - The majority of variants classified as P/LP in ClinVar and VUS by the aiVCE were mainly due to:
    - ACMG criteria that cannot be automatically applied, e.g., criteria related to case-level data
    - Incomplete penetrance genes such as BRCA1/2, GJB2, etc., for which current guidelines have limited applicability
    - Exclusion of PP5/BP6 rules
  - Interestingly, even though aiVCE only automates some of the ACMG classification rules, 127/127 (100%) variant classifications were in agreement with ≥1 of the laboratories

Table 3. Benchmarking aiVCE using the ClinVar 'Discordant' dataset of MSDs resolved after laboratory reassessment

Lab assessment / aiVCE	Actionable	Non-Actionable
Actionable	32 (25.2%)	26 (20.5%)
Non-Actionable	3 (2.4%)	66 (52.0%)

Table 4. Benchmarking aiVCE using the ClinVar 'Discordant' dataset of MSDs resolved after laboratory reassessment – classification distribution

ClinVar / aiVCE	P	VUS	B
P/LP	**32**	23	3
VUS	3	**42**	4
VUS/B	0	**5**	**4**
LB/B	0	0	**11**

CONCLUSIONS

- Variant classification can diverge among laboratories and evolve over time, mainly due to new/additional evidence associated with the case. As such, variant reinterpretation is of great importance
- Benchmarking results demonstrate the aiVCE's ability to classify variants at a high rate of concordance with expert curation
- 100% of the discordant variants were classified consistently with ≥1 of the laboratories
- AI-based tools can greatly reduce the turnaround-time associated with classifying unambiguous variants and thus allow more time for the manual assessment required for challenging cases
- The aiVCE can be designed to monitor and alert laboratories of variants with significant classification changes at predetermined time intervals

References

1. Niehaus A, et al. A survey assessing adoption of the ACMG-AMP guidelines for interpreting sequence variants and identification of areas for continued improvement. *Genet Med* 2019; Jan 23. doi: 10.1038/s41436-018-0432-7.
2. Harrison SM, et al. Scaling resolution of variant classification differences in ClinVar between 41 clinical laboratories through an outlier approach. *Hum Mutat* 2018;39:1641-1649.