Benchmarking an Automated Variant Classification Engine (aVCE) Algorithm Using ClinVar: Results of a Time-Capsule Experiment

Yaron Einhorn, MSc[1]; Oron Lev, MSc[1]; Moshe Einhorn, MSc[1]; Amir Trabelsi, BSc[1]; Nurit Paz-Yaacov, PhD[1]; Susan J. Gross, MD[1,2]

[1]GENOOX, Tel Aviv, Israel; [2]Icahn School of Medicine at Mount Sinai, New York NY, USA

ACCEPTED ABSTRACT

Introduction: DNA sequencing technology has evolved rapidly with the advent of high-throughput next-generation sequencing (NGS). To address challenges in NGS interpretation, a novel algorithm, which integrates human DNA sequences with phenotyping, has been developed, based on the American College of Medical Genetics and Genomics (ACMG) standards and guidelines (Richards S, et al. *Genet Med* 2015;17:405-24). Current guidelines published jointly by the Association for Molecular Pathology (AMP) and College of American Pathologists (CAP) strongly advocate for validation of pipeline tools and algorithms (Roy S, et al. *J Molecular Diag* 2017;doi: 10.1016/j.jmoldx.2017.11.00). To validate this novel automated Variant Classification Engine (aVCE), we performed a blinded time-capsule experiment to predict the ability of this algorithm to classify variants that were only uploaded to the ClinVar database after the time capsule cutoff date.

Methods: The ClinVar database is a publicly available archive of reports that details relationships among human variations and phenotypes, with supporting evidence. The aVCE was 'trained' on the ClinVar database (version 30-06-17). Variants with Reference/Submission ClinVar (RCV/SCV) creation dates before and after 01-07-16 were marked as 'Train' and 'Test,' respectively. Variants with ≥2 ClinVar stars were included in the 'Test' set. Using ACMG standards and guidelines for interpreting sequence variants, the aVCE was applied to the 'Test' set to classify variants as pathogenic (*P*), likely pathogenic (*LP*), uncertain significance (*VUS*), likely benign (*LB*), and benign (*B*). In accordance with the ACMG standards and guidelines, the aVCE algorithm has additional tiers for subclassification of *VUS* into 'variant of uncertain significance, leaning benign (*VUS-LB*), weak leaning pathogenic (*VUS-WLP*), and strong leaning pathogenic (*VUS-SLP*). Results also were characterized from a clinical perspective, i.e., clinically 'actionable' (*P/LP*) versus 'non-actionable' (*VUS/LB/B*) variants and benchmarked against the ClinVar classifications to determine performance characteristics (sensitivity and specificity).

Results: When compared against ClinVar submissions from clinical laboratories and high-certainty entries, the proprietary aVCE classified clinically 'actionable' (*P/LP*) and 'non-actionable' (*VUS/LB/B*) variants with very high sensitivity (99.29%, 1262/1271) and specificity (100%).

Conclusions: The aVCE algorithm, even without input from clinical databases specific to the 'Test' set, could predict with very high sensitivity and specificity whether a variant in the future would be categorized as clinically 'actionable' versus 'non-actionable.' Algorithms that apply the latest computational methodologies to ACMG guidelines may assist variant scientists with classification and interpretation of variants, including those with limited clinical information.

INTRODUCTION

American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) 2015 Standards and Guidelines for Variant Classification[1]

- Harmonize methods
- Reduce ambiguity between clinical laboratories
- Weighted rules related to:
  - Variant frequency
  - Variant type
  - Association to previous reports for pathogenicity
  - Consistency with inheritance model
- Require accessing/searching of multiple databases

Potential Benefits of Advanced Computational Methodologies

- Aid scientists in accurately applying ACMG-AMP standards
- Streamline data extraction related to phenotype, molecular sequence, and variant characteristics from existing databases
- Efficiently assimilate information from published reports of clinical aspects of variant classification
- Systematically and continually update information
- Remove current roadblocks in classifying variants, including automation of database and bioinformatics management

Novel automated Variant Classification Engine (aVCE)

- Based on ACMG-AMP standards and guidelines
- Utilizes AI technology
- Integrates knowledge acquired from multiple databases and published literature on an ongoing basis
- Determines internal numeric classification score to facilitate VUS subclassification

OBJECTIVES

Primary

- Validate the aVCE by performing a blinded time-capsule experiment to predict the ability of this algorithm to classify variants that were only uploaded to the ClinVar database after the time capsule cutoff date

Secondary

- Discern reasons underlying incongruence between aVCE and ClinVar
- Uncover areas in current classification guidelines that may benefit from further research

METHODS

automated Variant Classification Engine (aVCE)

- Provide automatic implementation of ACMG classification rules per currently available:
  - Population, disease, sequence databases
  - Published literature
- Classify variants as:
  - *B* Benign
  - *LB* Likely Benign
  - *VUS* Variant of Uncertain Significance
    - *VUS-LB* – Variant is classified as *VUS* according ACMG guidelines. More evidence found to support the variant as being *B*
    - *VUS-WLP* – Weak evidence for *P* but not enough for being classified as *LP* according to ACMG-AMP guidelines
    - *VUS-SLP* – Strong evidence for *P* but not enough for being classified as *LP* according to ACMG-AMP guidelines

    Optional *VUS* subclassification – Some laboratories choose to subclassify *VUS*, particularly for internal use, a practice not considered inconsistent with ACMG-AMP standards and guidelines
  - *LP* Likely Pathogenic
  - *P* Pathogenic

Validation Experiment

- ClinVar database
  - All normalized variants with Reference Accession Version (RCV) or Submission Accession Version (SCV) creation dates before 01-01-17 employed for building/training the aVCE
  - All other variants not overlapping with 'Training" dataset, including those with RCV/SCV creation dates after 01-07-16, were considered the 'Test' dataset for aVCE benchmarking
  - To avoid false positives in the 'Test' dataset, variants with <2 ClinVar scoring stars were removed, as were *VUS*[2]
- aVCE applied to the 'Test' dataset for variant classification
  - (*P*, *LP*, *VUS-SLP*, *VUS-WLP*, *VUS*, *VUS-LB*, *LB*, *B*)
- aVCE results characterized clinically:
  - 'Actionable' versus 'Non-actionable'[3]
    - 'Actionable' = *P* + *LP*
    - 'Actionable' = *P* + *LP* + *VUS-SLP*
    - All other variants considered 'Non-actionable'
  - Sensitivity/specificity of aVCE versus ClinVar database

RESULTS

A. aVCE Performance Characteristics

- Final dataset

All Variants, N = 1,689

'*Actionable*' variants, n = 1,271 | '*Non-actionable*' variants, n = 418

- aVCE demonstrated robust sensitivity and specificity in classifying variants that were only uploaded to the ClinVar database after the time capsule cutoff date **(Tables 1 and 2)**.

Table 1. Benchmarking an automated Variant Classification Engine (aVCE) using a time capsule of the ClinVar database

	aVCE	*B*	*LB*	*VUS*	*LP*	*P*
ClinVar						
B		164	9	167	0	0
LB		1	1	76	0	0
LP		0	0	**3**	4	5
P		0	0	**6**	1250	3

	aVCE	'Actionable'	'Non-actionable'	Sensitivity	Specificity
ClinVar		(*P/LP*)	(*VUS/LB/B*)		
'Actionable'		1262	**9**	0.9929	1
'Non-actionable'		0	418		

Table 2. Benchmarking an automated Variant Classification Engine (aVCE) employing subclassification using a time capsule of the ClinVar database

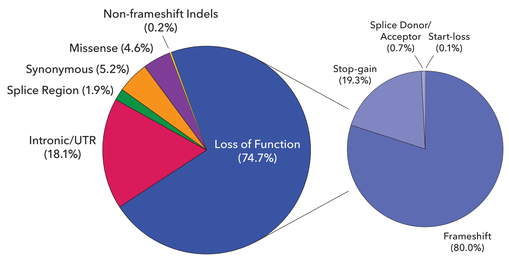
	aVCE	*B*	*LB*	*VUS-LB*	*VUS*	*VUS-WLP*	*VUS-SLP*	*LP*	*P*
ClinVar									
B		164	9	69	98	0	0	0	0
LB		1	1	55	16	**2**	**3**	0	0
LP		0	0	**1**	0	0	2	4	5
P		0	0	0	0	2	4	1250	3

	aVCE	'Actionable'	'Non-actionable'	Sensitivity	Specificity
ClinVar		(*P/LP/VUS-SLP*)	(*VUS/LB/B*)		
'Actionable'		1268	**3**	0.9976	0.9928
'Non-actionable'		**3**	415		

B. General Variant Effects

- 74.7% of variants represented LOF, most commonly frameshift and stop-gain effects **(Figure 1)**.
- All LOF were *P* variants in ClinVar
- All intronic/untranslated region (UTR) and synonymous effects were *B* variants in ClinVar
- Most (71/77) missense variants were *B*, while 6/77 were *P*, in ClinVar

Figure 1. Distribution of variant general effects and ClinVar classification



C. Discordant Variants (Table 1)

- 'Actionable' = *P* + *LP*
  - 9 discordant variants between ClinVar and aVCE

6 ClinVar *P/LP* variants classified as *VUS-SLP* by aVCE	2 ClinVar *P* variants classified as *VUS-WLP* by aVCE	1 ClinVar *LP* variant classified as *VUS-LB* by aVCE
Example: Variant (*P* for "GLYCOGEN STORAGE DISEASE" per ClinVar) met the PM1, PM2, PP2, and PP3 rules Based on strong evidence for pathogenicity, but not enough for *LP*, the aVCE aggregated prediction score resulted in *VUS-SLP* subclassification	**Example:** Very rare frameshift variant – also a type of indel – that occurred in a gene not documented to have a LOF pathogenic variant – PVS1 rule not met aVCE aggregated prediction score resulted in *VUS-WLP* subclassification	**Variant:** Variant (*LP* for "ALPORT SYNDROME" per ClinVar) met the PM2, PP2, and PP3 rules for pathogenicity – also appeared in a single individual in a homozygous state in gnomAD exomes aVCE aggregated prediction score resulted in *VUS-LB* subclassification

- 'Actionable' = *P* + *LP* + *VUS-SLP*
  - Only 6 discordant variants between ClinVar and aVCE

3 ClinVar *LB* variants classified as *VUS-SLP* by aVCE	2 ClinVar *P* variants classified as *VUS-WLP* by aVCE	1 ClinVar *LP* variant classified as *VUS-LB* by aVCE
2 variants – PM1, PM2, PP2 rules met **1 variant** – PM2, PP2, PP3 rules met	**See above**	**See above**

D. Variants and ACMG Rules

Table 3. Distribution of 1,271 'Actionable' variants by aVCE application of ACMG rules

		aVCE application of ACMG rules	
		Met	Unmet
ACMG rule/brief descriptor		n (%)	
PVS1	null variant where LOF known to cause disease	1,260 (99.1%)	11 (0.9%)[1]
PS1	same amino acid change as a known pathogenic variant	0	1,271 (100%)
PM1	mutational hot spot and/or critical, well-established functional domain	3 (0.2%)	1,268 (99.8%)[2]
PM2	absent from control databases or with extremely low frequency	1,271 (100%)	0
PM4	protein length changes due to in-frame deletions/insertions and stop losses	2 (0.2%)	1,269 (99.8%)
PM5	novel amino acid change at the same codon as a pathogenic variant	0	1,271 (100%)[3]
PP2	missense variant in gene with low rate of benign missense variation	5 (0.4%)	1,266 (99.6%)
PP3	multiple lines of computational (in silico) data support deleterious effect	7 (0.05%)	1,256 (99.5%)
PP5	reputable source reported *P*, but unable to perform independent evaluation	0	1,271 (100%)
BA1	allele frequency >5% in control databases	0	1,271 (100%)
BS1	allele frequency > expected for disorder in control databases	0	1,271 (100%)
BS2	observed in a healthy adult for disorder with full penetrance at early age	0	1,271 (100%)[4]
BP1	missense variant in gene for which truncation known to cause disease	0	1,271 (100%)
BP3	in-frame deletions/insertions in repetitive region with no known function	0	1,271 (100%)
BP4	multiple lines of computational (in silico) data suggest no impact	0	1,271 (100%)
BP6	reputable source reported *B*, but unable to perform independent evaluation	0	1,271 (100%)
BP7	synonymous (silent) variant for which splicing algorithm predicts no impact AND nucleotide highly conserved	0	1,271 (100%)

[1] 2, [2] 3, [3] 1, and [4] 2 variants flagged by aVCE (see Table 5)

Table 4. Distribution of 418 'Non-actionable' variants by aVCE application of ACMG rules

		aVCE application of ACMG rules	
		Met	Unmet
ACMG rule/brief descriptor		n (%)	
PVS1	null variant where LOF known to cause disease	0	418 (100%)
PS1	same amino acid change as a known pathogenic variant	0	418 (100%)
PM1	mutational hot spot and/or critical, well-established functional domain	13 (3.1%)	405 (96.9%)[1]
PM2	absent from control databases or with extremely low frequency	229 (54.8%)	189 (45.2%)[2]
PM4	protein length changes due to in-frame deletions/insertions and stop losses	0	418 (100%)
PM5	novel amino acid change at the same codon as a pathogenic variant	0	418 (100%)
PP2	missense variant in gene with low rate of benign missense variation	22 (5.3%)	396 (94.7%)
PP3	multiple lines of computational (in silico) data support deleterious effect	3 (0.7%)	415 (99.3%)
PP5	reputable source reported *P*, but unable to perform independent evaluation	0	418 (100%)
BA1	allele frequency >5% in control databases	138 (33.0%)	280 (67.0%)[3]
BS1	allele frequency > expected for disorder in control databases	29 (6.9%)	389 (93.1%)[4]
BS2	observed in a healthy adult for disorder with full penetrance at early age	236 (56.5%)	182 (43.5%)[5]
BP1	missense variant in gene for which truncation known to cause disease	7 (1.7%)	411 (98.3%)
BP3	in-frame deletions/insertions in repetitive region with no known function	1 (0.2%)	417 (99.8%)
BP4	multiple lines of computational (in silico) data suggest no impact	60 (14.4%)	358 (85.6%)
BP6	reputable source reported *B*, but unable to perform independent evaluation	0	418
BP7	synonymous (silent) variant for which splicing algorithm predicts no impact AND nucleotide highly conserved	87 (20.8%)	331 (79.2%)

[1] 4, [2] 6, [3] 6, [4] 16, [5] 43 variants flagged (see Table 5)

Table 5. Details of variants flagged by aVCE

ACMG rule/brief descriptor		No. of variants	Reason for flag
PVS1	null variant where LOF known to cause disease	2 Actionable	Null variant where LOF is not known to cause disease
PM1	mutational hot spot and/or critical, well-established functional domain	3 Actionable 4 Non-actionable	Region with a larger number of *P* than *B* variants, but not significantly higher
PM2	absent from control databases or with extremely low frequency	6 Non-actionable	A single outlier database with common frequency (>5%), while all other databases report very rare (<1%)
PM5	novel amino acid change at the same codon as a *P* variant	1 Actionable	Novel amino acid change within the same codon of a variant that was only reported (not confirmed) as *P*
BA1	allele frequency >5% in control databases	6 Non-actionable	A single outlier database with common frequency (>5%), while all other databases report very rare (<1%)
BS1	allele frequency > expected for disorder in control databases	16 Non-actionable	Frequency of 1%-1.5% in control public databases (somewhat higher than the very rare threshold of 1% for PM2 rule; rule met threshold is >1.5%)
BS2	observed in a healthy adult for disorder with full penetrance at early age	2 Actionable 43 Non-actionable	1) A single outlier database with common frequency; all other databases indicate very rare 2) A single individual appearing as a homozygous in public control database

- Interesting findings generated by the aVCE warranting further consideration

PSV1 Rule	PM1 Rule	PP3/BP4 Rules
Rule not met for 2 variants despite being LOF (gene not recognized as one where LOF is known disease mechanism) Going forward, as databases are continually updated, the aVCE will be trained to identify any LOF variant for such genes as meeting the PVS1 rule	2 different variants *P/LP* for very rare diseases according to ClinVar appeared in a homozygous state in allegedly healthy individual in control databases Could result from: ■ False positive in ClinVar classification ■ False positive in control database ■ Contamination of an affected individual in control database ■ Not 100% penetrance or the existence of another protective variant	7 *P* missense (n=6) and splice region (n=1) variants and 102 *B* missense (n=71) and splice region (n=31) variants ■ aVCE correctly called the PP3 rule for all 7/7 *P* variants compared with 3/102 *B* variants ■ aVCE correctly called the PP4 rule for 60/102 *B* variants and none of the *P* variants. ■ Remainder classified as *VUS* by the aVCE based on the PP3/BP4 rules not being met aVCE's aggregated prediction score was sensitive and specific in classifying variants

CONCLUSIONS

- The aVCE algorithm, even without input from clinical databases specific to the 'Test' set, could predict with very high sensitivity and specificity whether a variant in the future would be categorized as clinically 'actionable' versus 'non-actionable'
- In instances of discordance, the aVCE tended to under-call a variant as *VUS* rather than label a variant *LP* or *P* with insufficient evidence
- Results support the ongoing use of the ACMG rules of evidences as a standard for variant classification
- Knowledge derived from powerful computational methodologies can augment the human expertise and judgment still required to deduce final variant classifications
- Innovative approaches may allow for major advancements in variant classification, including those with limited clinical information, characterized by:
  - up-to-the minute database access
  - consistent weighting
  - rapid delivery of clinically meaningful information
- Such advances can:
  - aid clinical and research laboratory professionals in the current era characterized by increased complexity of variant analysis and interpretation
  - prove useful in future refinements of classification guidelines

ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics	aVCE	automated Variant Classification Engine	BS	benign strong	NGS	next-generation sequencing
AI	artificial intelligence	*B*	benign	CAP	College of American Pathologists	*P*	pathogenic
AMP	Assoc. for Molecular Pathology	BA	benign stand-alone	*LB*	likely benign	PM	pathogenic moderate
		BP	benign supporting	LOF	loss of function	PP	pathogenic supporting
				LP	likely pathogenic	PS/VS	pathogenic strong/very strong
RCV/SCV	*Reference/Submission Accession Version*	*UTR*	*untranslated region*	*VUS*	variant of uncertain significance	*VUS-LB*	variant of uncertain significance-leaning benign
VUS-SLP	variant of uncertain significance-strong leaning pathogenic	*VUS-WLP*	variant of uncertain significance-weak leaning pathogenic				

REFERENCES

1. Richards S, et al. *Genet Med* 2015;17(5):405-424.
2. Yang S, et al. *Genet Med* 2017;19(10):1118-1126.
3. Harrison SM, et al. *Genet Med* 2017;19(10):1096-1104.