Using Artificial Intelligence for Detection of Hotspot Regions: ACMG/AMP PM1 Criterion

Yaron Einhorn, MSc;[1] Adaia Kamshov, MD, MSc;[1] Nurit Paz-Yaacov, PhD;[1] Moshe Einhorn, MSc;[1] Michael Bespalov, BSc;[1] Susan Gross, MD, FACMG;[1,2] Elaine Lyon, PhD, FACMG[1,3]

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

ABSTRACT

Background: The 2015 ACMG/AMP sequence variant interpretation guideline provided a framework for classifying variants based on several benign (B) and pathogenic (P) evidence criteria, including a pathogenic criterion for missense variants located in a hotspot region (PM1). Recently published recommendations call for increased standardization of, and more conservative, scoring for calling the PVS1 rule for LOF variants (*Hum Mutat*. 2018;39:1517-1524. doi: 10.1002/humu.23626). However, there are current analytical challenges specific to hotspot regions, e.g., no existing hotspot region databases, or gold standard.

Aim: We hypothesized that an artificial intelligence-based variant classification engine (aiVCE) can be applied to aid in the detection of hotspot regions in a gene and sought to benchmark such an algorithm.

Methods: Within the aiVCE, for each exon/domain, a sliding window first extracts candidate regions between each pair of B variants or at the edges of the region, in order for it to be clear of B variants. Candidate regions without P variants within them are ignored. Within each candidate region, the AI model further detects the inner borders of P variants contained within, and the number of P variants. Based on the density and number of P variants, the model evaluates each region for the presence of a hotspot and then assigns a weight to the PM1 rule as 'supporting,' 'moderate,' or 'strong.' The AI-based weighting algorithm differentiates between the inner P region and the region between the P and B variants. In the absence of a gold standard for hotspot regions, we benchmarked the aiVCE's ability to detect hotspot regions against the different gene/disease-specific expert panels' (EP's) recommendations for PM1 and their published curated variant datasets (Table).

Results: Among the 131 variants where the aiVCE applied the PM1 rule, 110 (83.9%) were classified as P or likely pathogenic (LP) by the EPs, 14 (14.5%) were classified as uncertain, and only 2 (1.5%) were classified as B/LB (Table). Among the 110 P variants that met PM1 rule by the aiVCE, 76 were called by the aiVCE only. The EPs applied the PM1 rule to 60 variants, 55 (91.6%) of which were also called by the aiVCE (all were P/LP), while a single P/LP variant was called by the EPs only. The PM1 rule was applied 42 times in the PAH gene for P/LP variants by the aiVCE suggesting they are located in a hotspot region, while the rule was never applied by the EPs. Similarly, 4 P variants were found only by the aiVCE to be in a hotspot region for hearing loss genes.

Conclusions: We herein demonstrate AI-based algorithms can aid in the detection of exonic and domain hotspot locations in genes. Due to the lack of gold standard databases for applying PM1 rule, An AI-based hotspot detection is highly desirable. Our results show that the identification of the hotspot was consistent with the final classification of a variant by the Expert Panels. As such, it can be valuable tool for making future investigations of gene-based variant classification.

Introduction

American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) Standards and Guidelines – PM1 Criterion (Missense Variants Located in Hotspot Region)

- The 2015 ACMG/AMP standards and guidelines for sequence variant interpretation provided a framework for classifying variants based on several benign (B) and pathogenic (P) evidence criteria[1]
  - Includes a P criterion for missense variants located in a hotspot region (PM1)
- Recent publications provide refinements to scoring for some rules (e.g., calling for LOF variants[2])
- However, as no gold standards, recommendations, or hotspot region databases exist, this evidence criterion may benefit from innovative computational algorithms

Clinical Genome Resource (ClinGen) Expert Gene/Disease Panels (EPs)[3]

- Tasked with defining application of ACMG/AMP guidelines in specific genes/ diseases
- Recently, the 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[4]

Artificial Intelligence-Based Variant Classification Engine (aiVCE)

- Data-driven; based on ACMG/AMP classification guidelines
- Automates majority of ACMG/AMP classification rules
- Sequence 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

- Describe methodology for automatically predicting hotspot regions using an AI-based clustering algorithm
- Assess aiVCE at the rule level and compare hotspot detection performance to that of the EP decisions, utilizing the unique ClinGen Expert Curated Human Genetic Data database

Methods

Detection of Hotspot Regions (ACMG/AMP PM1 Criterion) using aiVCE

- Within the aiVCE, for each exon/domain, a sliding window initially extracts candidate regions between each pair of B variants or at the edges of the region to be clear of B variants
- Candidate regions without P variants ignored
- Within each candidate region, aiVCE further detects inner borders of P variants contained within, and determines number of P variants
- Based on density/number of P variants, the aiVCE evaluates each region for the presence of hotspots and then assigns a weight to the PM1 rule as 'supporting,' 'moderate,' or 'strong'
- The AI-based weighting algorithm differentiates between the inner P region and the region between the P and B variants

Comparison of aiVCE vs ClinGen Expert Curated Human Genetic Data for Hotspot Detection

- In the absence of a gold standard for hotspot regions, we compared the aiVCE's ability to detect hotspot regions with the different gene/disease-specific EPs' recommendations for PM1 and their published curated variant datasets (**Table 1**)
- To assess whether EP work can be assisted via automation, we compared whether or not the PM1 rule was met by the aiVCE and the EP for each variant that was classified by one of the EPs as P or B

Table 1. ClinGen EPs' gene/disease-specific published curated variant datasets utilized for assessing the aiVCE's ability to detect hotspot regions

Gene	Published curated variant dataset
CDH1 (tumor suppressor)	Lee K, et al. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum Mutat* 2018;39:1553-68.
PTEN (tumor suppressor)	Mester JL, et al. Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum Mutat* 2018;39:1581-92.
Hearing loss	Oza AM, et al. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat* 2018; 39:1593-613.
PAH (phenylalanine hydroxylase)	Zastrow DB, et al. Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): The ClinGen IEM Working Group and the Phenylalanine Hydroxylase Gene. *Hum Mutat* 2018;39:1569-80.
MYH (DNA repair)	Kelly MA, et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med* 2018;20:351-9.
RASopathy (germline mutation)	Gelb BD, et al. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genet Med* 2018;20:1334-45.

Results

- Overall, the EPs applied the PM1 rule to 60 variants, 55 (91.6%) of which were also called by the aiVCE (all were P/LP) (**Table 2**)
- Of the 56 variants that were called as P/LP by EP, and for which the PM1 rule was met by the EP –
  - aiVCE found 55 (98.2%) variants for which the PM1 rule was met
  - Only a single P/LP variant was called by the EPs only
- Among the 131 variants where the aiVCE applied the PM1 rule (**Figure 1**):

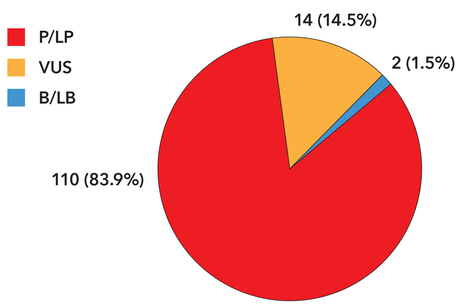

Figure 1. EP classification of the 131 variants where the aiVCE applied the PM1 rule

- Among the 110 P variants that met the PM1 rule by the aiVCE, 76 were called by the aiVCE only
- The PM1 rule was applied 42 times in the PAH gene for P/LP variants by the aiVCE, suggesting they are located in a hotspot region, while the rule was never applied by the EPs
- Similarly, 4 P variants were found only by the aiVCE to be in a hotspot region for hearing loss genes

Table 2. Assessment of aiVCE ability to detect hotspot regions using ClinGen EP recommendations for PM1 gene datasets

	PM1 by EP and AI	PM1 applied by EP only			PM1 applied by aiVCE only		
Gene	P/LP[1]	P/LP	VUS	B/LB	P/LP	VUS	B/LB
CDH1	0	0	0	0	0	0	0
PTEN	**1**	0	0	0	**4**	2	0
Hearing loss	**2**	0	0	0	**4**	0	0
PAH	0	0	0	0	**42**	8	0
MYH7	**44**	0	4	0	0	8	**1**
RASopathy	**8**	**1**	0	0	**5**	1	**1**
Total	**55**	**1**	4	0	**55**	19	**2**

[1]VUS/B/LB not called by both EP and AI.

CONCLUSIONS

- Determining whether a variant resides in a hotspot region, which requires clinical judgment and interpretation, is a difficult task due to lack of criterion-specific guidelines
- aiVCE methodology provides a more structured and evidence-based framework for predicting variant location in a hotspot region
- Compared with EPs, the aiVCE demonstrated high agreement (92-98%) in detecting hotspot region variants; thus, the aiVCE can aid detection of hotspot locations in the genome
- Many variants were found to be in hotspot regions by aiVCE, but not EP, most likely owing to:
  - Thresholds were set for this study to be less stringent, calling PM1 with less strength as not to miss any P variants, to allow variant scientists to manually review and determine whether variants are of clinical concern based on severity (thresholds can be customized based on gene of interest)
  - Determining variant location in a hotspot region is challenging, requiring clinical judgment and interpretation, and vulnerable to missed calls

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. Abou Tayoun AN, et al. Recommendations for interpreting loss of function PSV1 ACMG/AMP variant criterion. *Hum Mutat* 2018;39:1517-24.
3. 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.
4. 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