Journal of the American Heart Association

ORIGINAL RESEARCH

Performance of a Protein Language Model for Variant Annotation in Cardiac Disease

Aviram Hochstadt, MD, MPH; Chirag Barbhaiya, MD; Anthony Aizer, MD, MSc; Scott Bernstein, MD; Marina Cerrone, MD; Leonid Garber, MD; Douglas Holmes, MD; Robert J. Knotts, MD; Alex Kushnir, MD, PhD; Jacob Martin, MD; David Park, MD, PhD; Michael Spinelli, MD; Felix Yang, MD; Larry A. Chinitz, MD; Lior Jankelson, MD, PhD

BACKGROUND: Genetic testing is a cornerstone in the assessment of many cardiac diseases. However, variants are frequently classified as variants of unknown significance, limiting the utility of testing. Recently, the DeepMind group (Google) developed AlphaMissense, a unique artificial intelligence–based model, based on language model principles, for the prediction of missense variant pathogenicity. We aimed to report on the performance of AlphaMissense, accessed by VarCardio, an open web-based variant annotation engine, in a real-world cardiovascular genetics center.

METHODS AND RESULTS: All genetic variants from an inherited arrhythmia program were examined using AlphaMissense via VarCard.io and compared with the ClinVar variant classification system, as well as another variant classification platform (Franklin by Genoox). The mutation reclassification rate and genotype–phenotype concordance were examined for all variants in the study. We included 266 patients with heritable cardiac diseases, harboring 339 missense variants. Of those, 230 (67.8%) were classified by ClinVar as either variants of unknown significance or nonclassified. Using VarCard.io, 198 variants of unknown significance (86.1%, 95% CI, 80.9–90.3) were reclassified to either likely pathogenic or likely benign. The reclassification rate was significantly higher for VarCard.io than for Franklin (86.1% versus 34.8%, $P<0.001$). Genotype–phenotype concordance was highly aligned using VarCard.io predictions, at 95.9% (95% CI, 92.8–97.9) concordance rate. For 109 variants classified as pathogenic, likely pathogenic, benign, or likely benign by ClinVar, concordance with VarCard.io was high (90.5%).

CONCLUSIONS: AlphaMissense, accessed via VarCard.io, may be a highly efficient tool for cardiac genetic variant interpretation. The engine's notable performance in assessing variants that are classified as variants of unknown significance in ClinVar demonstrates its potential to enhance cardiac genetic testing.

Key Words: AlphaMissense ■ artificial intelligence ■ genetic testing ■ genotype ■ phenotype ■ VarCard.Io ■ variants of unknown significance

Genetic testing for monogenic disease has transformed the care of patients affected by a wide spectrum of cardiovascular conditions by enabling the identification of specific mutations that explain individual phenotypes.[1] The enhanced precision that comes with genetic analysis results in improved inference of risk, tailoring of gene-specific targeted therapy (ie, nadolol in *LQT1*), and upstream prevention using cascade screening in family members. In recent cohorts, as many as 50% of patients with genetically linked cardiovascular conditions are found to be positive for a reportable genetic variant.[1–3] A major limitation of genetic testing remains the finding of

Correspondence to: Lior Jankelson, MD, PhD, Director, Inherited Arrhythmia Program, Clinical Cardiac Electrophysiology, Leon H. Charney Division of Cardiology, NYU Grossman School of Medicine, NYU Langone Health, 560 First Avenue, Kimmel Pavilion 4th Floor, New York, NY 10016.
Email: lior.jankelson@nyulangone.org

This manuscript was sent to Jacquelyn Y. Taylor, PhD, PNP-BC, RN, FAHA, FAAN, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Preprint posted on MedRxiv June 05, 2024. doi: https://doi.org/10.1101/2024.06.04.24308460.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.124.036921

For Sources of Funding and Disclosures, see page 6.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Variants of unknown significance are found frequently in genetic testing for potentially inherited cardiac diseases; their interpretation may significantly change patient management.
- AlphaMissense has been recently shown to predict cardiac missense pathogenicity with high accuracy, a new tool developed using AlphaMissense findings. VarCard.io was found to correctly reclassify variants of unknown significance in a real-world cardiac genetic clinic data set.

What Are the Clinical Implications?

- VarCard.io can be used by clinicians as an adjunct tool to ascertain the clinical relevance of genetic testing results to further enhance the utility of genetic testing.

Nonstandard Abbreviations and Acronyms

AA	amino acid
CSV	clinically significant variant
LB	likely benign
LP	likely pathogenic
VUS	variant of unknown significance

a variant of unknown significance (VUS), occurring in up to 40% of genetic tests.[4–7] Recently, the DeepMind group (Google, Mountain View, CA) developed AlphaFold and AlphaMissense,[8,9] artificial intelligence–based models, for the prediction of protein folding and variant pathogenicity, respectively. Here, we report the performance of AlphaMissense, interfaced via VarCard.io,[10] an open web-based variant annotation engine allowing cDNA and protein change queries aligned with Clingen-derived gene–disease correlation, and its performance in a real-world cardiovascular genetics center.

METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request and New York University Langone Health Institutional Review Board committee approval.

AlphaMissesne and VarCard.Io

The development and testing of AlphaMissesne have been reported elsewhere.[9] Briefly, AlphaMissense predicts the pathogenicity of all possible amino acid (AA) changes in all known protein sequences. The model is trained by using an unsupervised learning approach, implementing a language model architecture. In an initial pretraining step, deep complex representations of protein sequence data are built by masking and unmasking AAs at random along a given sequence. By training the model to predict the masked AAs, the model learns the fundamental properties of the human proteome and the compatibility of any AA in any position. In the second step, the model is fine-tuned on a set of variants that are labeled on the basis of either being highly frequent or completely absent from human and primate populations. To allow easy interaction with AlphaMissense predictions, our team, led by the senior author (L.J.), built VarCard.io, a free nonprofit web engine that allows query of all missense variants by gene name and cDNA or AA change and extracts the AlphaMissense annotation of each variant as either likely benign (LB), likely pathogenic (LP), or ambiguous. In addition, VarCard.io extracts the gene-disease correlation for the queried gene, as assessed by ClinGen.

Patients and Annotations

All patients with missense variants found in probands presenting to the New York University Inherited Arrhythmia program were included. All missense variants found in each patient were included. We compared the annotation of AlphaMissense via VarCard.io to the ClinVar[11] database, as well as to the annotation of an independent commercial classification engine, Franklin[12] (Genoox, Palo Alto, CA). ClinVar annotations were recorded as per American College of Medical Genetics and Genomics guidelines to be either pathogenic, LP, VUS, benign, or LB. If the variant did not appear on the ClinVar database, it was treated in this study as a VUS unless otherwise specified. If a variant had a mixed classification (ie, >1 entry in ClinVar), it was treated as a VUS if at least 1 annotation was a VUS.

Assessing Clinical Genotype–Phenotype Relationship

All genes in the database were assessed using the ClinGen[13] framework (ClinGen, National Institutes of Health, Bethesda, MD) as to their known gene–disease correlation. Patient records were reviewed, and the clinical phenotype was assessed by using all clinical data available, including patient's history, ECG tracings, provocative tests, and imaging (eg, echocardiography, cardiac magnetic resonance imaging, calcium pyrophosphate imaging). The analysis was done by 2 physicians trained in assessing cardiogenetic cases (A.H. and M.C.) and the senior author (L.J.), who is the director of our institution's cardiogenetic program. All decisions were reached via consensus of

all 3 assessors, and the phenotype analysis was performed before the collection of VarCard.io's results. The genotype–phenotype correlation was considered as positively concordant if (1) a patient had a variant in a gene with known gene–disease correlation compatible with the designated clinical phenotype for this gene by ClinGen, and (2) the variant was classified as pathogenic/LP (positive concordance). Alternatively, negative concordance was determined if (1) a variant was found in a gene not compatible with the patient's clinical phenotype by ClinGen, and (2) the variant was classified as benign/LB. Otherwise, the genotype–phenotype was considered as discordant (ie, a patient with a pathogenic/LP variant in a gene not compatible with the patients' phenotype). Patients were excluded from analysis if their phenotype was indeterminable due to missing data. Patients included in the study provided written informed consent before inclusion, which was approved by the New York University Langone Health Institutional Review Board committee in accordance with the Helsinki Declaration.

Comparison of Variants' Functional Data With Varcard.io's Results

The *KCNH2* gene, whose channel (*HERG*) functional characteristics are well studied, was taken as a case study. Previously published results[14,15] of the *KCNH2* variant's channel current in comparison with wild type was compared with the VarCard.io results to assess Varcard.io's ability to predict the functional implications of an *KCNHN2* variant.

Statistical Analysis

Discrete variables are reported as numbers and rates, and 95% CIs are calculated using the Clopper–Pearson method. The statistical significance of rate differences was assessed using the χ^2 test. Results were considered significant when P values were <0.05. All calculations were done using R version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

We reviewed 623 probands in our original database, of which 266 patients were identified as meeting the inclusion criteria of having at least 1 genetic variant that was classified as a missense variant in a gene recognized to be associated with a cardiac phenotype by ClinGen (Figure 1). These patients had a total of 339 different genetic variants. The most common phenotype was long QT syndrome, followed by nonischemic cardiomyopathy and hypertrophic cardiomyopathy (Figure 1). The most frequently involved genes

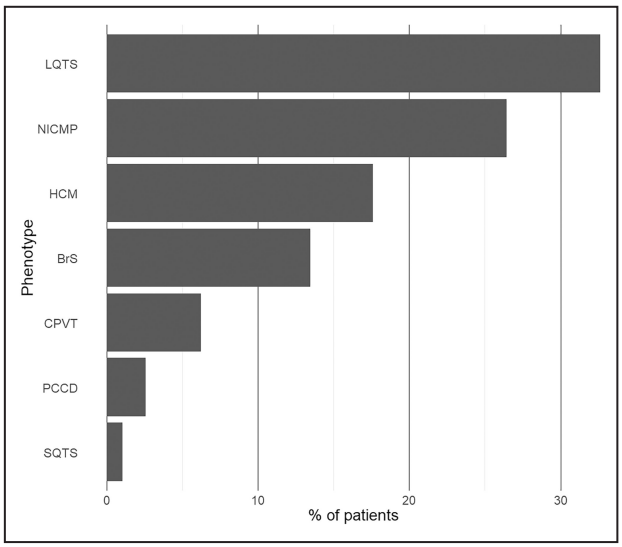

Figure 1. Frequency graph of patients' most frequent phenotypes in the cohort.
BrS indicates Brugada syndrome; CPVT, catecholaminergic polymorphic ventricular tachycardia; HCM, hypertrophic cardiomyopathy; LQTS, long QT syndrome; NICMP, nonischemic cardiomyopathy; PCCD, progressive cardiac conduction defect; and SQTS, short QT syndrome.

were *SCN5A* (implicated in Brugada syndrome, dilated cardiomyopathy, and long QT syndrome), followed by *KCNQ1*, *KCNH2*, and *RyR2* (Figure 2).Of the 339 variants tested in VarCard.io, 328 variants had ClinVar entries, of which 110 (33.5%) variants had a clinically significant variant annotation of pathogenicity (ie, pathogenic, LP, benign, or LB without VUS entries), and the rest were classified as VUS or had mixed classification (Table).

Comparison of ClinVar Annotation to VarCard.io

For the 230 (67.8%) VUSs (ie, variants with at least 1 ClinVar entry as VUS or not reported), 198 variants were reclassified by VarCard.io as either LP or LB, yielding a reclassification rate of 86.1% (95% CI, 80.9–90.3). These reclassifications were to LB in 111 variants (56.1% [95% CI, 48.8–63.1]) and to LP in 87 variants (43.9% [95% CI, 36.9–51.2]; Figure 3). Only 32 variants (13.9%) had an ambiguous annotation by VarCard.io. Of the 10 most common genes in the cohort, *DSP* gene variants and *ANK2* gene variants had the highest reclassification rate (100%), and *KCNH2* and *KCNQ1* had the lowest reclassification rates of 37.5% and 18.8%, respectively (Table S1).

Of the 109 variants with a CSV ClinVar annotation (ie, variants without any VUS entry), 105 remained a CSV, and the other 4 (3.7%) were annotated as ambiguous by VarCard.io. The overall agreement rate between

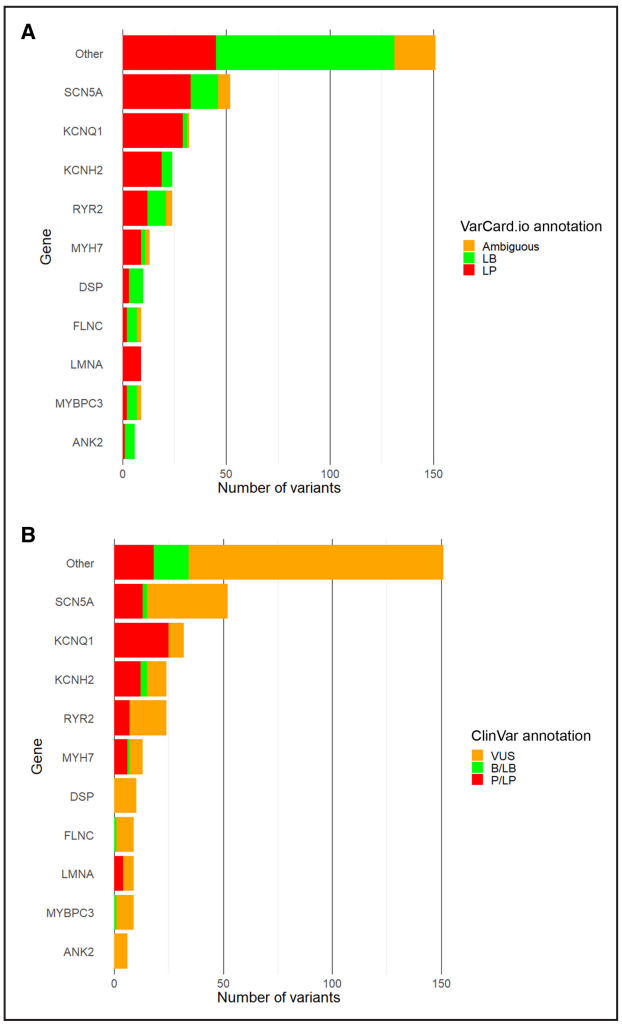

Figure 2. Variant frequency and annotation according to (A) VarCard.io and (B) ClinVar.
LB indicates likely benign; LP, likely pathogenic; and VUS, variant of unknown significance.

ClinVar and VarCard.io on CSVs (ie, same annotation in both systems) was 90.5% ([95% CI, 83.2–95.3]). This agreement rate was similar for 81 variants with ClinVar pathogenic/PL annotation and for 24 variants with ClinVar benign/LB annotation (91.4% and 87.5%, respectively; *P*=0.865).

VarCard.Io and ClinVar Genotype–Phenotype Concordance

Of the 303 variants with a CSV annotation by VarCard.io, 269 variants were related to a patient with a defined phenotype. Of these, 258 variants (95.9% [95% CI, 92.8–97.9]) had a concordant genotype–phonotype prediction: 138 (96.5% [95% CI, 92–98.9]) with positive concordance and 120 (95.2% [95% CI, 89.9–98.2]) with negative concordance. There were 34 variants (11.2%) in patients without sufficient clinical data to establish a defined phenotype and thus were not included in this analysis. For the significantly smaller group of patients with CSV annotation on ClinVar, of 69 patients with sufficient clinical data to establish a defined phenotype, 61 (88.4% [95% CI, 78.4–94.8]) had concordant genotype–phonotype prediction, and 8 (11.5% [95% CI, 5.1–21.6]) had a discordant prediction.

Comparison With Another Publicly Available Variant Classification Engine (Franklin)

Using Franklin on the 230 VUS variants appearing in ClinVar resulted in reclassification to either pathogenic/LP or benign/LB in only 80 variants, giving a reclassification rate of 34.8% (95% CI, 28.6–41.3), significantly lower than the 86.1% (95% CI, 80.9–90.3) VarCard.io reclassification rate (*P*<0.001). For the 100 variants with CSV annotation in both ClinVar and Franklin, agreement was 100% (95% CI, 96.4–100).

Relationship With Functional Channel Assessment

As the *KCNH2* gene is well studied, we compared VarCard.io's results with previously published[14,15] data regarding function assessment of the 24 *KCNH2* variants in our cohort. We were able to procure data for 11 variants, of which 3 were classified as LB and 8 as LP. All 8 LP variants had KCHN currents of <50% of wild-type channels. Of the 3 LB variants, 2 had a normal current and one borderline (p. Arg176Trp 48% current compared with wild type channels).

Table. Variant Classification According to ClinVar

ClinVar variant classification	N (%)	Total (%)	Comment
Only pathogenic/LP	85 (25.1)	85 (25.1)	CSV
Only benign/LB	24 (7.1)	24 (7.1)	
Pathogenic/LP and VUS	26 (7.7)	219 (64.6)	VUS/Mixed
VUS alone	122 (36.0)		
Benign/LB and VUS	67 (19.8)		
Mixed (>2 classes)	4 (1.2)		
Not reported in ClinVar	11 (3.2)	11 (3.2)	

CSV indicates clinically significant variant; LB, likely benign; LP, likely pathogenic; and VUS, variant of unknown significance.

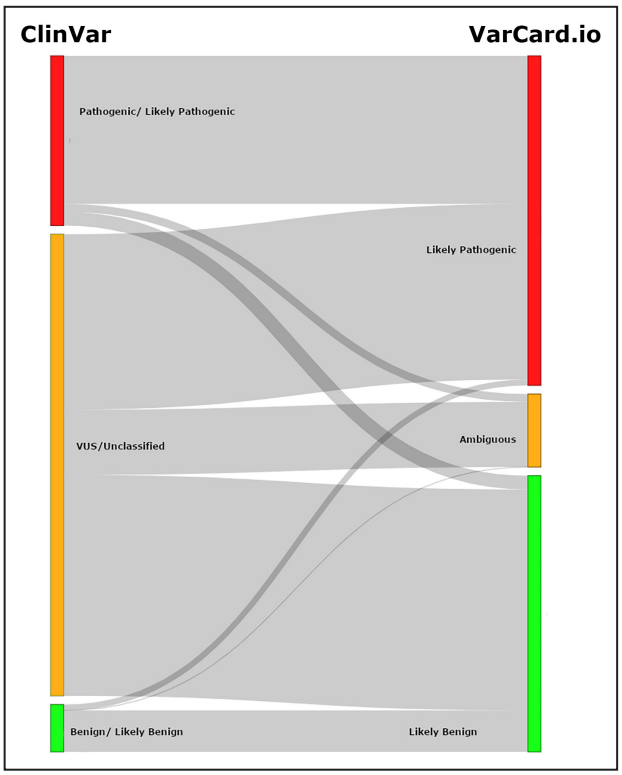

Figure 3. Sankey diagram showing reclassification of variant reclassification changes from ClinVar to VarCard.io. VUS indicates variant of unknown significance.

DISCUSSION

In this work, we aimed to evaluate the performance of AlphaMissense, interfaced through VarCard.io, an open searchable web-based engine for variant pathogenicity prediction, on a real-world cohort of patients with heritable arrhythmia and cardiomyopathy syndromes (central illustration). Our main findings are:

1. Using VarCard.io, we reclassify 86.1% of VUSs to either LP or LB classes. This reclassification rate was significantly greater than the reclassification rate of 35% observed in a widely used commercial variant annotation engine (Franklin).
2. Assigned annotations by VarCard.io were highly concordant with the observed clinical phenotypes, with 95.9% genotype–phenotype concordance.
3. We found a 90.5% agreement rate between ClinVar and VarCard.io on variants with pathogenic or benign annotations.

A diagnosis of VUS in a gene with cardiac manifestations, occurring in up to 40% of genetic tests,[4–7] is a pressing clinical problem that may have direct implications on patients' risk assessment, specific therapy, possible preventive strategies (such as in *PKP2* mutation carriers), family screening opportunities, and patients' anxiety.[16,17] As genetic panels become increasingly available and encompass more genes, the incidence of VUSs is also increasing, stressing the need for reliable and accessible variant annotation strategies.

The AlphaMissense variant pathogenicity model has the potential to facilitate the field of genetic testing by significantly enhancing the performance of computational VUS reclassification. The use of AlphaMissense was previously shown to result in the classification of 88.8% of 69.5 million variants present on gnomAD[9] to LP or LB, compatible with our finding of an 86.1% reclassification rate for ClinVar-reported VUSs.

The major difference in architecture between AlphaMissesne and prevailing popular computational models such as SIFT[18] and REVEL[19] is that these models are largely based on labels of pathogenicity guided by sequence homology between species and prevalence in population databases. This introduces inherent biases, as some relatively rare variants are pathogenic, and many cardiac conditions are characterized by age-dependent penetrance. In addition, traditional supervised artificial intelligence models are often trained on human-provided labels, thereby preserving a circular logic and bias.

The high performance achieved by AlphaMissense is the result of the model's unique self-supervised architecture, built and trained on the basis of the principles of a large language model. In the case of a language model, each word is represented by a vector (embedding) that captures semantic relationships between words, allowing the model to reason about language, that is, predict the compatibility of any word in any given context. In the case of AlphaMissense and variant prediction, vectors represent AAs, and their compatibility in each genomic position is computed on the basis of complex properties that capture their "context" within the protein. To further enhance performance, the model is then fine-tuned on a defined set of variants with well-established annotations based on extreme population frequency properties.

While our results demonstrate VarCard.io's utility, it is important to acknowledge certain limiting aspects. As AlphaMissense is an artificial intelligence model with predictions based on complex, convoluted representations, there is no simple way to derive linear mechanistic explanations for the different predictions and their pathogenesis in the clinical phenotype. In addition, AlphaMissense does not infer on penetrance or predict how a variant would interact with a patient's specific multiomic inventory to create the specific phenotype (ie, would not predict if a specific *SCN5A* variant would cause Brugada syndrome or long QT syndrome). Furthermore, assessing clinical

concordance of genes and phenotypes might be biased as both a gene's function and possibly pathogenicity and a patient's phenotype may be complex and not straightforward to assess for this study's purposes. Variants may also have incomplete penetrance and age-related penetrance; thus, a patient not having a phenotype at a specific time point does not necessarily mean a specific variant is not pathogenic in nature. Finally, as suggested by the American College of Medical Genetics and Genomics, computational predictions are 1 component in a matrix of criteria recommended for the clinical evaluation of genetic variants. Thus, although the PP3 criterion was recently upgraded from to "supporting" to "moderate" or "strong,"[20] it should still be considered within the context of other features.

In conclusion, despite the above limitations, VarCard.io, propelled by AlphaMissense, may be a highly efficient tool for cardiac genetic variant interpretation. The AlphaMissense notable performance in assessing variants that are classified as VUS in ClinVar demonstrates the potential of artificial intelligence as an important step forward in the field of clinical cardiovascular genetics.

ARTICLE INFORMATION

Received June 2, 2024; accepted September 13, 2024.

Affiliation

NYU Langone Health and the NYU Grossman School of Medicine, New York, NY.

Sources of Funding

None.

Disclosures

None.

Supplemental Material

Table S1

REFERENCES

1. Cirino AL, Harris S, Lakdawala NK, Michels M, Olivotto I, Day SM, Abrams DJ, Charron P, Caleshu C, Semsarian C, et al. Role of genetic testing in inherited cardiovascular disease: a review. *JAMA Cardiol*. 2017;2:1153–1160. doi: 10.1001/jamacardio.2017.2352
2. Behere SP, Weindling SN. Inherited arrhythmias: the cardiac channelopathies. *Ann Pediatr Cardiol*. 2015;8:210–220. doi: 10.4103/0974-2069.164695
3. Morales A, Hershberger RE. The rationale and timing of molecular genetic testing for dilated cardiomyopathy. *Can J Cardiol*. 2015;31:1309–1312. doi: 10.1016/j.cjca.2015.06.034
4. Mazzaccara C, Lombardi R, Mirra B, Barretta F, Esposito MV, Uomo F, Caiazza M, Monda E, Losi MA, Limongelli G, et al. Next-generation sequencing gene panels in inheritable cardiomyopathies and channelopathies: prevalence of pathogenic variants and variants of unknown significance in uncommon genes. *Biomolecules*. 2022;12:1417. doi: 10.3390/biom12101417
5. van Lint FHM, Mook ORF, Alders M, Bikker H, Lekanne D, Deprez RH, Christiaans I. Large next-generation sequencing gene panels in genetic heart disease: yield of pathogenic variants and variants of unknown significance. *Neth Heart J*. 2019;27:304–309. doi: 10.1007/s12471-019-1250-5
6. Young WJ, Maung S, Ahmet S, Kirkby C, Ives C, Schilling RJ, Lowe M, Lambiase PD. The frequency of gene variant reclassification and its impact on clinical management in the inherited arrhythmia clinic. *Heart Rhythm*. 2024;21:903–910. doi: 10.1016/j.hrthm.2024.01.008
7. Rehm HL, Alaimo JT, Aradhya S, Bayrak-Toydemir P, Best H, Brandon R, Buchan JG, Chao EC, Chen E, Clifford J, et al. The landscape of reported VUS in multi-gene panel and genomic testing: time for a change. *Genet Med*. 2023;25:25. doi: 10.1016/j.gim.2023.100947
8. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589. doi: 10.1038/s41586-021-03819-2
9. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381:eadg7492. doi: 10.1126/science.adg7492
10. VarCard.io—AI Diagnostic Tool. Accessed February 14, 2024. https://www.varcard.io/.
11. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res*. 2020;48:D835–D844. doi: 10.1093/nar/gkz972
12. Franklin. Accessed February 14, 2024. https://franklin.genoox.com/clinical-db/home.
13. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. ClinGen—the clinical genome resource. *N Engl J Med*. 2015;372:2235–2242. doi: 10.1056/NEJMsr1406261
14. Thomson KL, Jiang C, Richardson E, Westphal DS, Burkard T, Wolf CM, Vatta M, Harrison SM, Ingles J, Bezzina CR, et al. Clinical interpretation of *KCNH2* variants using a robust PS3/BS3 functional patch-clamp assay. *Human Genetics and Genomics Advances*. 2024;5:100270. doi: 10.1016/j.xhgg.2024.100270
15. Ng C-A, Perry MD, Liang W, Smith NJ, Foo B, Shrier A, Lukacs GL, Hill AP, Vandenberg JI. High-throughput phenotyping of heteromeric human ether-à-go-go-related gene potassium channel variants can discriminate pathogenic from rare benign variants. *Heart Rhythm*. 2020;17:492–500. doi: 10.1016/j.hrthm.2019.09.020
16. Muller RD, McDonald T, Pope K, Cragun D. Evaluation of clinical practices related to variants of uncertain significance results in inherited cardiac arrhythmia and inherited cardiomyopathy genes. *Circulation: Genomic and Precision Medicine*. 2020;13:e002789. doi: 10.1161/CIRCGEN.119.002789
17. Hellwig LD, Biesecker BB, Lewis KL, Biesecker LG, James CA, Klein WMP. Ability of patients to distinguish among cardiac genomic variant subclassifications. *Circulation: Genomic and Precision Medicine*. 2018;11:e001975. doi: 10.1161/CIRCGEN.117.001975
18. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–3814. doi: 10.1093/nar/gkg509
19. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*. 2016;99:877–885. doi: 10.1016/j.ajhg.2016.08.016
20. Pejaver V, Byrne AB, Feng B-J, Pagel KA, Mooney SD, Karchin R, O'Donnell-Luria A, Harrison SM, Tavtigian SV, Greenblatt MS, et al. ClinGen sequence variant interpretation working group. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet*. 2022;109:2163–2177. doi: 10.1016/j.ajhg.2022.10.013