Briefings in Bioinformatics, 2025, **26(5)**, bbaf545
https://doi.org/10.1093/bib/bbaf545
Review

The promises and pitfalls of automated variant interpretation: a comprehensive review

Mireia Costa*, Alberto García S., Ana León, Oscar Pastor

VRAIN, Universitat Politècnica de València, Camino de Vera s/n, E-46022 Valencia, Spain

*Corresponding author. VRAIN, Universitat Politècnica de València, Camino de Vera s/n, E-46022 Valencia, Spain. E-mail: micossan@vrain.upv.es

Abstract

The interpretation of DNA variants enables personalized medicine through precise diagnosis and treatment selection. To address the challenges of manual interpretation, a wide range of automated tools has been created. This study evaluates these tools through a comprehensive analysis of their features, methodologies, and performance assessment against the interpretations of the ClinGen Expert Panel for 256 variants associated with cardiomyopathies, hereditary cancer, or monogenic diabetes. Although the tools demonstrated high accuracy for clearly pathogenic/benign variants, they showed significant limitations with variants of uncertain significance (VUS). Despite current advances in automation for variant interpretation, our findings show that expert oversight is still needed when using these tools in a clinical context, particularly for VUS interpretation.

Keywords: DNA variant; variant interpretation; interpretation tools; variants of uncertain significance

Introduction

During the past decade, precision medicine has revolutionized clinical care by tailoring diagnoses and treatments to the unique characteristics of each patient, moving beyond the limitations of traditional, one-size-fits-all approaches. Genomics is central to this personalized approach, as it plays a critical role in understanding how DNA variants shape our physical traits, predispose us to certain health conditions, and affect our responses to specific treatments.

The effect of a variant can range from neutral to significant in terms of health impact, and the process of determining its significance for an individual's health is known as variant interpretation. This interpretation process follows specific guidelines to ensure consistent and accurate results. These guidelines have been widely adopted by geneticists and clinical experts [1], and provide a standardized framework for classifying variants into five clinically relevant categories: pathogenic, likely pathogenic, uncertain significance, benign, and likely benign.

One of the first and most widely adopted guidelines is the one developed by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) in 2015 [2]. As experts began to use these guidelines, they realized how important it was to adapt them to account for the nuances of specific genes, diseases, and types of variants. This natural progression toward specialization, known as "guideline adaptation" [3], has resulted in hundreds of adapted clinical guidelines. Examples include the AMP/ASCO/CAP guidelines for somatic variant interpretation [4] and the ACMG-ClinGen guidelines for copy number variant (CNV) analysis in constitutional genetics [5].

Despite their difference in scope, all guidelines share a common element: they establish criteria whose fulfillment determines the variant interpretation. However, evaluating these criteria is not easy, and it requires analyzing complex evidence distributed across thousands of heterogeneous data sources. In the early stages of adopting these guidelines, this was commonly a manual task that quickly proved to be intricate, time-consuming, and prone to human errors and inconsistencies among experts [6]. To address these challenges, two distinct computational approaches have emerged to assist in genetic variant interpretation.

First, *in silico predictors*. These predictors use computational algorithms, often incorporating artificial intelligence (AI) or complex statistical methods, to predict a variant's likelihood of causing a significant functional effect by assessing characteristics such as evolutionary conservation and physicochemical properties. However, their accuracy is variable and often depends on the variant type or specific gene affected. Consequently, current interpretation guidelines explicitly state that these tools are not a substitute for the interpretation process, but rather complementary evidence to be used cautiously.

Second, *variant interpretation automation tools*, which offer a more comprehensive solution due to their intention of replicating the entire variant interpretation process. These tools focus specifically on automating the evaluation of criteria defined within established clinical interpretation guidelines. To do so, these tools

Received: May 13, 2025. **Revised:** August 29, 2025. **Accepted:** September 20, 2025

collect, integrate, and assess diverse data from multiple sources in accordance with the specific conditions and evidence thresholds outlined in the guidelines. Essentially, these tools strive to replicate the human expert assessment, enhancing efficiency through automation. In fact, they often incorporate *in silico* predictors' data as one of the lines of evidence in their comprehensive evaluations. Currently, there is significant variability among these tools in the number of criteria that they automate, how they perform such automation, and the data sources that they use. Consequently, the reliability and consistency of these tools in clinical settings remain uncertain.

This work addresses this knowledge gap by characterizing and evaluating variant interpretation automation tools, given their design to facilitate the comprehensive, guideline-driven interpretation process crucial for robust clinical decision-making. We aim to do so by (i) presenting a comprehensive overview of existing tools that automate variant interpretation, highlighting their interpretation methodology, accepted input formats, technological features, and update frequency; and (ii) assessing the performance of the most broadly-applicable tools by comparing their interpretations with those provided by ClinGen Expert Panels using a set of 256 variants. For this assessment, the tools were selected based on their potential for widespread clinical adoption, including support for both GRCh37 and GRCh38 reference genomes, the availability of a user-friendly web interface, and the applicability to any clinical domain.

By achieving these objectives, we aim to provide a comprehensive overview of the current landscape of variant interpretation tools and an analysis of the practical applicability of these tools for routine clinical practice. To the best of our knowledge, this combined approach of comprehensive review and targeted performance assessment is novel in scope and depth.

The remainder of this paper is structured as follows: The Variant Interpretation Tools section presents the methodology used to identify currently existing tools; The Overview of the Variant Interpretation Tools section describes their functionality and technical specifications; The Case study section details a case study that compares the performance of a subset of tools; The Discussion section analyzes the implications of our findings for researchers and practitioners; and finally, the Key Points section highlights the main contributions and addresses future research directions.

Variant interpretation tools

To identify freely available tools automating variant interpretation, we conducted a Targeted Literature Review (TLR), a non-systematic, in-depth, and informative literature review to identify significant references and maximize rigor while minimizing bias.

The TLR search was conducted using PubMed (https://pubmed.ncbi.nlm.nih.gov), a comprehensive repository for biomedical literature containing over 36 million citations. We defined a syntactic query combining key terms related to variant interpretation, genomics, and automation (see Fig. 1 for further details).

The query is structured into four parts, each addressing a different selection criterion, highlighted in Fig. 1 with a specific color. In blue, we include the two terms commonly used to describe a DNA variant: "variant" and "variation." In orange, we target papers that mention interpretation-related terms, such as "classification"—often used synonymously with interpretation in this field—or derivatives of ACMG-AMP guidelines—as they are most widely used. In green, we narrow our focus to the DNA domain, incorporating terms such as "genomic" and "genetics." Lastly, the pink component of the query seeks articles that describe the implementation of the interpretation process by considering terms like "software," "tool," or "automation."

The titles and abstracts of the publications retrieved from PubMed were manually reviewed, and each article was classified into one of the following groups: "irrelevant," "not a tool," "preprint," or "tool" (see Table 1 for detailed definitions of each category). For the purposes of this study, only articles classified as "tool" were considered relevant.

We acknowledge that some tools may not have been documented in academic literature. To ensure all available tools were captured, we complemented the TLR with a web search using the Google search engine. We used "variant interpretation tool" as the query search, a broad term intended to encompass a wide range of potential tools. We limited the search to the first 10 pages of results to balance comprehensiveness and feasibility. Each result was manually evaluated against the criteria defined in Table 1, focusing on those websites that host tools designed to automate the evaluation of interpretation guidelines such as the ACMG-AMP.

To ensure the relevance and applicability of our findings, the tools identified through both search strategies were subject to a final filtering process according to the criteria outlined in Table 2. These criteria were designed to select tools that are (i) free to use, (ii) stand-alone in operation, (iii) currently available online, (iv) fully automated (i.e. no manual input is required), and (v) focused on automating the evaluation of an entire guideline, rather than focusing on a single criterion.

Following the methodology described above, we conducted the search on 30 July 2025. The PubMed query yielded a total of 1544 publications, which were categorized according to the criteria outlined in Table 1. As a result, 1129 were classified as "Not a tool," 363 as "Irrelevant," 26 as "Preprint," and 25 as "Tool." Complementing these results, the web search uncovered nine additional tools, seven of which had not been identified in the literature search. In total, 32 tools advanced to the next stage of our analysis.

Each of these 32 tools was manually analyzed according to the filtering criteria described in Table 2. Only tools that met all filtering criteria were selected for further analysis. The results of the filtering process, summarized in Fig. 2, led to the exclusion of 19 tools for the following reasons: four tools were not freely accessible (*open access* filter); four tools relied on existing automation rather than providing new approaches (*novelty* filter); four tools had nonfunctional web pages or code repositories (*availability* filter); six tools lacked full automation of the criteria (*automation* filter); and one tool only automated a single criterion of the ACMG-AMP guidelines (*completeness* filter). Consequently, 13 tools met all established criteria and were selected for further analysis. A full list of all 1544 articles and the 32 initially identified tools, including detailed exclusion reasons for the 19 tools that did not meet our selection criteria, is provided in [7]. Table 3 provides a clear overview of these tools, including their names, descriptions, scientific references, and URLs.

The selection process described in this section resulted in the identification of 13 tools that met all of our established criteria. These tools represent a diverse set of freely accessible, stand-alone, and fully automated solutions for variant interpretation based on recognized guidelines.

Overview of the variant interpretation tools

Building on the identification process outlined in the previous section, we now provide a detailed description of each of the

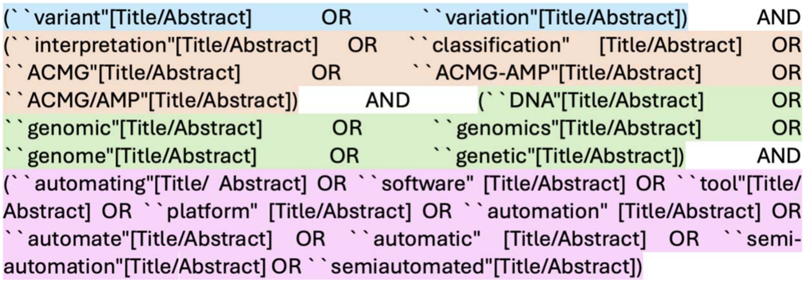

Figure 1. Diagram showing the structure of the PubMed search query, divided into four sections: variant-related terms, interpretation-related terms, genomic domain terms, and automation/tool-related terms.

Table 1. Classification criteria for assessing the relevance of publications identified in the literature review

Category	Explanation
Irrelevant	The article describes a tool that does not interpret variants using interpretation guidelines. e.g. articles that focus exclusively on AI methods for predicting variant impact.
Not a tool	The article does not describe a concrete tool, e.g. articles reviewing existing tools.
Preprint	The article is still in preprint status and has not yet been fully accepted.
Tool	The article describes a tool of interest.

Table 2. Filtering criteria for evaluating and selecting variant interpretation tools. These five criteria were used to assess the suitability and relevance of the initially identified tools for inclusion in the final analysis

Criterion	Explanation
Free access	The tool allows free access to its main functionalities.
Novelty	The tool provides a novel interpretation approach rather than using existing ones.
Availability	The tool's web page or code repositories are active and functional.
Automation degree	The tool fully automates the criteria evaluation, without requiring manual input for evidence.
Completeness	The tool focuses on automating entire interpretation guidelines rather than a single criterion.

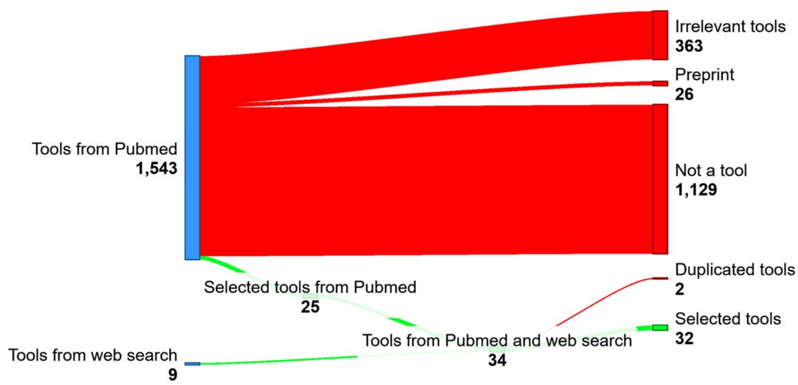

Figure 2. A flowchart summarizing the methodology and results obtained in the variant interpretation tool identification process. This figure details the number of tools identified and excluded at each step, providing a clear overview of how the final selection was reached.

13 tools that met our selection criteria. This analysis intends to offer a more in-depth understanding of each tool's strengths and potential applications, highlighting their key features and providing insights to assist clinicians in selecting the most appropriate option for their specific needs. Our analysis examines four key aspects of each tool: variant interpretation methodologies, input format requirements, technological features that influence their integration into clinical workflows, and the update frequency.

Interpretation methodology

First, we explored the variant interpretation methodology defined by each tool. We focused on five key aspects: the specific guidelines each tool adheres to and the clinical domains in which they are applicable, the number of criteria implemented by each tool, differences in criteria implementation across interpretation tools, and the different data sources utilized for variant interpretation.

Guidelines and domain applicability

Six tools (InterVar, Genebe, TAPES, ELLA, Franklin, and VarSome) automate the ACMG-AMP guidelines, enabling broad applicability across genetic conditions. Three tools automate the ACMG-AMP guidelines, but with modifications to adapt to the needs of concrete domains. PathoMAN is focused on the unique characteristics of hereditary cancer genes, CardioClassifier tailors the guidelines to meet the specific requirements of inherited cardiac conditions, and VIP-HL is designed for genetic hearing loss following the ClinGen Hearing Loss Expert Panel's adaptation of the ACMG-AMP guidelines.

Table 3. Overview of the 13 selected variant interpretation tools that meet our criteria

Tool	Description	URL
PathoMAN	Pathogenicity of Mutation Analyzer (PathoMAN) automates the curation of germline genomic variants in clinical cancer genetics following the ACMG-AMP guidelines [8].	https://pathoman.mskcc.org
VIP-HL	The Variant Interpretation Platform for genetic Hearing Loss (VIP-HL) aims to semi-automate the Hearing loss ACMG-AMP based variant interpretation guidelines [9].	http://hearing.genetics.bgi.com/
Cancer SIGVAR	Cancer SIGVAR implements the ACGS and ClinGen guidelines for automating the clinical interpretation of variants in hereditary cancer-related genes [10].	http://cancersigvar.bgi.com/
InterVar	Clinical interpretation of genetic variants by the ACMG/AMP 2015 guidelines for all genes and diseases [11].	https://wintervar.wglab.org
CardioClassifier	Automated and interactive web tool that supports disease-specific interpretation of genetic variants in genes associated with Inherited Cardiac Conditions [12].	http://www.cardioclassifier.org
Genebe	Online platform that streamlines the automated application of ACMG-AMP guidelines for assessment of pathogenicity of genetic variants [13].	https://genebe.net
TAPES	TAPES is an open-source tool designed to predict pathogenicity of variants in exome studies by implementing an automatization of the ACMG-AMP guidelines [14].	https://github.com/a-xavier/tapes
VIC	Computational Tool for interpreting the clinical impact of somatic variants following the AMP-ASCO-CAP 2017 Guidelines [15].	https://github.com/HGLab/VIC/
vaRHC	R package to automate the variant interpretation process for hereditary cancer genes. It uses the ACMG-AMP guidelines and gene-specific guidelines when available [16].	https://github.com/emunte/vaRHC
ClassifyCNV	Command-line tool that implements the ACMG-AMP guidelines to evaluate the pathogenicity of germline duplications and deletions [17].	https://github.com/Genotek/ClassifyCNV
ELLA	Tool for clinical interpretation of genetic variants, developed with a particular focus on speed, quality, and reproducibility.	https://allel.es
Franklin	Franklin by Genoox is a community-driven genomic data platform. One of its purposes is streamlining variant interpretation by automating the evaluation of the ACMG-AMP guidelines.	https://franklin.genoox.com
VarSome	Suite of bioinformatics tools for processing and annotation of NGS data that automates the ACMG/AMP recommendations for all genes and diseases	https://varsome.com

The remaining four tools implement guidelines other than ACMG-AMP. ClassifyCNV specializes in CNVs, following the ACMG-ClinGen guidelines for constitutional CNVs [5]. The other three tools focus on cancer variant interpretation: CancerSIGVAR implements the ACGS and ClinGen Sequence Variant Interpretation guidelines; VIC applies the AMP-ASCO-CAP 2017 guidelines for somatic variants [4]; and vaRHC uses gene-specific guidelines for cancer-related genes while applying ACMG-AMP guidelines for non-cancer genes.

This diversity in guideline adherence and adaptation reflects the complex nature of variant interpretation across different genetic conditions, emphasizing the importance of using tools that are appropriate for specific clinical or research contexts.

Number of criteria implemented

The utility of variant interpretation tools is significantly influenced by the extent of their guideline coverage, as tools that automate more criteria provide substantially more comprehensive analysis than those with limited coverage. Our analysis reveals considerable variation in this aspect across different tools.

As illustrated in Fig. 3, the majority of tools automate over 60% of their respective guideline criteria. However, automation rates vary among specialized tools. VIP-HL and VaRHC nearly reach this threshold, automating just under 60% of the criteria. More notably, ClassifyCNV and CardioClassifier lag behind, automating <50% of their respective criteria. In contrast, PathoMAN and VIC, both specialized in different areas of cancer genomics, have successfully automated 71.4% and 70% of their criteria, respectively.

Among tools implementing the ACMG-AMP guidelines, VarSome leads with an 82% automation rate. Franklin and TAPES follow closely at 71.42% each. ELLA, InterVar, and Genebe achieve

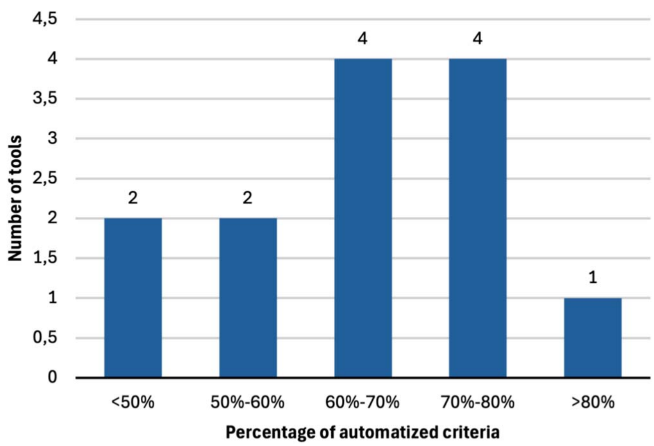

Figure 3. Distribution of variant interpretation tools based on the percentage of automated criteria relative to the guideline they automate. The x-axis represents the percentage ranges of automated criteria. The y-axis shows the number of tools falling within each range.

automation levels below 70%. Notably, some of the higher-performing tools benefit from commercial backing (VarSome and Franklin), which may contribute to their enhanced automation capabilities and resource availability.

As Fig. 4 illustrates, most of the tools that implement the ACMG/AMP criteria are focused on (i) evaluating a variant's frequency in specific populations (BA1 and BS1); (ii) assessing whether predictive tools consider the variant potentially damaging (PP3 and BP4); and iii) examining specific aspects of the variant type and its effects on gene or protein regions (PM1, PM4, and BP3). This consistent implementation suggests these criteria may be more straightforward to automate.

Criteria	CardioClassifier	Genebe	InterVar	TAPES	ELLA	PathoMAN	Franklin	VarSome
PVS1	✓	✓	✓	✓		✓	✓	✓
PS1	✓	✓	✓	✓			✓	✓
PS2				✓	✓		✓	✓
PS3				✓	✓	✓		✓
PS4	✓		✓		✓	✓	✓	
PM1	✓	✓	✓	✓	✓	✓	✓	✓
PM2	✓	✓	✓	✓		✓	✓	✓
PM3					✓			
PM4	✓	✓	✓	✓	✓	✓	✓	✓
PM5	✓	✓	✓	✓		✓	✓	✓
PM6					✓		✓	✓
PP1					✓			✓
PP2	✓	✓	✓	✓		✓	✓	✓
PP3	✓	✓	✓	✓	✓	✓	✓	✓
PP4			✓					
PP5		✓		✓	✓	✓	✓	✓
BA1	✓	✓	✓	✓	✓	✓	✓	✓
BS1	✓	✓	✓	✓	✓	✓	✓	✓
BS2		✓	✓	✓	✓	✓	✓	✓
BS3				✓	✓	✓		✓
BS4					✓			✓
BP1		✓	✓	✓	✓	✓	✓	✓
BP2					✓			
BP3	✓	✓	✓	✓	✓	✓	✓	✓
BP4	✓	✓	✓	✓	✓	✓	✓	✓
BP5								
BP6		✓	✓	✓		✓	✓	✓
BP7		✓	✓	✓		✓	✓	✓

Figure 4. Implementation of the 28 ACMG-AMP criteria across each tool. A precise definition of each criterion can be found in [2].

Conversely, other criteria prove more challenging to automate, as evidenced by their infrequent implementation. For instance, criteria evaluating evidence from functional studies (PS3 and BS3) are implemented by only half of the tools analyzed. Even less common are criteria assessing evidence from affected families (PP1 and BS4), implemented solely by ELLA and VarSome. The evaluation of other relevant variants in patient DNA (PM3 and BP2) is automated exclusively by ELLA. Finally, the assessment of alternative molecular causes of disease (BP5) is not implemented by any of the analyzed tools.

Criteria implementation

The selection of an appropriate tool for DNA variant interpretation requires careful consideration not only of the number of implemented criteria but also of the specific methodologies employed by each tool in evaluating these criteria. Despite apparent similarities in implemented criteria, differences in interpretation methodologies can lead to divergent interpretation outcomes. This comparison focuses specifically on tools implementing ACMG-AMP guidelines, as tools using other guidelines are unique in their respective frameworks, preventing a meaningful comparison.

A notable example of such divergence is observed in the evaluation of the PS3 criterion (PS3 criterion: well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product [2]) by TAPES and PathoMAN. Both tools utilize ClinVar, a widely recognized repository for variant interpretation [18]. However, their approaches differ substantially. On the one hand, TAPES considers the PS3 criterion fulfilled if the variant is interpreted in ClinVar as either “Pathogenic” or “Drug Response” and requires that ClinVar provides a confidence rating of three or four stars. On the other hand, PathoMAN only considers this criterion fulfilled if the variant is interpreted in ClinVar as “Pathogenic” but accepts a lower confidence threshold of two or more stars. Besides, it also includes manually curated literature evidence to complement the information obtained from ClinVar.

Another illustrative example is the implementation of the PM1 criterion (PM1 criterion: located in a mutational hot spot and/or critical and well-established functional protein domain (e.g. active site of an enzyme) without benign variation [2]). One of the aspects evaluated by this criterion is whether a variant is located within a mutational hotspot. VarSome and Genebe employ distinct approaches to this evaluation. VarSome defines a hotspot as a region containing at least four pathogenic variants within a 50-nucleotide span, centered on the variant of interest. In contrast, Genebe considers a broader protein-based context, examining a range of 15 amino acids surrounding the variant. Notably, Genebe does not specify a precise threshold for the number of pathogenic variants required to designate a hotspot.

These examples highlight the critical need for researchers and clinicians to consider the specific implementation details of variant interpretation tools. Such understanding is essential for making informed decisions in tool selection and for accurate interpretation of the results in the context of genetic variant analysis.

Data sources

The number of data sources utilized by each variant interpretation tool, as depicted in Fig. 5, varies significantly. A comprehensive list of all data sources used by each tool is provided in [7], detailing the specific databases, repositories, and resources accessed by each variant interpretation platform. Franklin stands

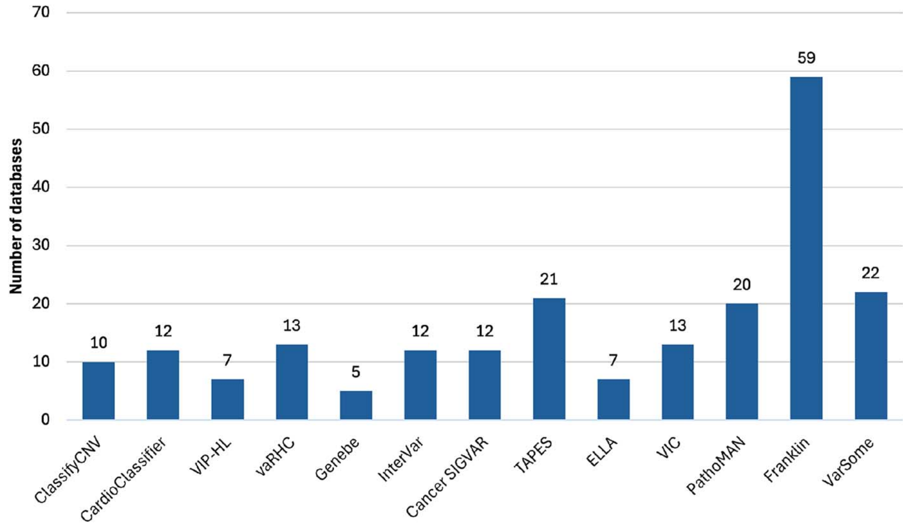

Figure 5. Comparison of data sources used by variant interpretation tools. The x-axis lists each tool. The y-axis indicates the number of data sources used.

out significantly, incorporating 59 data sources, which is nearly three times more than the next highest tools. VarSome and TAPES follow with 22 and 21 data sources, respectively, also providing a robust foundation for interpretation.

The majority of tools (61.5%) consider between 10 and 13 data sources for interpretation, including CardioClassififer, VIC, vaRHC, InterVar, and Cancer SIGVAR. This range appears to be a common benchmark in the field, possibly representing a balance between comprehensive coverage and computational efficiency. At the lower end of the spectrum, tools such as ClassifyCNV, VIP-HL, and ELLA utilize only seven data sources each, while Genebe considers just five.

The inclusion of multiple, varied genomic databases allows for a more holistic view of each variant, potentially uncovering specific information critical for accurate interpretation. Tools that use a greater number of reputable data sources may be better equipped to provide more context-rich and reliable variant interpretations across a wider range of clinical scenarios.

However, when choosing a tool, it is important to consider not only the number of data sources but also the specific ones used for performing the interpretation, as this may lead to differences in assessments between tools. An illustrative example of this occurs with the PM1 criterion of the ACMG/AMP guidelines (see footnote3). One of the aspects evaluated by this criterion is whether a variant is located within a functionally relevant protein domain. InterVar and VarSome employ different strategies for this evaluation. InterVar automates the criterion by determining if the variant resides within a functional domain as described in InterPro, a specialized database for protein domains. VarSome, on the other hand, uses UniProt, an alternative but equally pertinent database in the field of protein domains. The use of distinct data sources by these tools may result in varying PM1 criterion evaluations for identical variants, further reflecting the potential for tool-dependent outcome variability.

In this context, the transcript data source chosen by a tool is especially relevant. The reason is that the transcript data source selected can change a variant's predicted impact, potentially influencing clinical decisions [19]. Given this, ensuring that the variant interpretation tool matches the transcripts used to annotate the input data is critical.

There are two predominant transcript data sources: Ensembl transcripts and RefSeq. RefSeq, maintained by NCBI, focuses on manually curated, nonredundant sequences, prioritizing accuracy [20]. Ensembl, developed by EBI and Wellcome Trust Sanger Institute, aims for comprehensiveness with more automated annotation, including a broader range of transcript variants and species [21].

Our analysis reveals diversity in using transcript data sources across the evaluated tools. Five tools—ClassifyCNV, VIP-HL, InterVar, Cancer SIGVAR, and ELLA—exclusively use RefSeq, prioritizing its manually curated, nonredundant sequences. In contrast, only two tools, CardioClassifier and PathoMAN, rely solely on Ensembl, leveraging its comprehensive, automated annotation approach.

A significant number of tools offer more flexibility in their approach. vaRHC, TAPES, VIC, Franklin, and VarSome provide the option to use both RefSeq and Ensembl data sources, enabling a more versatile, multidisciplinary approach to variant interpretation. This dual-source strategy allows users to cross-reference and compare annotations from both databases, potentially leading to more robust interpretations. Notably, Genebe stands out by exclusively utilizing MANE (Matched Annotation from NCBI and EMBL-EBI), a new collaborative initiative between RefSeq and Ensembl. MANE aims to create a unified, high-confidence transcript set, potentially offering the best of both worlds in terms of curation and comprehensiveness.

The prevalence of tools offering multiple database options or adopting unified approaches such as MANE indicates a shift toward a more flexible stance in transcript selection, providing more flexibility to the users.

Type of input

This section focuses on the type of input accepted by each variant interpretation tool. First, we evaluated the genomic reference assembly versions supported by each tool (e.g. GRCh37 or GRCh38). Regarding the reference assembly, four of the tools support exclusively GRCh37, lacking compatibility with the more recent GRCh38 genomic reference sequence (i.e. PathoMAN, VIP-HL, Cancer SIGVAR, CardioClassifier). In contrast, eight tools offer support for both GRCh37 and GRCh38, providing greater

Table 4. Technological characteristics of Variant interpretation tools. This table summarizes the technological aspects of the 13 evaluated tools that are relevant for accessibility, maintainability, and integration

Tool	Programming language	Web interface	Last update
PathoMAN	Not specified	Yes	2019
VIP-HL	HTML5, JavaScript (VUE), Java and Python 3	Yes	2021
Cancer SIGVAR	Not specified	Yes	2021
InterVar	Python	Yes	2022
CardioClassifier	Perl and PHP	Yes	2017
Genebe	Python	Yes	2024
TAPES	Python	No	2023
VIC	Java	No	2021
vaRHC	R	No	2024
ClassifyCNV	Python	No	2021
ELLA	Python, JavaScript	No	2024
Franklin	Not specified	Yes	2024
VarSome	HTML5, JavaScript (React), Python 3 (Django) and C++	Yes	2024

flexibility for users working with different genomic assemblies (i.e. InterVar, TAPES, VIC, vaRHC, ClassifyCNV, Franklin, Genebe, and VarSome).

Second, we focused on whether the tools were able to process single variants described with their chromosomal position at their accepted assemblies or with transcript-level positions. Most tools support single variant input at the chromosome level, with seven tools (VIP-HL, Cancer SIGVAR, CardioClassifier, Genebe, vaRHC, Franklin, and Varsome) also accepting transcript locations.

However, three tools (TAPES, VIC, and ELLA) do not accept single variant search, only accepting Variant Call Format (VCF) files with multiple variants from a single patient. Six other tools, besides the single variant interpretation, can also handle multiple variants via VCF files: Cancer SIGVAR, InterVar, CardioClassifier, vaRHC, Franklin, and VarSome. However, notable distinctions exist. TAPES and ELLA require VCF files in the VEP annotation tool format, while Franklin and VarSome's VCF processing is limited to their premium versions. This variability in input format support highlights the differing levels of flexibility and availability across the tools.

These findings show the importance of considering input compatibility when selecting a variant interpretation tool for specific clinical or research applications. Indeed, the requirements differ significantly between interpreting a single variant and performing a complete genomic analysis of a patient. The genomic assembly in which the variants are described and the tool's accepted input are of utmost importance in selecting the most appropriate tool for a given task.

Technological features

The technological basis of variant interpretation tools is critical for their functionality, usability, and maintainability. These factors influence the performance of the tools and their ease of integration into existing workflows. We looked at two key technological aspects: the existence of a web interface and the programming language used. Each of these factors is important for different reasons: the existence of a web interface greatly impacts accessibility and ease of use; the programming language affects the future extensibility of the tool and community support; and the frequency of knowledge updates indicates how rapidly new scientific findings are integrated into the tool. Table 4 summarizes these characteristics for each of the 13 tools included in our study.

Eight out of 13 evaluated tools (61.5%) offer a web interface, including PathoMAN, VIP-HL, Cancer SIGVAR, InterVar, CardioClassifier, Genebe, Franklin, and VarSome. This feature significantly facilitates their adoption in clinical settings, as it allows for easy access and use without the need for local installation or advanced technical knowledge.

Conversely, the remaining tools (TAPES, VIC, vaRHC, ClassifyCNV, and ELLA) do not provide a web interface, instead requiring local installation and often utilizing command-line interfaces. This approach may present a significant barrier to adoption, particularly for clinical experts who may lack the technical expertise to install and operate such tools effectively.

Another relevant aspect, particularly for tools lacking a user-friendly web interface and requiring local installation, is the underlying programming language used in their development. The choice of programming language directly impacts a tool's compatibility with existing bioinformatics workflows, ease of local deployment, and long-term maintainability for technical users. Python, for instance, has emerged as a predominant language in bioinformatics, serving as the foundation for several tools in our analysis, including InterVar, Genebe, TAPES, ClassifyCNV, and ELLA. This is not merely a trend but a practical choice for pipeline developers and bioinformaticians, owing to its extensive libraries, relative ease of use, and strong community support, which collectively facilitate integration into diverse computational environments and enable custom modifications. However, some tools employ a multi-language process, such as VIP-HL and VarSome. While this approach may offer greater flexibility during initial development, it can introduce increased complexity in setup, maintenance, and the required technical expertise for integration into a specific laboratory's infrastructure.

Focusing on the tools that do not offer a web interface, we find a notable pattern. Three of these tools (TAPES, ClassifyCNV, and ELLA) use Python as their main programming language. This choice may benefit their adoption by institutions looking to install tools locally, as Python's popularity in bioinformatics means that many organizations already have the necessary expertise and infrastructure in place. The remaining two tools (VIC, vaRHC) use R and Java, respectively. While these languages are also widely used in scientific computing, they may present additional barriers to adoption compared with Python-based tools, depending on the specific expertise available within an organization.

Update frequency

Update frequency, defined here as the release of a new version as specified on the tool's webpage or the last update on its GitHub repository, should also be considered when selecting a tool, as it can impact its accuracy and relevance. A tool that is regularly updated is more likely to incorporate the latest research and data, leading to more accurate and relevant results. Our analysis revealed considerable variation in the frequency of updates among the tools.

The most recently updated tools were Genebe, vaRHC, ELLA, Franklin, and VarSome, all updated in 2024. This shows that these tools are actively committed to maintenance and continuous improvement, which is crucial in a field where knowledge evolves rapidly. In contrast, CardioClassifier was last updated in 2017, standing as the least recently updated tool. This raises concerns about the currency of its genomic knowledge and the reliability of its results.

Case study

To assess the performance of the variant interpretation tools, we conducted a comparative analysis against a reference standard. The choice of this standard was critical, as it directly impacted the validity and clinical relevance of the case study. For benchmarking purposes, we required a reference standard with the highest level of curation quality, ensuring every single interpretation has been manually curated by experts through a comprehensive review process.

While several interpretation data sources exist, many function as public repositories that aggregate variant interpretations from diverse sources, including both expert annotations and laboratory submissions, without manually validating their accuracy or reliability. These repositories often contain data of variable quality, which can lead to conflicts and inconsistencies that compromise their utility as a reference standard for tool evaluation. ClinVar, despite standing out as one of the most widely utilized and valuable resources in clinical practice, exemplifies this challenge. Despite being one of the most widely used and valuable resources in clinical practice [18], its comprehensive aggregation nature presents methodological limitations for tool evaluation.

This led us to select the ClinGen resource as our reference standard, as it stores interpretations performed exclusively by Expert Panels. These panels comprise multidisciplinary specialists who perform comprehensive literature reviews, functional studies analysis, and rigorous consensus-building processes that often require months of deliberation per variant. Moreover, these expert panels contribute their curated interpretations to ClinVar as the highest confidence submissions available in the repository.

From this resource, we selected 256 variants associated with one of the following clinical domains: Cardiomyopathy (87 variants), Hereditary Breast, Ovarian, and Pancreatic Cancer (71 variants), and Monogenic Diabetes (98 variants). In this dataset, we aimed to achieve a balanced representation across interpretation categories, targeting ∼20 variants for each interpretation (pathogenic, likely pathogenic, uncertain significance, likely benign, and benign), subject to availability.

Then, we focused on tool selection with the goal of identifying versatile solutions that could serve as standard resources across different clinical settings. To achieve this, we defined a set of selection criteria in consultation with clinical genetics experts. We prioritized tools that (i) support both GRCh38 and GRCh37 reference genome assemblies; (ii) provide a user-friendly web interface; and (iii) are designed for broad clinical application rather than disease- or gene-specific use.

First, the requirement for supporting both assemblies directly reflects the current landscape in clinical laboratories. Despite GRCh38's release in 2013, many laboratories continue to align their next-generation sequencing data to GRCh37. This persistence with the older assembly is driven by practical considerations, including significant migration costs, insufficient staff, and often, a lack of perceived benefits outweighing these burdens [22]. Compounding this, even when laboratories attempt to bridge assemblies, tools for lifting over coordinates may struggle with complex variants such as insertions and deletions (indels), variants in repetitive regions, or those within assembly gaps [23–25]. Therefore, tools offering support for both assemblies are more likely to accommodate current laboratory practices and have greater potential for widespread adoption. Second, the emphasis on web interfaces was driven by expert feedback, which highlighted that accessibility and ease of use are important factors for routine clinical adoption. Thus, web-based tools reduce implementation barriers and facilitate analysis in clinical practice. Finally, the focus on broad-spectrum tools rather than specialized solutions reflects the need for versatile platforms that can serve diverse use cases without requiring multiple tool implementations for different disease contexts.

Based on these criteria, four tools emerged as suitable candidates for our analysis: Franklin, Varsome, InterVar, and GeneBe (see Table 3 for a comprehensive overview of these tools).

We tested these tools with the 256 variants from ClinGen using their respective web interfaces. Our initial observation revealed that not all tools were capable of processing every selected variant, with each tool demonstrating different limitations. Franklin and GeneBe showed similar constraints; both tools are unable to process five variants. These problematic variants were all imprecise large variants (e.g. "NC_000011.9:g.(?_108137888)_(108225611_?)dup," or "NC_000011.10:g.(?_108287594)_(108287721_?)del") or were located in complex repetitive regions. Franklin faced an additional challenge with one variant for which only the chromosomal position and not the transcript position were available. VarSome demonstrated slightly better coverage, failing to process only two variants. Notably, these were the same imprecise large variants that posed challenges for Franklin and GeneBe, suggesting a common difficulty across multiple tools for addressing this type of variant. InterVar, however, exhibited the most significant limitations among the tools examined. It was unable to process 51 variants, a substantially higher number compared with the other tools. This limitation stems from the online version of InterVar being restricted to exonic variants that are not indels, which considerably narrows its applicability in variant interpretation.

To assess the performance of the variant interpretation tools, we first conducted a preliminary evaluation. In this initial assessment, we used a general accuracy metric based on exact matches between tool-provided interpretations and the corresponding ClinGen interpretations for all 256 variants. As Table 5 reflects, this metric revealed suboptimal performance across all tools, with Franklin achieving the highest accuracy at 58.4%.

Given these results, we implemented a more lenient approach, grouping the categories "benign" and "likely benign," as well as "pathogenic" and "likely pathogenic," as equivalent. The results following this correction (see *Corrected Accuracy* column) showed substantial improvement across all tools. Notably, GeneBe demonstrated the highest corrected accuracy at 82.07%,

Table 5. Performance metrics of variant interpretation tools in comparison with ClinGen Expert Panel interpretations. The "Accuracy" column shows the percentage of exact matches between tool interpretations and ClinGen interpretations. The "Corrected accuracy" column displays improved percentages after grouping "benign" with "likely benign" and "pathogenic" with "likely pathogenic" as equivalent interpretations

Tool	Accuracy (%)	Corrected accuracy (%)
Franklin	58.4	78.4
VarSome	51.97	81.89
InterVar	36.58	59.31
GeneBe	50.99	82.07

Table 6. Corrected accuracy of variant interpretation tools across different genetic contexts as defined by ClinGen Expert Panels (i.e. cardiomyopathy, hereditary cancer, and monogenic diabetes)

Tool	Cardiomyopathy (%)	Hereditary cancer (%)	Monogenic diabetes (%)
Franklin	80.23	92.65	66.67
VarSome	80.46	82.61	82.65
InterVar	70.88	69.77	42.68
GeneBe	82.55	82.35	81.44

indicating substantial alignment with ClinGen interpretations under these modified criteria.

To determine whether the observed differences in accuracy between the tools were statistically significant, we performed McNemar's test for each pair of tools, resulting in six comparisons. Given the multiple comparisons, we applied the Bonferroni correction to adjust our significance level from 0.05 to 0.00833 (0.05 divided by 6) to mitigate the risk of Type I errors.

This statistical analysis revealed that, despite apparent numerical differences in accuracy, most were not statistically significant. Specifically, the difference between GeneBe (the most accurate tool) and Varsome (the second most accurate tool) was not significant, with a P-value of 1.0. Similarly, the comparison between Franklin (the third most accurate tool) and Varsome resulted in a P-value of 0.243, and between Franklin and GeneBe, with a P-value of 0.222, indicating no significant differences despite the accuracy differences.

Significant differences were observed only when comparing InterVar (the least accurate tool) with the other tools, showing statistically significant differences with Franklin ($P = 2.93 \times 10^{-5}$), Varsome ($P = 7.39 \times 10^{-7}$), and GeneBe ($P = 2.77 \times 10^{-7}$). This indicates that only Intervar's performance differs significantly from the others.

On the other hand, Table 6 presents the corrected accuracy of each variant interpretation tool across the three expert panels included in this study. Interestingly, the performance of the tools varies across different genetic contexts.

VarSome and GeneBe exhibit consistent and comparable performance across all panels, with accuracy ranging from 80.46% to 82.65%. This uniformity indicates broad applicability across the studied contexts, establishing both tools as versatile options for variant interpretation across multiple genetic diseases.

Franklin performs similarly to VarSome and GeneBe in the Cardiomyopathy domain. However, it significantly outperforms these tools in Hereditary Cancer, achieving a remarkable accuracy of 92.65%. This suggests a strong alignment with ClinGen's interpretations in this specific domain. Such exceptional performance indicates that Franklin may be particularly well suited for interpreting variants associated with hereditary cancer syndromes. Conversely, Franklin underperforms in the Monogenic Diabetes domain, with an accuracy of 66.67%, notably lower than VarSome and GeneBe.

InterVar consistently performed poorly compared with the other tools across all domains. This is particularly evident in the Monogenic Diabetes category, where its accuracy drops to 42.68%, highlighting significant challenges in interpreting variants in this specific context.

These findings reflect the importance of selecting the appropriate variant interpretation tool based on the specific disease context, as no single tool uniformly outperforms the others across all categories.

To perform a more nuanced analysis, we evaluated each variant interpretation tool's performance across three main categories: Benign, Uncertain Significance, and Pathogenic (Table 7). Consistent with our previous analysis, we considered "benign" and "likely benign," as well as "pathogenic" and "likely pathogenic," as equivalent interpretations.

Table 7. Corrected performance of variant interpretation tools across different interpretation categories. The percentages indicate the corrected accuracy for each tool within the "Benign" and "Pathogenic" categories, and concordance for "Uncertain Significance" variants

Tool	Benign (%)	Uncertain significance (%)	Pathogenic (%)
Franklin	72.28	48.33	100
Varsome	97.62	30.00	98.18
InterVar	61.11	59.57	57.65
GeneBe	100	25.00	100

In this performance assessment, for pathogenic and benign variants, we report accuracy using ClinGen Expert Panels as our reference standard. ClinGen interpretations are supported by sufficient evidence to establish clinical actionability and represent definitive interpretations based on current scientific knowledge.

However, for variants of uncertain significance (VUS), we decided to use the term *concordance* rather than accuracy. This decision aims to acknowledge that these variants lack sufficient evidence for definitive interpretation, representing a deliberate state of uncertainty. Therefore, instead of measuring a tool's "correctness," our concordance metric measures how well the tools maintain the inherent uncertainty reported by ClinGen. This concordance is relevant because correctly maintaining VUS interpretations prevents premature clinical actions based on insufficient evidence, aligning with the principle of "do no harm" in medical practice. Furthermore, analyzing tool behavior with VUS provides valuable insights into interpretation tendencies and potential biases that may impact clinical decision-making, even without a definitive reference truth for these variants.

Most of the variant interpretation tools exhibited strong performance in interpreting pathogenic variants, with Franklin and GeneBe achieving 100% accuracy and Varsome closely following at 98.18%. Their effectiveness extended to benign variant interpretation, where all but one tool surpassed 70% accuracy. InterVar, however, underperformed in both categories, with accuracies of 57.65% for pathogenic and 40.28% for benign variants.

On the contrary, all tools struggled with VUS, indicating a common challenge in interpreting ambiguous cases. Interestingly,

Table 8. Distribution of interpretations assigned by variant interpretation tools to variants considered of uncertain significance by ClinGen Expert Panels

Tool	Benign (%)	Pathogenic (%)	Uncertain significance (%)
Franklin	1.67	50.00	48.33
Varsome	13.33	56.67	30.00
InterVar	6.38	34.04	59.57
GeneBe	20	55.00	25.00

InterVar showed the highest concordance for VUS at 59.57%, significantly outperforming its peers despite lower accuracy in other categories. This pattern suggests InterVar's tendency to interpret variants as uncertain, highlighting a potential bias in its algorithm.

The remaining tools demonstrated concordance rates below 50% for interpreting VUS. Due to this low concordance, we decided to investigate further how the interpretation tools interpret variants that ClinGen Expert panels consider to be of uncertain significance. Table 8 illustrates how each tool interprets the variants deemed uncertain by ClinGen.

Our analysis reveals that, except for InterVar, all tools show a tendency to interpret VUS as either pathogenic or likely pathogenic, which could lead to an increased rate of false positives. This trend highlights the need for cautious interpretation of the results generated by these tools to prevent unwarranted clinical interventions or treatments based on inaccurate interpretation.

Our findings underscore the importance of critical evaluation and expert validation of automated variant interpretations, particularly for those falling within the "uncertain significance" category. It also emphasizes the ongoing challenges in accurate variant interpretation and the need for continued refinement of computational prediction methods.

Discussion

Variant interpretation remains a complex, time-consuming process characterized by expert disagreement. Although automated interpretation tools have emerged as potential solutions for variant interpretation, our systematic identification and analysis of 13 freely accessible tools have revealed significant heterogeneity. This diversity spans multiple dimensions: scope of application, degree of automation, implementation strategies, data sources used, and the number of automated interpretation criteria.

The performance assessment of four broad-spectrum variant interpretation tools applied to 256 genetic variants, using ClinGen Expert Panels' evaluations as the benchmark, has revealed multiple insights that hold significant consequences for clinical application. While Franklin, VarSome, and GeneBe consistently demonstrated high accuracy (greater than 75%), InterVar showed significantly lower performance across all disease contexts (<70% in all cases).

Despite the general high accuracy of Franklin, VarSome, and GeneBe, it is important to note that we observed that tool performance is strongly context-dependent. For instance, GeneBe demonstrated higher accuracy for cardiomyopathy-associated variants, while Franklin and VarSome excelled in interpreting hereditary cancer and monogenic diabetes variants. These findings highlight the importance of tailoring tool selection to specific clinical contexts, whether using specialized tools or choosing the most suitable broad-application solution based on their demonstrated performance in specific disease domains, if such validation data are available.

Our analysis reveals two patterns that have different clinical implications. On the one hand, most tools (Franklin, VarSome, and GeneBe) demonstrate high overall accuracy with unambiguous cases, but they exhibit a systematic bias toward overclassifying VUS as pathogenic, with these tools interpreting >50% of VUS this way. On the other hand, InterVar shows the opposite pattern: while it has a more appropriate conservative behavior for VUS interpretation, it significantly underperforms in interpreting clearly pathogenic and benign variants.

These findings suggest that different tools present distinct types of clinical risks: high-performing tools risk overtreatment through VUS overclassification, while InterVar risks underdiagnosis through excessive conservatism in clear cases. Both patterns require expert oversight and highlight a fundamental challenge in automated variant interpretation, namely, achieving an optimal balance between sensitivity and specificity while maintaining clinical utility.

Our findings regarding VUS interpretation warrant additional methodological and clinical considerations. First, evaluating automated tools against VUS interpretation presents inherent limitations regardless of the reference standard chosen. Even expert panels like ClinGen explicitly acknowledge uncertainty in these interpretations, making any performance assessment fundamentally different from evaluating definitive pathogenic or benign variants. The observed low concordance rates should therefore be interpreted as classification tendencies, providing insights into tool behavior patterns that may influence clinical decision-making.

From a clinical perspective, maintaining the uncertainty inherent in VUS interpretation is crucial for appropriate patient care. The systematic tendency of most tools to interpret VUS as pathogenic or likely pathogenic (>50% in our analysis) represents a significant clinical concern, as it may lead to unnecessary interventions, psychological burden, and healthcare costs. Conversely, tools that maintain VUS interpretations, while potentially appearing less "decisive," actually preserve the appropriate clinical uncertainty that prevents premature medical actions based on insufficient evidence.

These findings highlight a fundamental challenge in automated variant interpretation: the need for balancing definitive answers with the current situation of insufficient evidence for some variants. Future tool development should prioritize maintaining appropriate uncertainty levels for ambiguous variants, rather than optimizing for apparent decisiveness that may, in fact, compromise clinical care.

In conclusion, our study represents an important step toward characterizing the landscape of variant interpretation tools. However, several limitations and future directions warrant discussion. While our findings provide valuable insights into current automated variant interpretation tools, a thorough evaluation of this rapidly evolving field presents ongoing methodological challenges. For instance, our selection criterion requiring web interfaces may have inadvertently excluded high-quality command-line tools. Although this criterion was established through meetings with clinical genetics experts who emphasized accessibility concerns, clinical genomics workflows inherently require substantial bioinformatics expertise for data processing, making the distinction between web-based and command-line tools less critical than initially considered. Future evaluations must address this limitation by including both modalities to provide a complete assessment of available tools regardless of their implementation format.

Building on these methodological considerations, a comprehensive analysis comparing general-purpose tools against domain-specific solutions remains necessary. Furthermore, given the rapid evolution of both available tools and existing ones, regular benchmarking against expert-curated datasets will be crucial for tracking improvements and identifying areas that need development.

Ultimately, significant progress has been made in automation, but the limitations of current tools require continued expert oversight. Looking ahead, the key challenge lies in striking an optimal balance between automation benefits and human expertise to fully realize precision medicine's potential in genomics. This will require close collaboration between tool developers, clinical experts, and researchers to ensure that automated systems effectively complement, rather than replace, human expertise in variant interpretation.

Key Points

- There is a diverse array of automated tools for the interpretation of genomic variants, ranging from broad-spectrum applications to highly specialized solutions.
- Most of the broad-spectrum tools analyzed achieve at least 80% accuracy for clearly pathogenic/benign variants but between 25% and 60% concordance rates for VUS.
- There is a widespread tendency to overclassify VUS as pathogenic (>50% in most cases).
- Despite advancements in automation, current limitations underscore the need for a hybrid approach where automated tools are used for an initial screening, but manual expert review is required for VUS and more complex cases.
- Further research is required to enhance VUS interpretation accuracy and standardize interpretation methodologies across different tools.

Acknowledgments

The authors thank anonymous reviewers for their valuable suggestions and feedback.

Author contributions

Mireia Costa (Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing), Alberto Garca (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Ana Len (Conceptualization, Writing—original draft, Writing—review & editing), and Oscar Pastor (Supervision, Writing—original draft, Writing—review & editing)

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work is supported in part by the Generalitat Valenciana through the CoMoDiD project (CIPROM/2021/023) and the predoctoral project granted to M.C. (ACIF/2021/117). Financed with Aid to First Research Projects (PAID-06-24), Vice-rectorate of Research of the Universitat Politècnica de València (UPV). This work is part of the Horizon Europe project BETTER, funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101136262.

Competing interests

No competing interest is declared.

Data availability

All the data is available as supplementary material.

References

1. Niehaus A, Azzariti DR, Harrison SM. *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;**21**:1699–701. https://doi.org/10.1038/s41436-018-0432-7
2. Richards S, Aziz N, Bale 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. https://doi.org/10.1038/gim.2015.30
3. Wang Z, Norris S, Bero L. The advantages and limitations of guideline adaptation frameworks. *Implementation Sci* 2018;**13**:05. https://doi.org/10.1186/s13012-018-0763-4
4. Li MM, Datto M, Duncavage EJ. *et al.* Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn* 2017;**19**:4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002
5. Riggs ER, Andersen EF, Cherry AM. *et al.* Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* 2020;**22**:245–57. https://doi.org/10.1038/s41436-019-0686-8
6. Amendola L, Ferar K, Biesecker L. *et al.* Variant classification concordance using the ACMG-AMP variant interpretation guidelines across nine genomic implementation research studies. *Am J Hum Genet* 2020;**107**:932–41. https://doi.org/10.1016/j.ajhg.2020.09.011
7. Costa M. Data for the "the promises and pitfalls of automated variant interpretation: a comprehensive review" paper. https://doi.org/10.5281/zenodo.16631994
8. Ravichandran V, Shameer Z, Kemel Y. *et al.* Toward automation of germline variant curation in clinical cancer genetics. *Genet Med* 2019;**21**:1668. https://doi.org/10.1038/s41436-019-0490-5
9. Peng J, Xiang J, Jin X. *et al.* VIP-HL: Semi-automated ACMG/AMP variant interpretation platform for genetic hearing loss. *Hum Mutat* 2021;**42**:1567–75. https://doi.org/10.1002/humu.24277
10. Li H, Liu S, Wang S. *et al.* Cancer SIGVAR: a semi-automated interpretation tool for germline variants of hereditary

cancer-related genes. *Hum Mutat* 2021;**42**:359–72. https://doi.org/10.1002/humu.24177

11. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet* 2017;**100**: 267–80. https://doi.org/10.1016/j.ajhg.2017.01.004
12. Whiffin N, Walsh R, Govind R. *et al*. CardioClassifier: disease- and gene-specific computational decision support for clinical genome interpretation. *Genet Med* 2018;**20**:1246–54. https://doi.org/10.1038/gim.2017.258
13. Stawiński P, Płoski R. Genebe.net: Implementation and validation of an automatic ACMG variant pathogenicity criteria assignment. *Clin Genet* 2024;**106**:119–26. https://doi.org/10.1111/cge.14516
14. Xavier A, Scott R, Talseth-Palmer B. TAPES: a tool for assessment and prioritisation in exome studies. *PLoS Comput Biol*, 2019;**15**:e07453, 10, https://doi.org/10.1371/journal.pcbi.1007453.
15. He M, Li Q, Yan M. *et al*. Variant interpretation for cancer (VIC): a computational tool for assessing clinical impacts of somatic variants. *Genome Med* 2019;**11**:08. https://doi.org/10.1186/s13073-019-0664-4
16. Munté E, Feliubadaló L, Pineda M. *et al*. vaRHC: an R package for semi-automation of variant classification in hereditary cancer genes according to ACMG/AMP and gene-specific ClinGen guidelines. *Bioinformatics* 2023;**39**:03. https://doi.org/10.1093/bioinformatics/btad128
17. Gurbich T, Ilinsky V. ClassifyCNV: a tool for clinical annotation of copy-number variants. *Sci Rep* 2020;**10**:11.
18. Landrum MJ, Lee JM, Benson M. *et al*. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;**46**:D1062–7. https://doi.org/10.1093/nar/gkx1153
19. Mccarthy D. *et al*. Choice of transcripts and software has a large effect on variant annotation. *Genome Med* 2014;**6**:26. https://doi.org/10.1186/gm543
20. O'Leary NA. *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2015;**44**:D733–45. https://doi.org/10.1093/nar/gkv1189
21. Harrison P. *et al*. Ensembl 2024. *Nucleic Acids Res* 2023;**52**:D891–9.
22. Lansdon LA, Cadieux-Dion M, Yoo B. *et al*. Factors affecting migration to GRCh38 in laboratories performing clinical next-generation sequencing. *J Mol Diagn* 2021;**23**:651–7. https://doi.org/10.1016/j.jmoldx.2021.02.003
23. Genovese G, Rockweiler NB, Gorman BR. *et al*. BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. *Bioinformatics* 2024;**40**. https://doi.org/10.1093/bioinformatics/btae038
24. Luu P-L, Ong P-T, Dinh T-P. *et al*. Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. *NAR Genomics Bioinf* 2020;**2**:08. https://doi.org/10.1093/nargab/lqaa054
25. Park K-J, Yoon Y, Park J-H. Evaluation of liftover tools for the conversion of Genome Reference Consortium Human Build 37 to Build 38 using ClinVar variants. *Genes* 2023;**14**:1875. https://doi.org/10.3390/genes14101875

Briefings in Bioinformatics, 2025, **26(5)**, bbaf545
https://doi.org/10.1093/bib/bbaf545
Review