Freie Universität Berlin

Bachelor thesis at the Department of Mathematics and Computer Science of Freie Universität Berlin

Bioinformatics

Automated Classification of Sequence Variants According to ACMG Criteria

Dzmitry Hramyka

First examiner: Dr. Manuel Holtgrewe

Second examiner: Prof. Dr. Martin Vingron

Berlin, August 18, 2024

Preface

The assessment of sequence variants, as guided by the ACMG Criteria established by Richards et al. in 2015, has become a standard for clinical interpretation within genomic medicine. This ruleset comprises several criteria corresponding to various consequences of mutations, enabling unified classification results. However, (i) the manual assessment of the ACMG criteria suffers from a lack of scalability; (ii) with the ongoing updates to the ruleset, exemplified by the ACGS criteria, there's a constant need for continual verification and review of assessments, reinforcing the need for the automation of the assessment process. Recent advances in bioinformatics have introduced tools such as AutoPVS1, InterVar, Genebe, and VarSome, designed to automate these assessments. Nevertheless, all of these services have certain limitations, which prevent them from being considered as complete up-to-date automation solutions. This work proposes the development of an open-source tool aimed at the comprehensive automation of the ACMG criteria, overcoming the limitations of existing systems by ensuring up-to-date variant interpretation.

Contents

1 Introduction

1.1 Motivation

The rapid advancement in broad genetic testing, especially through exome sequencing (ES) and genome sequencing (GS), has led to a marked increase in the detection and analysis of genetic variations [1]. ES primarily enriches the coding regions of genes, enabling the identification of variants that directly impact the coding sequence. Conversely, GS examines the complete deoxyribonucleic acid (DNA) sequence, capturing both coding and non-coding regions, thus providing a comprehensive view of genetic variations. These technologies have significantly enhanced the ability to detect both small (≤ 50 base pairs) and structural (>50 base pairs) genetic variants, leading to a substantial increase in the number of known genetic variants catalogued in genomic databases.

This significant growth in sequencing capabilities, along with the increased volume of data, presents both opportunities and challenges in clinical settings. Clinicians now have access to thousands of variants per individual, but this large volume requires sophisticated analytical tools to accurately differentiate pathogenic variants from benign ones. Correct interpretation of these variants is crucial, as they are often directly linked to disease phenotypes and can influence treatment plans.

The American College of Medical Genetics and Genomics (ACMG), in collaboration with the Association for Molecular Pathology (AMP), has established standards and guidelines for the clinical interpretation of sequence variants based on 28 criteria [2]. These guidelines offer a systematic framework for assessing the pathogenicity of variants, utilising a range of evidence, including population data, computational predictions, and functional analyses. Yet, the complexity and quantity of data produced by GS and ES, along with variations in classifications by different clinical experts, highlight the essential need for computational tools for automated variant interpretation.

A significant challenge in implementing a unified computational algorithm is that the standards were initially designed for human interpretation, not for automated systems. Nonetheless, some recent studies have introduced quantitative thresholds for terms such as "well-known" and "hot-spot," facilitating the semi-automated classification of some of these criteria [3]. Despite significant progress, existing tools like InterVar [4], AutoPVS1 [5], and VarSome [6] still have notable limitations. These platforms often do not cover all ACMG criteria comprehensively (as seen with AutoPVS1), are challenged by the need for frequent updates in response to new genetic insights (as with InterVar), and their proprietary nature can hinder widespread adoption and customization (as with VarSome). These issues underscore the urgent need for an open-source tool that can integrate the latest genetic insights, provide comprehensive coverage of the ACMG criteria, and offer the flexibility to adapt to future advancements.

1.2 Objectives

This thesis introduces AutoACMG, an open-source software developed in response to the evolving challenges in genomic variant interpretation. The tool automates the application of the selected ACMG (American College of Medical Genetics and Genomics) criteria, and incorporates the latest guideline updates from the ClinGen (Clinical Genome Resource) [3,6–11] for the automated criteria.

AutoACMG is designed to address the limitations of existing tools such as InterVar, which lacks recent updates, and VarSome, which has constraints due to being commercial software. AutoACMG provides a comprehensive, open-source platform that integrates the newest data sources and implements algorithms based on the latest guideline findings. The primary goal of this work is to develop a robust software tool that supports continuous updates and customization. The secondary goal is to evaluate AutoACMG against established methodologies like InterVar, AutoPVS1, and VarSome, focusing on its precision. Upon successful validation, AutoACMG is anticipated to be integrated into the REEV [12] software suite, thereby enhancing its functionality as a versatile tool for genetic variant analysis.

1.3 Organisation

The thesis is structured into three main sections. In Section 2, "Materials and Methods" fundamental terms and concepts related to the assessment of ACMG criteria are introduced. The technical implementation of the AutoACMG tool is described, and the methodologies employed for evaluating its performance against other software tools are presented. The next Section 3, "Results" provides a comprehensive summary of the AutoACMG tool's functionality. It presents the outcomes of the tool's application and includes a detailed analysis comparing AutoACMG to existing methodologies like InterVar, AutoPVS1, and VarSome. Finally, Section 4, "Discussion and Outlook" evaluates the software, discusses unresolved issues and challenges, and outlines potential future development ideas.

2 Materials and Methods

2.1 Terms and Concepts

2.1.1 Sequencing

Genome sequencing encompasses various methodologies used to explore the entire genetic makeup of an organism, focusing on both coding and non-coding regions of deoxyribonucleic acid (DNA). This process is fundamental in identifying the **genomic variations** (variants) that may influence health and disease. Within genome sequencing, **Genome Sequencing** (GS) and **Exome Sequencing** (ES) are critical techniques. GS provides a comprehensive overview of an entire genome, offering insights into both coding regions and the vast expanses of non-coding DNA that may regulate gene activity. ES, however, targets only the **exons**, or the sequences within genes that directly code for proteins, thus spotlighting the parts of the genome most likely to affect biological functions directly.

This thesis focuses on the human genome, which is diploid with 23 pairs of chromosomes. Specific locations on these chromosomes are called loci, defined by their start and end positions. Differences at the same loci on two chromosomes (one maternal and one paternal) are called alleles, representing genetic diversity among individuals. Further in this thesis genes are defined as functional units of the genome, consisting of exons and introns, where exons include encoded protein segments and untranslated regions (UTR), while introns are noncoding and removed during RNA processing. Transcription converts genes into transcripts, with protein-coding genes subsequently translated into proteins.

Zygosity is essential in genetics, describing the variation in allele pairings at a specific locus on a chromosome. The three primary types of zygosity are **homozygous**, **heterozygous**, and **hemizygous**. A homozygous genotype has two identical alleles at a locus, leading to consistent expression of a particular trait. A heterozygous genotype has two different alleles at a locus, where typically the **dominant** trait is expressed, but it can also result in a combination of traits or codominance. **Recessive** traits require both alleles to be identical to be expressed. Hemizygosity occurs when only one allele is present at a locus, usually due to deletion or, in males, for genes on the X chromosome that lack a corresponding allele on the Y chromosome.

The **transcript**, an RNA copy of a gene's code, is central to gene expression. It undergoes processing at **splice sites** within the gene where non-coding introns are excised and the coding exons are spliced together. This splicing results in a continuous **messenger RNA** (**mRNA**) sequence that will direct the assembly of amino acids into proteins.

Reference genomes, like **GRCh37** and **GRCh38**, established by the **Genome Reference Consortium**, serve as essential tools for aligning and identifying genetic variants. These reference genomes provide a standardised framework, allowing researchers and clinicians to accurately compare genetic sequences and detect variations consistently.

2.1.2 Variants

In genetics, **variants** represent differences between a donor's or patient's genome and a reference genome sequence. These variants are broadly classified into **sequence variants** and **structural variants**. Sequence variants alter a small number of nucleotides and typically have localised effects on the genome. Structural variants, on the other hand, involve more extensive changes to the DNA structure, impacting larger segments (greater than 50 base pairs) of the genome. These structural changes can significantly affect gene function and regulation by disrupting genes, altering gene dosage, or changing the spatial organisation of the genome, such as by destroying isolator motifs and affecting genome architecture. Structural variants are generally rarer due to their larger size. The exact distribution of genetic variants in the human genome is illustrated in Figure 1.

Building on the general classifications of genetic variants, sequence variants include **single nucleotide variants** (**SNVs**), and insertions or deletions (**indels**). SNVs, which involve changes to a single nucleotide, are the most prevalent types of genetic variations and constitute the majority of variations observed in human genomes. Indels, which encompass both the addition and loss of a small number of bases, although less common, still represent a significant portion of sequence variants.

In contrast, structural variants involve more substantial modifications, including **deletions**, **insertions**, and complex sequence alterations, and are less common than sequence variants. Other types of structural variants, such as **substitutions**, **translocations**, and more complex rearrangements, are very rare but still present in genomes.

2.2.4 Concepts in Classification Algorithm

In genetic analysis, cellular processes are crucial in predicting the impact of variants. Variants are broadly categorised into **loss-of-function** (LoF) and **gain-of-function** (GoF) groups. LoF variants reduce or eliminate a gene's function, often leading to significant phenotypic consequences, especially in essential genes. These include nonsense variants, which introduce premature stop codons and trigger nonsense-mediated decay (NMD) to degrade faulty mRNA, preventing harmful protein production. **Missense** variants, on the other hand, substitute one amino acid for another, potentially disrupting protein function. However, these variants can also result in **synonymous** DNA changes that do not alter the protein, thus tending to be benign.

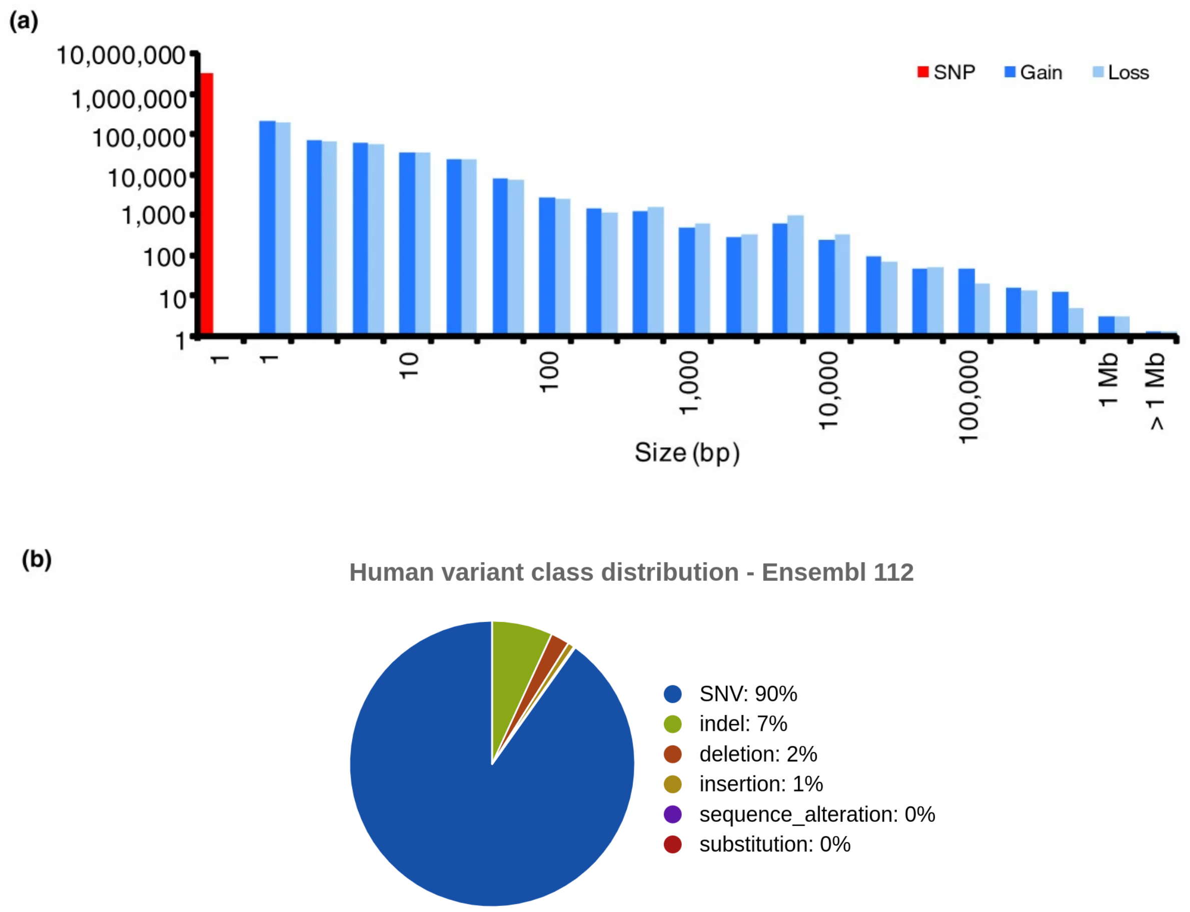

Figure 1. Distribution of Genetic Variants in the Human Genome.
1a: Size Distribution of Genetic Variants — Showcases a non-redundant spectrum of SNV and copy number variation (CNV) sizes with a detailed breakdown of the proportion of genomic gains to losses [13].
1b: Type Distribution of Genetic Variants — Illustrates the variety of genetic variants categorised by the type [14].

GoF variants, though rarer and harder to predict, result in increased or novel gene activity, sometimes causing diseases through abnormal pathway activation (Chronic Myeloid Leukaemia [15]). Gene-disease associations often depend on the specific type of variant within a gene; for instance, different conditions can arise from nonsense versus missense variants in the same gene.

Additionally, certain genomic regions influence variant impact. **Tandem repeat** regions, where nucleotide motifs repeat consecutively, are linked to genetic disorders when repeat numbers vary (Huntington's disease [16]). Variants in **UniProt** domains, specific protein regions tied to particular functions, can significantly alter protein function and disrupt cellular processes, tending to be pathogenic (BRCA1-associated RING domain in BARD1 [17]).

2.2 Theoretical Assumptions

2.2.1 Variant Consequences

ACMG guidelines include separate rulesets for classifying copy number variants (CNVs) [18] and sequence variants [2]. In this thesis, the focus is on the AutoACMG algorithm implementation, which is designed specifically for the classification of sequence variants. A critical aspect of this classification involves considering the position of the variant relative to the transcript structure, as this can significantly influence gene function and expression. The consequences of these variants, as depicted in Figure 2, span various classes and impacts based on their location and nature.

Nonsense and **frameshift** variants, such as "stop gained" and "frameshift" mutations, introduce disruptions in the protein product. Stop gained variants insert premature stop codons into the coding sequence, leading to truncated proteins that are often non-functional. Conversely, frameshift variants result from insertions or deletions whose lengths are not divisible by three (the length of a codon) altering the normal reading frame and profoundly changing the amino acid sequence and the resulting protein.

Splicing variants affect the initial stages of protein synthesis and include changes in the 5' UTR regions, "start lost" mutations, and alterations at splice donor and acceptor sites. These variants can influence the splicing process and potentially alter the final protein product by modifying how exons and introns are read. "Start lost" variants specifically compromise the initiation of translation, by impacting the start codon of some exon.

Other important variations within the coding sequences of exons, such as **missense** mutations, **inframe insertions**, and **inframe deletions**, can affect protein functionality to some extent. Missense variants change a single amino acid in the protein, while inframe insertions and deletions add or remove amino acids without disrupting the reading frame. These types of variants are generally less pathogenic compared to nonsense and frameshift variants but can still lead to altered phenotypes.

Additionally, regulatory and terminal modifications such as **3' UTR** variants and **"stop lost"** variants have mostly pathogenic impact on protein products. 3' UTR variants can affect gene regulation and mRNA stability, while "stop lost" variants extend the protein beyond its normal endpoint, potentially introducing new amino acid sequences with varied effects [19,20].

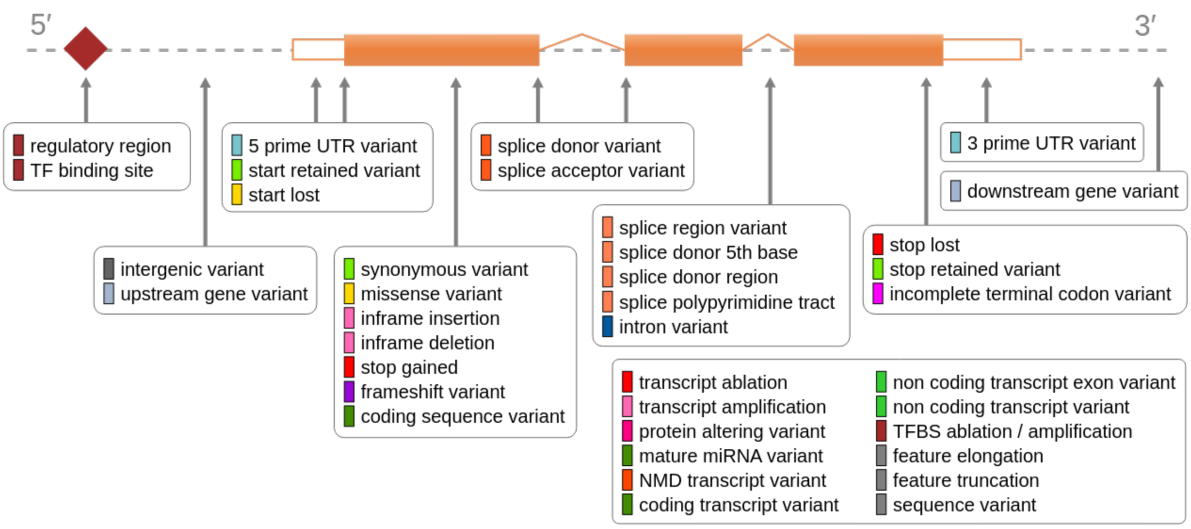

Figure 2. Location of various variant types relative to the transcript structure — Illustrates the positioning of various genetic variants within the transcript structure, highlighting how these variants correlate with gene function. [14]

2.2.2 ACMG Ruleset

To standardise the clinical interpretation of genetic variants, the **American College of Medical Genetics and Genomics** introduced standards for interpreting sequence variations in 2000 and 2007 [21,22]. With the rapid expansion of available genetic data, there was a clear need to refine these standards. In response, ACMG, in collaboration with the **Association for**

Molecular Pathology (AMP), published updated guidelines in 2015 [2]. These guidelines introduced a five-tiered classification system for variant interpretation—benign (B), likely benign (LB), uncertain significance (US), likely pathogenic (LP), and pathogenic (P). This system is underpinned by 28 specific criteria designed to provide a comprehensive assessment of variants. The sections below detail the ACMG Ruleset more extensively.

2.2.2.1 ACMG Guidelines

The core principle of the ACMG guidelines is to assess variants from multiple perspectives, ultimately classifying them into one of five categories. The 28 criteria derive from diverse data sources, including population data, in silico predictive data, functional data, and segregation data, as illustrated in Figure 3. To ensure a balanced evaluation, each criterion is assigned a level of evidence—Very Strong (PVS1), Strong (PS1-4), Moderate (PM1-6), or Supporting (PP1-5) for pathogenic assessments, and Stand Alone (BA1), Strong (BS1-4), or Supporting (BP1-6) for benign assessments. The weight of each criterion may be adjusted based on the strength of the clinical evidence presented.

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder *BA1/BS1* **OR** observation in controls inconsistent with disease penetrance *BS2*			Absent in population databases *PM2*	Prevalence in affecteds statistically increased over controls *PS4*	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product *BP4* Missense in gene where only truncating cause disease *BP1* Silent variant with non predicted splice impact *BP7*	Multiple lines of computational evidence support a deleterious effect on the gene /gene product *PP3*	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before *PM5* Protein length changing variant *PM4*	Same amino acid change as an established pathogenic variant *PS1*	Predicted null variant in a gene where LOF is a known mechanism of disease *PVS1*
Functional Data	Well-established functional studies show no deleterious effect *BS3*		Missense in gene with low rate of benign missense variants and path. missenses common *PP2*	Mutational hot spot or well-studied functional domain without benign variation *PM1*	Well-established functional studies show a deleterious effect *PS3*	
Segregation Data	Non-segregation with disease *BS4*		Co-segregation with disease in multiple affected family members *PP1*	Increased segregation data		
De novo Data				*De novo* (without paternity & maternity confirmed) *PM6*	*De novo* (paternity & maternity confirmed) *PS2*	
Allelic Data		Observed in *trans* with a dominant variant *BP2* Observed in *cis* with a pathogenic variant *BP2*		For recessive disorders, detected in *trans* with a pathogenic variant *PM3*		
Other Database		Reputable source w/out shared data = benign *BP6*	Reputable source = pathogenic *PP5*			
Other Data		Found in case with an alternate cause *BP5*	Patient's phenotype or FH highly specific for gene *PP4*			

Figure 3. Evidence Level and Data Source Distribution of ACMG Criteria. This chart categorises ACMG criteria by evidence type and strength for benign (left) and pathogenic (right) assertions, detailing categories like benign strong (BS), benign supporting (BP), and various pathogenic levels (PM, PP, PS, PVS), along with factors such as family history (FH), loss of function (LOF), and minor allele frequency (MAF). [2]

The final classification within the five-tier system is determined by a combination of these criteria. For instance, to achieve a classification of **"Pathogenic"**, several configurations are possible: (i) One Very Strong (PVS1) and either ≥1 Strong, ≥2 Moderate, 1 Moderate plus ≥1 Supporting, or ≥2 Supporting criteria; (ii) ≥2 Strong criteria; or (iii) 1 Strong plus either ≥3

Moderate, 2 Moderate plus ≥2 Supporting, or 1 Moderate plus ≥4 Supporting criteria. The complete methodology for these assignments is detailed in Table 1.

Table 1. Rules for combining criteria to classify sequence variants. [2]

Category	Criteria
Pathogenic	1 Very Strong (PVS1) AND ≥1 Strong (PS1–PS4)
Pathogenic	1 Very Strong (PVS1) AND ≥2 Moderate (PM1–PM6)
Pathogenic	1 Very Strong (PVS1) AND 1 Moderate (PM1–PM6) AND 1 Supporting (PP1–PP5)
Pathogenic	1 Very Strong (PVS1) AND ≥2 Supporting (PP1–PP5)
Pathogenic	≥2 Strong (PS1–PS4) OR 1 Strong (PS1–PS4) AND (≥3 Moderate (PM1–PM6) OR 2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5) OR 1 Moderate (PM1–PM6) AND ≥4 Supporting (PP1–PP5))
Likely Pathogenic	1 Very Strong (PVS1) AND 1 Moderate (PM1–PM6)
Likely Pathogenic	1 Strong (PS1–PS4) AND 1-2 Moderate (PM1–PM6)
Likely Pathogenic	1 Strong (PS1–PS4) AND ≥2 Supporting (PP1–PP5)
Likely Pathogenic	≥3 Moderate (PM1–PM6)
Likely Pathogenic	2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5)
Likely Pathogenic	1 Moderate (PM1–PM6) AND ≥4 Supporting (PP1–PP5)
Benign	1 Stand-Alone (BA1)
Benign	≥2 Strong (BS1–BS4)
Likely Benign	1 Strong (BS1–BS4) AND 1 Supporting (BP1–BP7)
Likely Benign	≥2 Supporting (BP1–BP7)

2.2.2.2 Criteria Description

Each of the 28 ACMG criteria is associated with specific attributes of a variant, based on the data source. Below is a brief overview of each criterion.

PVS1: Loss of Function Variants
PVS1 assess null and loss-of-function variants (e.g., nonsense, frameshift mutations) that disrupt gene function. This includes evaluation potential escape from nonsense-mediated decay and checking if splice site variants are in critical domains.

PS1 Same Amino Acid Change
PS1 assigns strong evidence if a missense variant results in the same amino acid change as a known pathogenic variant, unless influenced by direct DNA interaction.

PS2/PM6: De Novo Variants
PS2 and PM6 provide strong evidence when de novo mutations occur in dominant disorder-linked genes and are absent in both parents, correlation with the patient's clinical presentation.

PS3/BS3: Functional Studies Evidence
Functional studies demonstrating a deleterious effect on gene/protein function support pathogenicity (PS3), while those showing no adverse effects suggest benignity (BS3).

PS4, PM2, BA1, BS1, BS2: Allele Frequency Data
Population frequency data is essential for differentiating benign from pathogenic variants. Common variants in healthy individuals are benign (BS1), especially if their frequency exceeds disease prevalence (BA1). Rare variants in the general population but prevalent in affected individuals are likely pathogenic (PM2). Zygosity evaluations add benign evidence for recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) conditions (BS2). High relative risk in Mendelian disorders supports pathogenicity (PS4).

PM1: Mutational Hot Spot
Missense variants in critical protein domains known to be essential for function are considered moderately pathogenic. These regions, termed mutational hot spots, have all identified missense variants shown to be pathogenic.

PM3, BP2: Cis/Trans Testing
Testing if variants occur in cis (same gene copy) or trans (different gene copies) helps assess pathogenicity. Two heterozygous variants in a gene for a recessive disorder, where one is pathogenic, indicate moderate pathogenicity for the other in trans (PM3). Conversely, finding the second variant in cis supports benign evidence (BP2).

PM4, BP3: Protein Length Changes
Alterations in amino acids, particularly in stop codons, can disrupt protein function by changing protein length. Moderate pathogenic evidence (PM4) is applied to large, conserved in-frame deletions/insertions, while smaller, non-conserved changes support benign evidence (BP3).

PM5: Novel Missense
A novel missense variant at the same position as another pathogenic missense change is moderate evidence of pathogenicity. Different amino acid changes can lead to varying phenotypes, and a novel change more conserved than a known pathogenic change may not be pathogenic.

PP1, BS4: Segregation Analysis
Segregation analysis determines if a genetic variant co-segregates with a disease phenotype within a family, indicating a potential link. Effective segregation with disease phenotypes in diverse families provides moderate to strong pathogenic evidence, while a lack of segregation strongly suggests benignity.

PP2, BP1: Variant Spectrum
The variant spectrum criterion considers known distributions of pathogenic and benign variations within a gene. For genes where missense mutations commonly cause disease and benign variants are rare, a novel missense change is supporting evidence for pathogenicity. In genes typically affected by truncating variants, missense changes are likely benign.

PP3, BP4: In Silico Analysis
In silico predictions are crucial in variant classification, where computational evidence must be carefully evaluated. Multiple computational models concurring in their predictions provide supportive evidence for pathogenicity or benign nature.

PP4: Phenotype Matching
A patient’s phenotype aligning with the clinical features of a gene can provide supportive evidence under specific conditions. If the phenotype closely matches a well-defined syndrome with minimal overlap with other conditions and the gene shows high clinical sensitivity, this supports the variant's pathogenicity (PP4). This can be strengthened if the gene shows limited benign variation in large population studies and the family history aligns with the gene's inheritance pattern.

PP5, BP6: Reputable Source Validation
Pathogenicity classifications from reputable clinical laboratories are often cited in genetic databases. Recent classifications from such sources are considered supporting evidence.

BP5: Alternate Locus Observations
Variants observed alongside an alternative genetic cause of disease generally suggest benignity. In dominant disorders, a variant found with a known pathogenic variant might still contribute to disease severity. In recessive disorders, a novel variant's benign classification requires cautious interpretation and additional evidence.

BP7: Synonymous Variants
Synonymous variants, traditionally considered benign, require careful interpretation due to potential splicing impacts. These changes can disrupt gene function, particularly in genes where loss of function causes disease. If computational predictions and evolutionary conservation do not suggest splicing impact, variants are likely benign.

The original ACMG criteria were designed for manual evaluation by experts, and certain criteria remain challenging to automate fully. Specifically, criteria such as **PS2/PM6** (de novo mutations), **PS3/BS3** (functional studies), **PM3/BP2** (cis/trans testing), and **PP1/BS4** (segregation analysis) have not been automated due to their reliance on complex clinical data, experimental evidence, and familial information that require expert interpretation.

2.2.2.3 Separate Guidelines for PVS1

Among the ACMG criteria, **PVS1** stands out for its complexity and importance in classifying null variants that typically cause loss of function. These include nonsense mutations, frameshift indels, and canonical splice site alterations, which disrupt gene function often through mechanisms like nonsense-mediated decay. In 2018, a decision tree was introduced to refine the assessment of these variants [7], evaluating factors such as the variant's gene location, its impact on splicing, and the presence of alternative start codons that might mitigate the loss of function. This tree assigns a tailored PVS1 strength rating, from **"Very Strong"** to **"Supporting"**.

The updated guidelines enhance PVS1 assessments by integrating criteria to determine the evidence strength for classifying null variants. Key considerations include whether truncating variants **escape NMD**, the location of the variant within **the final exon or the last 50 base pairs of the penultimate exon**, and the potential production of **truncated proteins**. These guidelines also emphasise assessing the biological relevance of the transcript and the functional significance of the affected region, evaluating the possibility of exon skipping or cryptic splice site use, and the necessity for detailed functional assays for variants that escape NMD but may still produce functional proteins. The refined guidelines introduce varying strengths for PVS1, supporting more precise and context-sensitive interpretations of null variants across different genetic contexts and diseases.

2.2.2.4 ClinGen Modifications

Building upon the existing American College of Medical Genetics and Genomics (ACMG) guidelines, the Clinical Genome Resource (ClinGen) has made several significant modifications to refine and update the criteria used for sequence variant interpretation.

One of the notable changes introduced by ClinGen involves the discontinuation of the **PP5** and **BP6** criteria, which were previously used to classify variants based on reputable source information [23]. The removal of these criteria is part of an effort to ensure that genetic variant classifications are based on transparent and replicable evidence, rather than on potentially unverifiable sources.

Additionally, ClinGen has introduced an exception list for the **BA1** criterion [9], which is used to classify a variant as benign if it is common in a healthy population. The exception list clarifies situations where a variant previously thought to be benign due to its frequency may still be considered for pathogenic classification under specific circumstances, enhancing the accuracy of variant interpretation.

Significant advancements have also been made in the interpretation of **splicing recommendations** [11]. ClinGen has developed guidelines that integrate both predicted and observed impacts on splicing, offering a more comprehensive framework for assessing variants that may affect RNA splicing processes. These guidelines aim to improve predictions of the functional effects of intronic and exonic changes that could disrupt normal splicing.

ClinGen also has addressed the calibration of computational tools used to predict the pathogenicity of missense variants (**PP3** and **BP4**) [3]. These guidelines offer a framework for evaluating the performance of in silico tools, aiming to standardise the use of computational evidence in the variant classification process. By calibrating these tools against known variant datasets, ClinGen seeks to enhance the reliability of predictions made regarding variant pathogenicity.

2.2.2.5 VCEP Gene-Specific Curated Modifications

The Variant Curation Expert Panels (VCEP) within ClinGen have refined the ACMG guidelines to improve the accuracy of sequence variant interpretation through **gene-specific modifications**. These adjustments cater to the unique characteristics of individual genes, providing more precise thresholds and general guidelines for criteria assessment. However, these VCEP guidelines are not implemented in the current version of AutoACMG due to the lack of machine-readable specifications.

2.3 Data

2.3.1 Data for Predictions

The successful implementation of the AutoACMG tool relies heavily on comprehensive genetic data from various databases. For gene-focused information, **OMIM** (Online Mendelian Inheritance in Man) provides detailed gene-phenotype relationships, **Decipher** (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources) helps in interpreting variants in rare diseases, and **Orphanet** offers information on orphan drugs and rare diseases.

Variant-focused databases include **gnomAD** (Genome Aggregation Database), which provides population frequency data by aggregating genome and exome sequencing data, and **dbSNP** (Single Nucleotide Polymorphism Database), which catalogues short genetic variations. **UniProt** offers extensive data on protein sequences and functions, essential for understanding the impact of amino acid changes. The **Human Phenotype Ontology** (HPO) provides a structured vocabulary of phenotypic abnormalities encountered in human disease, which is helpful for correlating genetic variations with clinical data. Additionally, **dbNSFP** integrates scores from various computational predictive tools such as **REVEL**, **PolyPhen**, **CADD**, **BayesDel**, **PrimateAI**, **FATHMM**, and **PhyloP**, all of which predict the functional effects of variants.

Access to comprehensive resources is essential for implementing the AutoACMG tool effectively to predict the clinical significance of genetic variants. Such diverse and extensive data ensures that AutoACMG's interpretations are well-supported by empirical evidence, thereby enhancing the reliability and utility of genetic testing in clinical and research settings.

2.3.2 Data Sources

In AutoACMG, the REEV web service [12] acts as a gateway to aggregate and access genetic data from various databases. The **"annonars"** microservice pulls gene-specific and variant-specific data from sources such as gnomAD, dbSNP, and dbNSFP, alongside computational scores that are critical for variant interpretation. Additionally, **"mehari"**, another microservice within REEV, provides detailed transcript-specific information including HGVS notations, variant effect predictor (VEP) consequences, and feature tags, such as whether a transcript is MANE selected.

For the resolution of sequence variants from HGVS notations, the **"dotty"** microservice is utilized, ensuring that variants are accurately interpreted based on the latest genomic assemblies. An essential component of variant prediction, particularly for assessing **splicing implications**, involves the MaxEnt package [24,25], which utilises RefSeq sequences for both GRCH37 and GRCH38. Furthermore, to aid in the identification of **repetitive masked regions** and **critical functional domains** within proteins, we utilise preprocessed Uniprot and RepeatMasker (RMSK) tracks from the UCSC genome browser [26].

2.3.3. Data Limitations

In AutoACMG, although a wide range of genetic data is used for variant interpretation, certain types of data are not accessible, limiting the full automation of ACMG criteria. Specifically, we lack critical **familial** and **clinical** information such as maternity/paternity validation, **functional studies** assessing gene or protein function, **prevalence** differences between affected individuals and control groups, **testing of parents** to confirm de novo status of variants, **segregation** analysis within families, and detailed family medical histories. These data are essential for a comprehensive assessment of variant pathogenicity and typically require direct clinical input and validation.

Consequently, the absence of detailed clinical and familial data limits the scope of AutoACMG, making it a semi-automated classification algorithm. The manual evaluation by clinicians is necessary for criteria PS2, PS3, PS4, PM3, PM6, PP1, PP4, BS3, BS4, BP2, and BP5. Table 2 outlines these ACMG criteria and explains why each cannot be fully automated.

2.4 Selection of tools

A variety of **computational tools** have already been developed to automate the application of ACMG guidelines. Among these, AutoPVS1 has significantly influenced the implementation of the PVS1 criteria in AutoACMG. The methodologies and threshold definitions provided by the documentation of VarSome [6] and the InterVar [4] paper have also been helpful in refining algorithmic approaches for assessing variant pathogenicity.

Further exploration through a Google search on July 15, 2024, for **"automatic ACMG guidelines"** reveals several other notable tools including Franklin by Genoox [27], MAGI-ACMG [28], and GenOtoScope [29], each offering functionalities on automated guideline application. Another inquiry via Google Scholar using the same query highlights systems such as GeneBe.net [30], vaRHC [31], AutoCNV [32], CardioVAI [33], and VIP-HL [34]. Additionally, a PubMed search on the same date and query found MARGINAL [35], another tool for automatic classification of variants in BRCA1 and BRCA2 genes.

Table 2. ACMG Criteria Requiring Manual Evaluation. Outline ACMG criteria that require manual evaluation, detailing the specific data sources needed and explaining why automation is not possible for these criteria.

ACMG Criteria	Information Sources/Databases Used	Notes on Non-Automatability
PS2, PS3, PS4	Literature and functional studies	These criteria require clinical validation of variant pathogenicity through family studies, which cannot be automated due to the need for patient and family-specific data.
PM3, PM6	Internal family databases	Testing for de novo status or cis/trans phase requires specific parental data which is often not available in public databases, necessitating manual verification.
PP1, PP4	Family studies, clinical reports	Segregation analysis and phenotype correlation require detailed and often confidential family medical histories and phenotype data that are not typically accessible or automatable.
BS3, BS4	Functional studies, clinical databases	Functional assays and their interpretation often need detailed laboratory results and expert analysis to determine the impact on gene or protein function, which cannot be automated.
BP2, BP5	Literature, clinical databases	These criteria involve assessment of benign impact often requiring detailed clinical insights and longitudinal studies, which go beyond genetic data alone.

The selection of comparative tools for AutoACMG was strategic, emphasising open-source availability, robust documentation, and ease of testing to ensure a comprehensive evaluation. InterVar and GeneBe stood out as the principal concurrent tools, both offering well-documented frameworks that are accessible and testable, aligning with the open-source principles of AutoACMG.

2.5 Implementation of AutoACMG

2.5.1 Software, Tools and External Services

In developing AutoACMG, a range of specialised **software tools**, **programming resources**, and **external services** were utilised. The project primarily employed Python 3.12, with version control managed through GitHub and local repositories handled using git. Continuous integration and deployment (CI/CD) were streamlined using GitHub Actions, automating workflows for improved software testing and deployment. Documentation was built using Sphinx and hosted on "Read the Docs", providing a comprehensive interface for both end-users and developers.

The command-line interface (CLI) for AutoACMG was crafted using Typer for ease of use, and Loguru for enhanced logging. Pydantic handled data validation from external application programming interfaces (APIs), ensuring data processing correctness. The software's code quality was maintained using MyPy for static type checking, Isort for import sorting, Flake8 for coding style enforcement, and Black for code formatting. Testing was conducted using Pytest to manage and execute comprehensive test suites, essential for a tool like AutoACMG.

External data and services integration was achieved through **REEV microservices** — annonars, mehari, and dotty. The annonars microservice pulled gene-specific and variant-specific data from databases such as gnomAD, dbSNP, and dbNSFP, including computational scores. Mehari provided transcript-specific information like HGVS notations and VEP consequences, identifying if a transcript is MANE selected. The dotty microservice resolved sequence variants from HGVS notations based on the latest genomic assemblies.

For domain-specific functionalities, PyTabix was utilized for searching UniProt and RMSK conservation domains, while SeqRepo was used for efficient data retrieval from RefSeq GRCh37 and GRCh38 human genomes. MaxEntpy facilitated splicing predictions using RefSeq sequences retrieved via SeqRepo.

2.5.3 AutoACMG Implementation

Software Architecture

The AutoACMG tool is engineered to offer dual functionalities through a Command Line Interface (CLI) and an Application Programming Interface (API), accommodating various user needs. The tool's architecture is designed to facilitate the analysis of genetic variants by implementing the ACMG criteria through structured computational steps.

At its core, AutoACMG operates through a series of critical steps. It begins by **resolving the variant** using a combination of regular expressions matches for canonical representations and the "dotty" microservice for variants expressed in HGVS (Human Genome Variation Society) and rsID notations, which refer to reference SNP identifiers in databases. Once the variant is resolved, AutoACMG methodically **processes each ACMG criterion**. During this prediction phase, necessary data is retrieved and algorithmic evaluations are conducted to classify the variant according to ACMG guidelines. Figure 4 graphically presents all the core functionalities of AutoACMG in the internal infrastructure diagram.

Command Line Interface

The CLI component of AutoACMG provides the **"classify"** command, which accepts a variant name as a required positional argument and an optional genome release version, defaulting to GRCh38 if unspecified. This interface logs detailed steps of the prediction process and returns the prediction result. This output includes details for each of the 28 ACMG criteria, with properties such as name, prediction, summary, and description. The name specifies the criterion, while prediction indicates its status: **"met"** (criteria triggered), **"unmet"** (criteria not triggered), **"deprecated"** (following ClinGen guidelines for PP5 and BP6), **"not applicable"** (PS1 & PM5 for missense variants only), or **"not set"** (prediction failure). An example of this prediction result can be found in Appendix under the name "Figure A1. Example output".

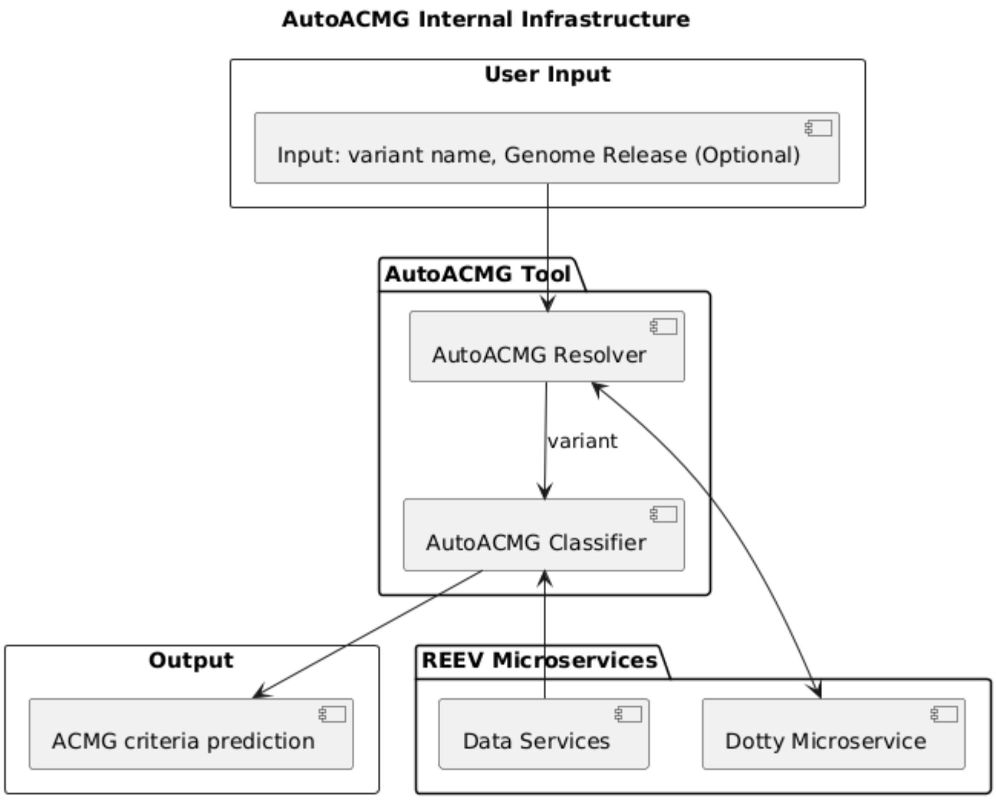

Figure 4. AutoACMG Internal Infrastructure. The diagram shows the workflow of the AutoACMG tool, where user inputs are processed by the AutoACMG Resolver and Dotty Microservice to resolve sequence variants, which are then classified by the AutoACMG Classifier using data from REEV Microservices, resulting in ACMG criteria prediction outputs.

Application Programming Interface

At the core of the API is the **"AutoACMG"** class, which can be initialised with parameters for the variant name and genome release. This class includes methods such a **"resolve_variant"**, which processes and returns the sequence variant in its canonical form, an **"predict"**, which performs the classification of ACMG criteria and returns the prediction result.

PVS1 Criterion Implementation

The AutoACMG tool employs a decision tree (Figure A2 in Appendix), aligned with the 2018 guidelines for PVS1 criterion evaluation, focusing on the evidence strength for variant classification. This section details the implementation of key decision blocks within the tool.

The **"undergo_nmd"** function assesses whether variants undergo nonsense-mediated decay (NMD), a key factor for evaluating PVS1 in nonsense or frameshift variants. The function evaluates if the variant is located in the last exon or within the last 50 nucleotides of the penultimate exon; variants outside these regions are predicted to undergo NMD, while those within are not. Details are available in Pseudocode 1.

The **"in_bio_relevant_tx"** function determines if a variant is in a biologically relevant transcript by checking for the **"ManeSelect"** tag, which signifies the transcript as a major isoform according to the MANE project. This function ensures that the variant analysis focuses on clinically relevant transcripts in the AutoACMG tool.

Pseudocode 1. Determining Nonsense-Mediated Decay (NMD) Status of Variants

```
Input: variant_position, gene_name, genomic_strand, exons
Output: True if the variant undergoes NMD, False otherwise
Begin
    If gene_name is "GJB2" Then
        Return True
    Calculate exon_lengths as (exon_end_position - exon_start_position + 1)
for all exons
    If single exon Then
        Return True
    Set nmd_cutoff as sum of exon_lengths minus the last exon and minimum
from second last exon and 50
    If variant_position <= nmd_cutoff Then
        Return True
    Else
        Return False
End
```

The **"crit4prot_func"** function evaluates whether a truncated or altered region is critical for protein function by examining the presence of pathogenic variants downstream of the new stop codon. It calculates the affected region based on the variant's position within the gene's coding sequence and counts pathogenic variants within this region, using data from **ClinVar**. If more than 5% of the variants in this region are pathogenic, the region is considered critical. Further details are available in Pseudocode 2.

Pseudocode 2. Assessing Impact on Protein Function

```
Input: variant, exons, genomic_strand
Output: True if the altered region is critical for the protein function,
otherwise False
Begin
    Calculate start_pos and end_pos of the affected exon region based on
variant, exons and strand
    Fetch pathogenic_variants and total_variants in the range start_pos to
end_pos
    If total_variants is 0 Then
        Return False  // No variant found
    Calculate frequency of pathogenic_variants as pathogenic_variants /
total_variants
    If frequency of pathogenic_variants > 0.05 Then
        Return True
    Else
        Return False
End
```

The **"lof_freq_in_pop"** function determines if the frequency of Loss-of-Function (LoF) variants in a gene's exon is common in the general population, by assessing their prevalence within the exon's genomic range. It computes the location of the exon based on the variant's position and counts both total LoF variants and those considered frequent. A LoF variant is classified as frequent if more than **10%** of identified LoF variants within the region are common in the population. This assessment influences the application of the PVS1 criterion within the ACMG guidelines, as variants with high LoF frequencies may be interpreted as benign. Details are available in Pseudocode 3.

Pseudocode 3. Assessing Frequency of LoF Variants in the General Population

```
Input: variant, exons, strand
Output: True if the LoF variant frequency is greater than 0.1%, False
otherwise
Begin
    Calculate start_pos and end_pos of the affected exon region based on
variant, exons and strand
    Fetch frequent_lof_variants and total_lof_variants in the range start_pos
to end_pos
    If total_lof_variants is 0 Then
        Return False
    Calculate frequency of frequent_lof_variants as frequent_lof_variants /
total_lof_variants
    If frequency of frequent_lof_variants > 0.1 Then
        Return True
    Else
        Return False
End
```

The **"lof_rm_gt_10pct_of_prot"** function evaluates whether a Loss-of-Function (LoF) variant eliminates more than **10%** of a protein, using the variant's position and the total protein length. It directly calculates the proportion of the protein affected by the variant and returns if more than 10% of the protein is removed, aligning with the criterion that significant deletions in protein structure are likely to impact function. Details are available in Pseudocode 4.

Pseudocode 4. Evaluating if the LoF Variant Removes More Than 10% of Protein

```
Input: prot_pos, prot_length
Output: True if the LoF variant removes more than 10% of the protein, False
otherwise
Begin
    Calculate percentage_removed as prot_pos / prot_length
    If percentage_removed > 0.1 Then
        Return True
    Else
        Return False
End
```

The **"exon_skip_or_cryptic_ss_disrupt"** function evaluates whether a genetic variant causes exon skipping or disrupts cryptic splice sites. For exon skipping, the function checks if the exon's length where the variant is located is divisible by three; if not, it predicts exon skipping. In the case of splice variants, it identifies potential cryptic splice sites by extracting the sequence around the variant and calculating **MaxEnt scores**. If a cryptic splice site has a high enough MaxEnt score to be considered significant and its distance from the variant isn't divisible by three, the function predicts disruption of the splice site. This detailed assessment helps ascertain if a variant can alter the normal splicing process, potentially leading to pathogenic outcomes. Details are available in Pseudocode 5.

The **"alt_start_cdn"** function determines whether a sequence variant results in an alternative start codon in any transcript besides the primary one. This is done by comparing the start positions of coding sequences across various transcripts. If any transcript presents a start position different from the primary transcript's start codon, the function concludes that an alternative start codon has been introduced. Details are available in Pseudocode 6.

The **"up_pathogenic_vars"** function evaluates if there are pathogenic variants upstream from the nearest potential in-frame start codon. It first identifies this start codon, then fetches and counts any pathogenic variants between this codon and the variant's position. If pathogenic variants exist within this specified range, the function returns "true", indicating the presence of **upstream pathogenic variants**. This process involves checking the sequence orientation and calculating the range based on the first and last exons, along with the closest alternative start codon, to determine the correct genomic span for searching pathogenic variants. Details are available in Pseudocode 7.

Pseudocode 5. Evaluating if Exon Skipping or Cryptic Splice Site Disruption Alters the Reading Frame

```
Input: variant, exons, consequences, strand
Output: True if the variant causes exon skipping or cryptic splice site
disruption, False otherwise
Begin
    Calculate start_pos and end_pos of the affected exon region based on
variant, exons and strand
    If (end_pos - start_pos) % 3 != 0 Then
        Return True  // Exon skipping predicted
    // Check for cryptic splice site disruption
    Find all cryptic_sites using MaxEnt splice site prediction
    For each cryptic_site in cryptic_sites Do
        If abs(cryptic_site.position - seqvar.pos) % 3 != 0 Then
            Return True  // Cryptic splice site disruption predicted
    Return False  // Preserves reading frame
End
```

Pseudocode 6. Evaluate possibility of alternative start codon

```
Input: transcripts_info, variant
Output: True if variant introduces an alternative start codon, False
otherwise
Begin
    Choose main_transcript based on variant  // Choose MANE transcript or
the longest one
    For each transcript in transcripts_info Do
            If alt_start_position of transcript != start_position of
main_transcript Then
            Return True
    Return False
End
```

Pseudocode 7. Evaluate importance of upstream region

```
Input: transcripts_info, variant, exons, genomic_strand
Output: True if pathogenic variants are found upstream of the closest
potential in-frame start codon, False otherwise
Begin
    Choose main_transcript based on variant  // Choose MANE transcript or
the longest one
    For each transcript in transcripts_info Do
            If alt_start_position of transcript != start_position of
main_transcript Then
            Set alt_start_pos as alt_start_position
            Break
    Determine end_pos of the exon where alt_start_codon is located using
variant, genomic_strand, and exons
    Count pathogenic_variants in the range from alt_start_pos to end_pos
    If pathogenic_variants > 0 Then
        Return True
    Return False
End
```

Other Criteria Implementation

Just as with the PVS1 criteria, the AutoACMG tool applies a systematic approach for predicting other ACMG criteria, involving accurate data parsing and algorithmic evaluation. This section details the algorithmic steps utilised for these predictions.

For the **PS1** and **PM5** criteria, the prediction process begins by retrieving the primary variant's details and extracting the primary amino acid change. The tool then iterates over all potential alternative bases for the variant, gathering information for these alternative variants. For each alternative, it parses the amino acid change and assesses whether this alternative variant has been previously determined as pathogenic. If the amino acid change in the alternative variant matches that of the primary variant and is deemed pathogenic, the PS1 criterion is met. Conversely, if the amino acid change differs but the variant is still pathogenic, the PM5 criterion is applied. The detailed implementation is provided in Pseudocode 8.

Pseudocode 8. PS1 and PM5 evaluation

```
Input: variant
Output: Prediction result for PS1 and PM5
Begin
    PS1, PM5 are set to False per default
    Parse Amino Acid substitution primary_aa_change from variant
    For each alt_base in {A, C, G, T} different from primary_aa_change Do
        If variant with alt_base is pathogenic Then
             Parse Amino Acid substitution alt_aa_change from variant_info
with alt_base
            If alt_aa_change == primary_aa_change Then
                Set PS1 to True
            If alt_aa_change != primary_aa_change Then
                Set PM5 to True
    Return PS1 and PM5
End
```

The **PM1** criterion in the AutoACMG tool is predicted through an analysis of pathogenic variants within specific genomic ranges related to the sequence variant. The process begins by counting pathogenic variants in the 50 base pair proximity to the variant. If **four or more** pathogenic variants are detected within this range, the PM1 criterion is met, suggesting a significant likelihood of pathogenicity due to the variant's location within a hotspot.

Additionally, the tool assesses whether the variant is situated within a UniProt domain, which is a critical region for protein function. If the variant lies within a UniProt domain, the tool then counts the pathogenic variants within this domain. Meeting the PM1 criterion in this context requires the presence of **at least two** pathogenic variants within the domain. The detailed implementation is provided in Pseudocode 9.

Furthermore, the **BS2** criterion considers the zygosity and penetrance of the variant; it is satisfied if the variant is observed in a healthy adult at a zygosity expected to cause disease if the variant were pathogenic. This evaluation involves checking the allele count and zygosity in genetic databases, ensuring the variant does not cause disease in a fully penetrant manner at an early age. The detailed implementation is provided in Pseudocode 10.

Pseudocode 9. PM1 evaluation

```
Input: variant, variant_info
Output: Prediction result for PM1
Begin
    Set PM1 to False by default
    If variant is in mitochondrial genome Then
        Return PM1 is False  // PM1 is not applicable
    Count pathogenic variants in the range (variant.position - 25) to (variant.position + 25)
    If pathogenic_count >= 4 Then
        Return PM1 is True
    Check if the variant is in a UniProt domain
    If not in a UniProt domain Then
        Return PM1 is False  // PM1 is not met
    Count pathogenic variants in the UniProt domain range
    If pathogenic_count >= 2 Then
        Return PM1 is True
    Else
        Return PM1 is False
End
```

Pseudocode 10. PM2, BA1, BS1 and BS2 evaluation

```
Input: variant
Output: Prediction result for PM2, BA1, BS1, BS2
Begin
    Initialize PM2, BA1, BS1, BS2 as False by default
    Retrieve allele frequency (af) for the variant
    If af is None Then
        Set PM2 to True  // Absent from controls
    Else If af > 0.05 Then
        Set BA1 to True  // Allele frequency > 5%
    Else If af >= 0.01 Then
        Set BS1 to True  // Allele frequency > 1%
    Else
        Set PM2 to True  // Allele frequency <= 1%
    Check zygosity and penetrance
    If af >= 0.01 and variant is observed in a healthy individual with relevant disorder Then
        Set BS2 to True
    Return PM2, BA1, BS1, BS2
End
```

The prediction of **PM4** and **BP3** criteria within the AutoACMG tool focuses on the type and location of in-frame deletions/insertions and stop-loss variants. For PM4, the tool identifies changes in protein length due to in-frame deletions or insertions that are not located within repeat regions or stop-loss variants, indicating significant alterations to protein structure that could affect function. If the variant causes a stop-loss, PM4 is automatically met, considering its potential to prolong the protein beyond its normal termination point, often resulting in functional disruption.

Conversely, **BP3** is triggered when in-frame deletions or insertions occur within repeat regions, suggesting that these alterations are less likely to impact protein function significantly due to their repetitive and potentially non-functional context. The tool utilises genomic libraries and the RepeatMasker track to ascertain the presence of a variant within these regions effectively. The detailed implementation is provided in Pseudocode 11.

Pseudocode 11. PM4 and BP3 evaluation

```
Input: variant
Output: Prediction result for PM4 and BP3
Begin
    Initialize PM4 and BP3 as False by default
    If variant is a stop-loss Then
        Return PM4 is True and BP3 is False
    If variant is in-frame deletion/insertion Then
        Check if the variant is in a repeat region
        If variant is not in a repeat region Then
            Set PM4 to True
        Else
            Set BP3 to True
    Return PM4 and BP3
End
```

The prediction of **PP2** and **BP1** criteria within the AutoACMG tool involves evaluating the frequency and type of missense variants within a specific gene segment. For PP2, a missense variant is considered supportive of pathogenicity if it occurs within a region where missense mutations are frequently pathogenic relative to benign ones. Conversely, BP1 is met when the frequency of benign missense variants significantly exceeds pathogenic variants, indicating that missense changes at this location are typically benign.

The process begins by fetching the gene transcript data to establish the range for variant assessment. Variants are then retrieved from this range, and each is evaluated for its missense consequence and clinical significance according to ClinVar records. The tool calculates the ratio of pathogenic to total missense variants and benign to total missense variants. If the pathogenic ratio exceeds a predefined threshold (**0.808**), PP2 is assigned, suggesting the region's susceptibility to harmful mutations. Similarly, if the benign ratio goes beyond a set threshold (**0.569**), BP1 is assigned, indicating that missense changes in this region are generally benign. The detailed implementation is provided in Pseudocode 12.

Pseudocode 12. PP2 and BP1 evaluation

```
Input: variant, transcript_info
Output: Prediction result for PP2 and BP1
Begin
    Initialize PP2 and BP1 as False by default
    Fetch transcript data for the variant
    If consequence is not missense Then
        Return PP2 is False and BP1 is False  // Only applicable for missense
variants
     Calculate start_pos and end_pos of the affected exon based on
transcript_info
     Count pathogenic_count, benign_count, and total_count of missense
variants in the range start_pos to end_pos
    Calculate pathogenic_ratio as pathogenic_count / total_count
    Calculate benign_ratio as benign_count / total_count
    If pathogenic_ratio > 0.808 Then
        Set PP2 to True
    If benign_ratio > 0.569 Then
        Set BP1 to True
    Return PP2 and BP1
End
```

For the prediction of **PP3** and **BP4** criteria within the AutoACMG tool, the approach relies on integrating computational predictions to assess variant pathogenicity or benign impact. PP3 is supported when computational tools predict a variant to be damaging, while BP4 is supported when computational predictions indicate benign impact.

The procedure begins by checking if computational tools like MetaRNN, BayesDel, and SpliceAI — chosen for their prior use in GeneBe — indicate a pathogenic prediction for the variant under consideration. If a computational tool assigns a score above a defined threshold, suggesting that the variant may affect the protein function or splicing, PP3 is met. Conversely, for BP4, the assessment checks if the tools classify the variant as benign, meaning the computational scores fall below a certain benign threshold, indicating that the variant is likely benign.

For **BP7**, which evaluates synonymous variants unlikely to affect splicing, the AutoACMG tool checks for pathogenic variants within a **2 base pairs** range. If such variants are found, BP7 is not met as it suggests a potential functional impact. Next, the tool assesses the proximity to canonical splice sites. Variants within 2bp of splice sites are excluded from BP7 due to the risk of splicing alterations. Lastly, the tool uses SpliceAI to predict splice site alterations. If SpliceAI indicates that the variant might create or disrupt splice sites, BP7 is not met. The detailed implementation is provided in Pseudocode 13.

Pseudocode 13. BP7 evaluation

```
Input: variant
Output: Prediction result for BP7
Begin
    Initialize BP7 as False by default
    If variant is in mitochondrial DNA Then
        Return BP7 is False
    Check for pathogenic variants within 2bp of the variant position
    If pathogenic variants are found Then
        Return BP7 is False
    Check if the variant is within 2bp of a splice site
    If the variant is within 2bp of a splice site Then
        Return BP7 is False
    Predict splice site alterations using SpliceAI
    If the variant is a splice site alteration Then
        Return BP7 is False
    Return BP7 is True
End
```

2.6 Comparison and Validation

2.6.1 Variant Selection

To validate AutoACMG and compare its performance with other automated classification tools, a custom dataset of variant classifications was compiled. This dataset comprises **168 variants**, selected from the AutoPVS1 supplementary materials and the ClinGen Evidence Repository [36]. The ClinGen Evidence Repository provides a curated collection of variants, each with detailed criteria assignments, reflecting the expert-driven classification activities of ClinGen. In contrast, the selection from AutoPVS1 was specifically focused on variants relevant to the PVS1 criterion, with additional criteria incorporated based on their mentions in the accompanying paper. The use of these two distinct sources enhances the dataset's value, allowing for a comprehensive evaluation of AutoACMG across a wide range of ACMG criteria, ensuring that the tool is rigorously tested on diverse and independently validated data.[1]

2.6.2 Comparison with Other Softwares

For the comparative analysis, **InterVar** and **GeneBe** were selected alongside **AutoACMG**, with VarSome excluded due to commercial restrictions. The primary focus was on evaluating the agreement between each tool and reference ClinGen classification using **Cohen's kappa**, similarly as highlighted in et al. (2024) [30]. This statistical coefficient measures the inter-annotator agreement for categorical items, such as ACMG criteria. Cohen's kappa was calculated for each ACMG criterion to determine the level of agreement between AutoACMG and each comparator tool. Standard interpretation thresholds for kappa values were used: values above 0.75 indicated excellent agreement, 0.60 to 0.74 indicated substantial agreement, 0.40 to 0.59 indicated moderate agreement, and 0.21 to 0.39 indicated fair agreement.

Additionally, the analysis included counting True Positives, False Positives, and False Negatives for each criterion across the three algorithms. Based on these counts, metrics such as **Precision**, **Recall**, and **F-score** were calculated to further compare the tools' performance. Precision measures the proportion of true positive results among all positive results predicted by the tool, reflecting the accuracy of the positive predictions. Recall indicates the proportion of true positive results among all actual positive cases, demonstrating the tool’s ability to identify positive cases. The F-score is the harmonic mean of Precision and Recall, providing a balanced measure that considers both false positives and false negatives. Detailed calculations are presented in Figure 5.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \quad F1\text{-}score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 5. Formulas for calculating Precision, Recall, and F1-score.

3 Results

3.1 Comparative Analysis of Algorithms

The agreement level between the results of each algorithm and the reference ClinGen assertions for individual ACMG criteria are presented with Cohen's kappa values in Table 3. This analysis considered only the presence of the criteria, without accounting for differences in reported strength levels.

AutoACMG showed excellent agreement for criteria such as **PVS1** (0.976) and **PM4** (1.0), reflecting its robust implementation for these well-defined criteria. However, for criteria like PM2 and PP2, the kappa values were lower (0.19 and -0.012, respectively), highlighting areas for further refinement. GeneBe consistently outperformed both AutoACMG and InterVar across most criteria, indicating the benefits of its more comprehensive approach by implementation of separate VCEP modifications.

[1] The complete list is available at the following link: https://github.com/bihealth/auto-acmg/blob/main/src/bench/comparison_criteria_custom.csv

Table 3. Cohen's Kappa Values for ACMG Criteria. This table presents Cohen's kappa values for each ACMG criterion, comparing the level of agreement between AutoACMG, InterVar, and GeneBe tools with the ClinGen reference assertions. The kappa values indicate the degree of concordance, categorised into different levels of agreement: excellent (≥0.75), substantial (0.60-0.74), moderate (0.40-0.59), and fair (0.21-0.39). The table also includes information on the number of variants and their respective genes used for the evaluation of each criterion, as well as the number of variants with specific VCEP modifications.

Criteria	Number of variants	VCEP modifications	Number of genes	GeneBe Kappa [%]	AutoACMG Kappa [%]	Intervar Kappa [%]
PVS1	24	9 (37,5%)	8	97,5	**97,6**	25,6
PS1	12	8 (66,7%)	7	**90,3**	69,8	0
PM1	48	48 (100%)	16	**94,2**	58,7	28,9
PM2	105	87 (82,9%)	21	**95,1**	19,4	37,2
PM4	3	0 (0%)	1	**100**	**100**	0
PM5	33	24 (72,7%)	14	**96,3**	81,7	9,4
PP2	48	40 (83,3%)	12	**97,1**	-1,2	-6,3
PP3	71	32 (45,1%)	23	**100**	79,7	62,4
BA1	37	37 (100%)	18	**94,9**	26,8	8,2
BS1	15	15 (100%)	7	**88,7**	31,5	-8,4
BS2	22	22 (100%)	9	**97,4**	38,9	8,7
BP1	15	1 (6,7%)	2	**100**	-1,2	32,3
BP3	3	0 (0%)	1	**100**	**100**	0
BP4	33	31 (93,9%)	16	**100**	41,3	60,8
BP7	22	17 (77,3%)	12	**100**	26,6	78,3

Table 4 presents the performance metrics — Precision, Recall, and F1-score — offering a detailed view of each tool's accuracy and reliability in variant classification. AutoACMG demonstrated high precision for **PVS1** (0.96), **PM4** (1.0), **BA1** (1.0), and **BP3** (1.0), indicating high specificity for these criteria. The precision for these criteria is attributed to their strict definitions, minimising the likelihood of false positive predictions when the necessary data is available. In contrast, InterVar's precision for **PM4** (N/A) and **BP3** (N/A) was significantly lower due to the lack of correct positive predictions. GeneBe performed near excellent for all the criteria.

For recall, AutoACMG showed high values for **PVS1** (1.0), **PP3** (0.901), **BP3** (1.0), and **BP7** (1.0), indicating its effectiveness in covering positive predictions for these criteria. However, the high recall for BP7 was achieved at the cost of a high number of False Positive predictions, likely due to the high number of variants with VCEP specifications in the testing dataset. The F1-score, which balances both precision and recall, provides a comprehensive measure of performance. AutoACMG achieved high F1-scores for PVS1 (0.99) and PM4 (1.0), indicating a well-rounded performance in these criteria. GeneBe's F1-scores were consistently high, particularly for PM2 (0.981) and PP2 (0.979), demonstrating its superior overall performance. InterVar, on the other hand, showed lower F1-scores across most criteria, reflecting its comparatively lower reliability in variant classification.

In summary, while AutoACMG excelled in several key criteria, particularly for **PVS1**, **PM4**, and **BP3**, GeneBe showed overall superior performance across a broader range of criteria. InterVar generally lagged behind in both precision and recall, emphasising the importance of

Table 4. Evaluation Metrics for ACMG Criteria. This table details the performance metrics for AutoACMG, InterVar, and GeneBe in predicting ACMG criteria. Metrics include Precision, Recall, and F1-score values for each criterion. The extended version of metrics can be found in Appendix under the name "Table A3. Evaluation Metrics for ACMG Criteria".

Criteria	**GeneBe**			**AutoACMG**			**InterVar**		
	Precision [%]	**Recall [%]**	**F1 [%]**	**Precision [%]**	**Recall [%]**	**F1 [%]**	**Precision [%]**	**Recall [%]**	**F1 [%]**
PVS1	**100**	95,8	97,9	96	**100**	**99**	**100**	16,7	28,6
PS1	**100**	**83,3**	**90,9**	69,2	75	72	N/A	0	N/A
PM1	**95,8**	**95,8**	**95,8**	62,5	83,3	71,4	40,5	93,8	56,6
PM2	**97,2**	**99**	**98,1**	84,6	31,4	45,8	73,9	80,9	73,2
PM4	**100**	**100**	**100**	**100**	**100**	**100**	N/A	0	N/A
PM5	94,3	**100**	**97,1**	87,1	82	84,3	**100**	6,1	11,4
PP2	**100**	**95,8**	**97,9**	0	0	N/A	17,6	6,3	9,2
PP3	**100**	**100**	**100**	86,4	90,1	88,3	69,8	94,4	80,2
BA1	94,8	**97,3**	**96**	**100**	18,9	31,8	**100**	5,4	10,3
BS1	**92,9**	**86,7**	**89,7**	26,5	86,7	40,6	0	0	N/A
BS2	**95,7**	**100**	**97,8**	37,2	72,7	49,2	21,1	18,2	20
BP1	**100**	**100**	**100**	2,4	6,7	3,3	26,3	100	41,7
BP3	**100**	**100**	**100**	**100**	**100**	**100**	N/A	0	N/A
BP4	**100**	**100**	**100**	55,2	48,5	51,6	85,7	54,5	66,7
BP7	**100**	**100**	**100**	26,2	**100**	41,5	85	77,3	81

updating the software and adapting it to newer guidelines, such as ClinGen and VCEP modifications.

4 Discussion and Outlook

4.1 Interpretation of Results

AutoACMG predictions generally showed higher accordance with the reference assertions than InterVar but did not reach the excellent results of GeneBe. This difference is particularly evident in criteria like **PM2** and **PP2**, where AutoACMG's kappa values were notably lower (0.19 and -0.012, respectively). The reduced performance in these criteria can be attributed to the absence of gene-specific thresholds defined by **VCEP modifications**, which were not implemented in AutoACMG. Similarly, the BA1, BS1 and PM2 criteria, which requires consideration of population frequency data, benefits significantly from VCEP's gene-specific cutoffs that were applied in GeneBe but not in AutoACMG.

Furthermore, in criteria such as **BP7**, AutoACMG showed a high recall (1.0) but at the expense of precision (0.262), leading to a lower F1-score. This outcome suggests that while AutoACMG effectively identifies all true positives, it also generates a higher rate of false positives, likely due to the **lack of splicing scores**, used for splicing alteration assessment. In contrast, AutoACMG excelled in criteria with strict definitions, such as PVS1 (with an F1-score of 0.99), where the robust implementation of null variant rules led to high accuracy.

In summary, while AutoACMG demonstrates strong performance in well-defined criteria, its overall accuracy is limited by the absence of **VCEP-specific guidelines**, as seen in the comparison with GeneBe, which implements these gene-specific modifications more effectively. This highlights the need for further refinement and integration of VCEP rules to enhance the precision and reliability of AutoACMG.

4.2 Technical Limitations and Challenges

The AutoACMG tool, while robust, faces several technical limitations and challenges that need to be addressed to enhance its predictive accuracy and reliability. One significant limitation is the quality and completeness of data obtained from external APIs. For example, the lower performance in criteria such as **BA1** (stand-alone, common variant), **BS1** (frequent in control population), and **PM2** (absent from controls) can be attributed to the high number of genes with special thresholds defined in VCEP rules. The reliance on default thresholds rather than gene-specific cutoffs, as specialised by experts in corresponding VCEPs, underscores the need for more context-aware algorithms.

Additionally, some criteria in AutoACMG rely on preset thresholds that could benefit from further tuning and refinement. For instance, the **PM1** criterion (variant in a hotspot region) and **PP2** criterion (missense variants in genes with a low rate of benign missense variation) showed lower concordance, indicating that the current thresholds might not be optimal. Fine-tuning these thresholds based on more extensive datasets and expert input could significantly improve the predictive performance of these criteria.

In summary, the main technical limitations and challenges faced by AutoACMG include the need for better data quality from APIs, the necessity of refining thresholds for certain criteria (implementation of VCEP rules), and the lack of detailed algorithmic descriptions for others.

Addressing these issues will be crucial for improving the tool's performance and reliability in clinical genetic variant analysis.

4.3 Future Work

4.3.1 Criteria Prediction Improvements

Future work will focus on implementing VCEP gene-specific rules to enhance the accuracy of AutoACMG's predictions. This will involve integrating these specialised guidelines to improve the reliability of the tool for a wider range of genetic variants. Secondly, refining thresholds for criteria such as PM1 and PP2 will be a priority, ensuring that the tool's predictions align more closely with expert assessments and current research findings. Lastly, efforts will be directed toward improving the implementation of the BP7 criterion by validating and incorporating more advanced splicing alteration detection tools. These enhancements will collectively ensure that AutoACMG continues to provide precise and reliable genetic variant classifications.

4.3.2 Technical Perspectives

Finally, integrating the AutoACMG package into the REEV software tool will streamline its usage and enhance its capabilities, offering users a comprehensive solution for genetic variant analysis. This integration will utilise REEV's robust infrastructure to facilitate more efficient data retrieval and variant classification, significantly improving upon the current use of InterVar.

Abbreviations

ES: Exome sequencing
GS: Genome sequencing
RNA: Ribonucleic acid
DNA: Deoxyribonucleic acid
ACMG: American College of Medical Genetics and Genomics
AMP: Association for Molecular Pathology
ClinGen: Clinical Genome Resource
UTR: Untranslated regions
mRNA: Messenger RNA
GRC: Genome Reference Consortium
SNV: Single nucleotide variant
indel: Insertion or deletion
CNV: Copy number variation
LoF: Loss-of-function
GoF: Gain-of-function
NMD: Nonsense-mediated decay
B: Benign
LB: Likely benign
US: Uncertain significance
LP: Likely pathogenic
P: Pathogenic
PVS: Pathogenic very strong
PS: Pathogenic strong
PM: Pathogenic moderate
PP: Pathogenic supporting

BA: Benign stand alone
BS: Benign strong
BP: Benign supporting
VCEP: Variant Curation Expert Panels
VEP: Variant effect predictor
MANE: Matched Annotation from NCBI and EMBL-EBI
CI/CD: Continuous integration and deployment
CLI: Command-line interface
API: Application programming Interface
TP: True positives
FP: False positives
FN: False negatives
N/A: Not applicable

References

[1] Lappalainen T, Scott A J, Brandt M and Hall I M 2019 Genomic Analysis in the Age of Human Genome Sequencing *Cell* **177** 70–84

[2] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody W W, Hegde M, Lyon E, Spector E, Voelkerding K and Rehm H L 2015 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.* **17** 405–24

[3] Pejaver V, Byrne A B, Feng B-J, Pagel K A, Mooney S D, Karchin R, O'Donnell-Luria A, Harrison S M, Tavtigian S V, Greenblatt M S, Biesecker L G, Radivojac P, Brenner S E, Biesecker L G, Harrison S M, Tayoun A A, Berg J S, Brenner S E, Cutting G R, Ellard S, Greenblatt M S, Kang P, Karbassi I, Karchin R, Mester J, O'Donnell-Luria A, Pesaran T, Plon S E, Rehm H L, Strande N T, Tavtigian S V and Topper S 2022 Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria *Am. J. Hum. Genet.* **109** 2163–77

[4] Li Q and Wang K 2017 InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines *Am. J. Hum. Genet.* **100** 267–80

[5] Xiang J, Peng J, Baxter S and Peng Z 2020 AutoPVS1: An automatic classification tool for PVS1 interpretation of null variants *Hum. Mutat.* **41** 1488–98

[6] Kopanos C, Tsiolkas V, Kouris A, Chapple C E, Albarca Aguilera M, Meyer R and Massouras A 2019 VarSome: the human genomic variant search engine *Bioinformatics* **35** 1978–80

[7] Abou Tayoun A N, Pesaran T, DiStefano M T, Oza A, Rehm H L, Biesecker L G, Harrison S M and Svi) C S V I W G (ClinGen 2018 Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion *Hum. Mutat.* **39** 1517–24

[8] Biesecker L G, Byrne A B, Harrison S M, Pesaran T, Schäffer A A, Shirts B H, Tavtigian S V and Rehm H L 2024 ClinGen guidance for use of the PP1/BS4 co-segregation and PP4 phenotype specificity criteria for sequence variant pathogenicity classification *Am. J. Hum. Genet.* **111** 24–38

[9] Ghosh R, Harrison S M, Rehm H L, Plon S E, Biesecker L G, and ClinGen Sequence Variant Interpretation Working Group 2018 Updated recommendation for the benign stand-alone ACMG/AMP criterion *Hum. Mutat.* **39** 1525–30

[10] Brnich S E, Abou Tayoun A N, Couch F J, Cutting G R, Greenblatt M S, Heinen C D, Kanavy D M, Luo X, McNulty S M, Starita L M, Tavtigian S V, Wright M W, Harrison S M, Biesecker L G, Berg J S, Abou Tayoun A N, Berg J S, Biesecker L G, Brenner S E, Cutting G R, Ellard S, Greenblatt M S, Harrison S M, Karbassi I, Karchin R, Mester J L, O'Donnell-Luria A, Pesaran T, Plon S E, Rehm H, Tavtigian S V, Topper S, and On behalf of the Clinical Genome Resource Sequence Variant Interpretation Working

Group 2019 Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework *Genome Med.* **12** 3
[11] Walker L C, Hoya M de la, Wiggins G A R, Lindy A, Vincent L M, Parsons M T, Canson D M, Bis-Brewer D, Cass A, Tchourbanov A, Zimmermann H, Byrne A B, Pesaran T, Karam R, Harrison S M, Spurdle A B, Biesecker L G, Harrison S M, Tayoun A A, Berg J S, Brenner S E, Cutting G R, Ellard S, Greenblatt M S, Kang P, Karbassi I, Karchin R, Mester J, O'Donnell-Luria A, Pesaran T, Plon S E, Rehm H L, Strande N T, Tavtigian S V and Topper S 2023 Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup *Am. J. Hum. Genet.* **110** 1046–67
[12] Hramyka D, Sczakiel H L, Zhao M X, Stolpe O, Nieminen M, Adam R, Danyel M, Einicke L, Hägerling R, Knaus A, Mundlos S, Schwartzmann S, Seelow D, Ehmke N, Mensah M A, Boschann F, Beule D and Holtgrewe M 2024 REEV: review, evaluate and explain variants *Nucleic Acids Res.* gkae366
[13] Pang A W, MacDonald J R, Pinto D, Wei J, Rafiq M A, Conrad D F, Park H, Hurles M E, Lee C, Venter J C, Kirkness E F, Levy S, Feuk L and Scherer S W 2010 Towards a comprehensive structural variation map of an individual human genome *Genome Biol.* **11** R52
[14] Harrison P W, Amode M R, Austine-Orimoloye O, Azov A G, Barba M, Barnes I, Becker A, Bennett R, Berry A, Bhai J, Bhurji S K, Boddu S, Branco Lins P R, Brooks L, Ramaraju S B, Campbell L I, Martinez M C, Charkhchi M, Chougule K, Cockburn A, Davidson C, De Silva N H, Dodiya K, Donaldson S, El Houdaigui B, Naboulsi T E, Fatima R, Giron C G, Genez T, Grigoriadis D, Ghattaoraya G S, Martinez J G, Gurbich T A, Hardy M, Hollis Z, Hourlier T, Hunt T, Kay M, Kaykala V, Le T, Lemos D, Lodha D, Marques-Coelho D, Maslen G, Merino G A, Mirabueno L P, Mushtaq A, Hossain S N, Ogeh D N, Sakthivel M P, Parker A, Perry M, Piližota I, Poppleton D, Prosovetskaia I, Raj S, Pérez-Silva J G, Salam A I A, Saraf S, Saraiva-Agostinho N, Sheppard D, Sinha S, Sipos B, Sitnik V, Stark W, Steed E, Suner M-M, Surapaneni L, Sutinen K, Tricomi F F, Urbina-Gómez D, Veidenberg A, Walsh T A, Ware D, Wass E, Willhoft N L, Allen J, Alvarez-Jarreta J, Chakiachvili M, Flint B, Giorgetti S, Haggerty L, Ilsley G R, Keatley J, Loveland J E, Moore B, Mudge J M, Naamati G, Tate J, Trevanion S J, Winterbottom A, Frankish A, Hunt S E, Cunningham F, Dyer S, Finn R D, Martin F J and Yates A D 2024 Ensembl 2024 *Nucleic Acids Res.* **52** D891–9
[15] Melo J, Gordon D, Cross N and Goldman J 1993 The ABL-BCR fusion gene is expressed in chronic myeloid leukemia *Blood* **81** 158–65
[16] McMurray C T 2010 Mechanisms of trinucleotide repeat instability during human development *Nat. Rev. Genet.* **11** 786–99
[17] Ghimenti C, Sensi E, Iandolo D, Cipollini G, Ricci S, Conte P, Bevilacqua G and Caligo M 2000 Germline mutations of BRCA1-associated RING domain (BARD1) gene in breast and/or ovarian families negative for BRCA1 and 2 alterations *Breast Cancer Res.* **2** P1.15
[18] Riggs E R, Andersen E F, Cherry A M, Kantarci S, Kearney H, Patel A, Raca G, Ritter D I, South S T, Thorland E C, Pineda-Alvarez D, Aradhya S and Martin C L 2020 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.* **22** 245–57
[19] Chen J-M, Férec C and Cooper D N 2006 A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes I: general principles and overview *Hum. Genet.* **120** 1–21
[20] Preussner M, Gao Q, Morrison E, Herdt O, Finkernagel F, Schumann M, Krause E, Freund C, Chen W and Heyd F 2020 Splicing-accessible coding 3′UTRs control protein stability and interaction networks *Genome Biol.* **21** 186
[21] Kazazian H.H., Boehm C.D., Seltzer W.K. 2000 ACMG recommendations for

standards for interpretation of sequence variations *Genet. Med.* **2** 302–3
[22] Richards C S, Bale S, Bellissimo D B, Das S, Grody W W, Hegde M R, Lyon E and Ward B E 2008 ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007 *Genet. Med.* **10** 294–300
[23] Biesecker L G, Harrison S M, and ClinGen Sequence Variant Interpretation Working Group 2018 The ACMG/AMP reputable source criteria for the interpretation of sequence variants *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20** 1687–8
[24] Yeo G and Burge C B 2004 Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals *J. Comput. Biol.* **11** 377–94
[25] Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, Dörk T, Burge C and Gatti R A 2004 Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: Maximum entropy estimates of splice junction strengths *Hum. Mutat.* **23** 67–76
[26] Nassar L R, Barber G P, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez J N, Hinrichs A S, Lee B T, Lee C M, Muthuraman P, Nguy B, Pereira T, Nejad P, Perez G, Raney B J, Schmelter D, Speir M L, Wick B D, Zweig A S, Haussler D, Kuhn R M, Haeussler M and Kent W J 2023 The UCSC Genome Browser database: 2023 update *Nucleic Acids Res.* **51** D1188–95
[27] Anon Genoox. *Franklin by Genoox*. Accessed July 23, 2024. https://franklin.genoox.com.
[28] Cristofoli F, Daja M, Maltese P E, Guerri G, Tanzi B, Miotto R, Bonetti G, Miertus J, Chiurazzi P, Stuppia L, Gatta V, Cecchin S, Bertelli M and Marceddu G 2023 MAGI-ACMG: Algorithm for the Classification of Variants According to ACMG and ACGS Recommendations *Genes* **14** 1600
[29] Melidis D P, Landgraf C, Schmidt G, Schöner-Heinisch A, Hardenberg S von, Lesinski-Schiedat A, Nejdl W and Auber B 2022 GenOtoScope: Towards automating ACMG classification of variants associated with congenital hearing loss *PLOS Comput. Biol.* **18** e1009785
[30] Stawiński P and Płoski R GeneBe.net: Implementation and validation of an automatic ACMG variant pathogenicity criteria assignment *Clin. Genet.* **n/a**
[31] Elisabet Munté, Lidia Feliubadaló, Marta Pineda, Eva Tornero, Maribel Gonzalez, José Marcos Moreno-Cabrera, Carla Roca, Joan Bales Rubio, Laura Arnaldo, Gabriel Capellá, Jose Luis Mosquera, Conxi Lázaro, vaRHC: an R package for semi-automation of variant classification in hereditary cancer genes according to ACMG/AMP and gene-specific ClinGen guidelines, *Bioinformatics*, Volume 39, Issue 3, March 2023, btad128, https://doi.org/10.1093/bioinformatics/btad128
[32] Fan C, Wang Z, Sun Y, Sun J, Liu X, Kang L, Xu Y, Yang M, Dai W, Song L, Wei X, Xiang J, Huang H, Zhou M, Zeng F, Huang L, Xu Z and Peng Z 2021 AutoCNV: a semiautomatic CNV interpretation system based on the 2019 ACMG/ClinGen Technical Standards for CNVs *BMC Genomics* **22** 721
[33] Nicora G, Limongelli I, Gambelli P, Memmi M, Malovini A, Mazzanti A, Napolitano C, Priori S and Bellazzi R 2018 CardioVAI: An automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases *Hum. Mutat.* **39** 1835–46
[34] Peng J, Xiang J, Jin X, Meng J, Song N, Chen L, Tayoun A A and Peng Z 2021 VIP-HL: Semi-automated ACMG/AMP variant interpretation platform for genetic hearing loss *Hum. Mutat.* **42** 1567–75
[35] Karalidou V, Kalfakakou D, Papathanasiou A, Fostira F and Matsopoulos G K 2022 MARGINAL: An Automatic Classification of Variants in BRCA1 and BRCA2 Genes Using a Machine Learning Model *Biomolecules* **12** 1552
[36] Rehm H L, Berg J S, Brooks L D, Bustamante C D, Evans J P, Landrum M J, Ledbetter D H, Maglott D R, Martin C L, Nussbaum R L, Plon S E, Ramos E M, Sherry S T and Watson M S 2015 ClinGen — The Clinical Genome Resource *N. Engl. J. Med.* **372** 2235–42

Appendix

Figure A1. Example output of AutoACMG prediction. Original file can be found at: https://github.com/bihealth/auto-acmg/blob/main/src/bench/NM_000257.3(MYH7)%3Ac.3036C%3ET_output.csv


```
{
   "pvs1":{
      "name":"PVS1",
      "prediction":"<AutoACMGPrediction.NotSet":1>,
      "summary":"",
      "description":"PVS1 prediction failed."
   },
   "ps1":{
      "name":"PS1",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Error occurred during PS1/PM5 prediction. Error: No valid
primary variant information for PS1/PM5 prediction.",
      "description":""
   },
   "ps2":{
      "name":"PS2",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "ps3":{
      "name":"PS3",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "ps4":{
      "name":"PS4",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "pm1":{
      "name":"PM1",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Counting pathogenic variants in the range of 50bp.The
range is 23892794 - 23892844. =>
                Found 0 Pathogenic variants. =>
                Found less than 4 pathogenic variants.Checking if the
variant is in a UniProt domain. =>
                Check if the variant is in a UniProt domain.
                Counting pathogenic variants in the UniProt domain. The
range is 23892755 - 23892932. =>
                Found less than 2 pathogenic variants in the UniProt
domain. PM1 is not met.",
      "description":""
   },
   "pm2":{
      "name":"PM2",
      "prediction":"<AutoACMGPrediction.Positive":5>,
      "summary":"Check allele frequency for the control population.
      Allele frequency: 0.001873719971626997.Allele frequency <= 1%: PM2 is
met.Check zygosity.",
```

```
      "description":""
   },
   "pm3":{
      "name":"PM3",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "pm4":{
      "name":"PM4",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Check consequences of the variant for PM4.
                Variant consequence is not indel or stop-loss. PM4 and BP3
are not met.Check if the variant is in a repeat region.
                Variant is in a repeat region.",
      "description":""
   },
   "pm5":{
      "name":"PM5",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Error occurred during PS1/PM5 prediction. Error: No valid
primary variant information for PS1/PM5 prediction.",
      "description":""
   },
   "pm6":{
      "name":"PM6",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "pp1":{
      "name":"PP1",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "pp2":{
      "name":"PP2",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Fetching transcript data. =>
            Error occurred during PP2 and BP1 prediction. Error: Transcript
data is not fully set. Cannot initialize the PVS1 class.",
      "description":""
   },
   "pp3":{
      "name":"PP3",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Checking Scores. =>
            An error occurred during prediction. Error: Missing dbNSFP
data.",
      "description":""
   },
   "pp4":{
      "name":"PP4",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "pp5":{
      "name":"PP5",
      "prediction":"<AutoACMGPrediction.Depricated":4>,
```

```
      "summary":"",
      "description":""
   },
   "ba1":{
      "name":"BA1",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Check allele frequency for the control population.
            Allele frequency: 0.001873719971626997.
            Allele frequency <= 1%: PM2 is met.Check zygosity.",
      "description":""
   },
   "bs1":{
      "name":"BS1",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Check allele frequency for the control population.
            Allele frequency: 0.001873719971626997.
            Allele frequency <= 1%: PM2 is met.Check zygosity.",
      "description":""
   },
   "bs2":{
      "name":"BS2",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Check allele frequency for the control population.
            Allele frequency: 0.001873719971626997.
            Allele frequency <= 1%: PM2 is met.Check zygosity.",
      "description":""
   },
   "bs3":{
      "name":"BS3",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "bs4":{
      "name":"BS4",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "bp1":{
      "name":"BP1",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"",
      "description":""
   },
   "bp2":{
      "name":"BP2",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "bp3":{
      "name":"BP3",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Check consequences of the variant for PM4.
            Variant consequence is not indel or stop-loss. PM4 and BP3 are
not met.Check if the variant is in a repeat region.
            Variant is in a repeat region.",
      "description":""
   },
   "bp4":{
```

```
      "name":"BP4",
      "prediction":"<AutoACMGPrediction.Negative":6>,
      "summary":"Checking Scores. =>
            An error occurred during prediction. Error: Missing dbNSFP
data.",
      "description":""
   },
   "bp5":{
      "name":"BP5",
      "prediction":"<AutoACMGPrediction.NotAutomated":3>,
      "summary":"",
      "description":""
   },
   "bp6":{
      "name":"BP6",
      "prediction":"<AutoACMGPrediction.Depricated":4>,
      "summary":"",
      "description":""
   },
   "bp7":{
      "name":"BP7",
      "prediction":"<AutoACMGPrediction.Positive":5>,
      "summary":"Checking for pathogenic variants in the range of 2bp. =>
            No pathogenic variants found in the range of 2bp. =>
            Checking for proximity to splice site. =>
            Variant is not within 2bp of a splice site. =>
            Predicting splice site alterations using SpliceAI. =>
            Variant is not a splice site alteration. BP7 is met.",
      "description":""
   }
}
```

Figure A2. PVS1 Decision Tree. Outlines criteria for PVS1 classification including considerations for in silico splicing predictions and nonsense-mediated decay (NMD). Details factors like splice site proximity and functional importance of gene domains. [7]

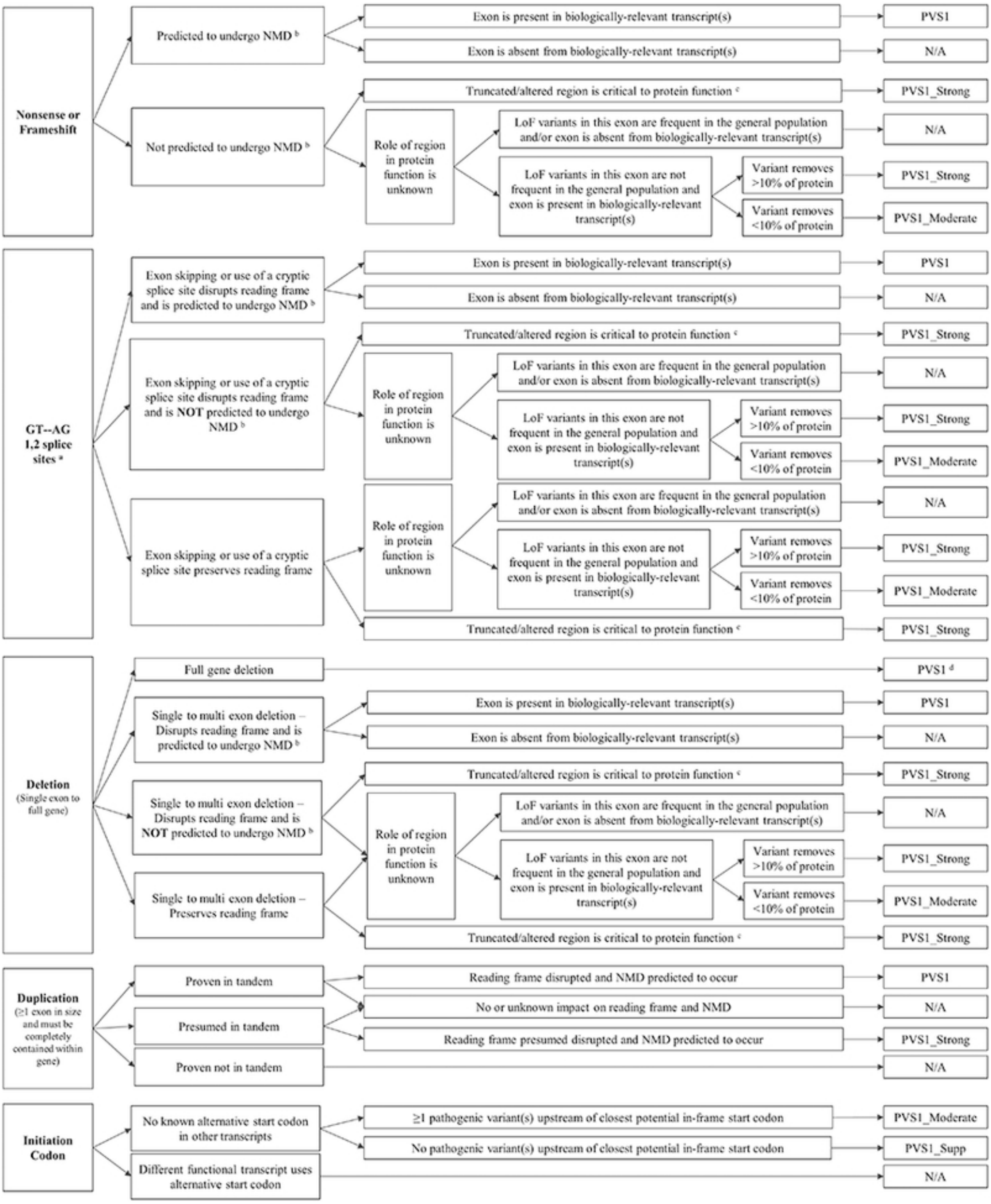

Table A3. Evaluation Metrics for ACMG Criteria. This table details the performance metrics for AutoACMG, InterVar, and GeneBe in predicting ACMG criteria. Metrics include True Positives (TP), False Positives (FP), and False Negatives (FN) for each criterion. Additionally, Precision, Recall, and F1-score values are provided, offering a comprehensive evaluation of each tool's accuracy and reliability in variant classification.

Criteria	AutoACMG						InterVar						GeneBe					
	TP	FN	FP	Precision	Recall	F1	TP	FN	FP	Precision	Recall	F1	TP	FN	FP	Precision	Recall	F1
PVS1	24	0	1	0.96	**1**	**0.99**	4	20	0	1	0.167	0.286	23	1	0	**1**	0.958	0.979
PS1	9	3	4	0.692	0.75	0.72	0	12	0	N/A	0	N/A	10	2	0	**1**	**0.833**	**0.909**
PM1	40	8	24	0.625	0.833	0.714	45	3	66	0.405	0.938	0.566	46	2	2	**0.958**	**0.958**	**0.958**
PM2	33	72	6	0.846	0.314	0.458	85	20	30	0.739	0.809	0.7732	104	1	3	**0.972**	**0.99**	**0.981**
PM4	3	0	0	**1**	**1**	**1**	0	3	0	N/A	0	N/A	3	0	0	**1**	**1**	**1**
PM5	27	6	4	0.871	0.82	0.843	2	31	0	**1**	0.061	0.114	33	0	2	0.943	**1**	**0.971**
PP2	0	48	1	0	0	N/A	3	45	14	0.176	0.063	0.092	46	2	0	**1**	**0.958**	**0.979**
PP3	64	7	10	0.864	0.901	0.883	67	4	29	0.698	0.944	0.802	71	0	0	**1**	**1**	**1**
BA1	7	30	0	**1**	0.189	0.318	2	35	0	**1**	0.054	0.103	36	1	2	0.948	**0.973**	**0.96**
BS1	13	2	36	0.265	0.867	0.406	0	15	12	0	0	N/A	13	2	1	**0.929**	**0.867**	**0.897**
BS2	16	6	27	0.372	0.727	0.492	4	18	15	0.211	0.182	0.2	22	0	1	**0.957**	**1**	**0.978**
BP1	1	14	50	0.02	0.067	0.03	15	0	42	0.263	1	0.417	15	0	0	**1**	**1**	**1**
BP3	3	0	0	**1**	**1**	**1**	0	3	0	N/A	0	N/A	3	0	0	**1**	**1**	**1**
BP4	16	17	13	0.552	0.485	0.516	18	15	3	0.857	0.545	0.667	33	0	0	**1**	**1**	**1**
BP7	22	0	62	0.262	**1**	0.415	17	5	3	0.85	0.773	0.81	22	0	0	**1**	**1**	**1**