

Variant Interpretation Using ACMG Guidelines

Submitted in partial fulfilment of the required for the award of
Master of Science Degree in Bioinformatics and Data Science

By

AKASH ANJALO (43739001)



**DEPARTMENT OF BIOINFORMATICS AND DATA
SCIENCE**

SCHOOL OF BIO AND CHEMICAL ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

**Accredited with Grade “A++” by NAAC JEPPIAAR
NAGAR, RAJIV GANDHISALAI, CHENNAI – 600119**

APRIL 2024



I

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with “A” grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Akash Anjalo (43739001) who carried out the project entitled “Variant interpretation using ACMG guidelines” under my / our supervision from January 2024 to April 2024

Internal Guide

External Guide (if Applicable)

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

II

DECLARATION

I, **AKASH ANJALO (43739001)**, hereby declare that the Project Report entitled “VARIANT INTERPRETATION USING ACMG GUIDELINES” done by me under the guidance of **Dr.ALEX ANAND** is submitted in partial fulfilment of the requirements for the award of Master of Science degree in Bioinformatics and Data Science.

DATE: 29/4/2024

AKASH ANJALO

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKRAVARTHI M.E., Ph.D. Dean of Arts and Science** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my project guide **Dr. ALEX ANAND** Department of Bioinformatics and Data Science for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Bioinformatics and Data Science** who were helpful in many ways for the completion of the project.

(AKASH ANJALO)

Table of Contents

Serial number	TITLE	Page number
1	Abstract	1
2	Introduction	2
3	Variants and Types	3
4	Variant Nomenclature tools	3
5	Integration Database	4
6	Population Database	5
7	Variant Database	5
8	Literature Search Tools	7
9	Computational Prediction	7
10	ACMG criteria • Aim • Learning objectives • Methods • Procedure • <u>Criteria for variants based on ACMG guidelines</u> • <u>Observation</u> • Interpretation • Significance and Application	9
11	References	17
12	Review Article	20
13	Certificates	25

Title: Variant Interpretation using ACMG Guidelines

Abstract:

This report provides a comprehensive overview of the interpretation of genetic variants according to the standards set forth by the American College of Medical Genetics and Genomics (ACMG). It delves into the diverse types of genetic variations, ranging from missense to null and silent variants, elucidating their significance in genotype-phenotype correlations. The importance of variant nomenclature standardization tools such as HGVS recommendations, VariantValidator, and Mutalyzer is emphasized for ensuring accurate variant descriptions.

Integrated databases such as Varcards, Mastermind, and wANNOVAR are explored as invaluable resources that consolidate variation data from multiple sources, including population databases like the Chinese Millionome Database and the Genome Aggregation Database. Furthermore, the roles of variant databases such as ClinVar, HGMD, and LOVD in annotating and classifying variations are elucidated.

Literature search tools like LitVar and PubMed are highlighted for their utility in gathering relevant information on genetic variations. Computational prediction tools such as MutationTaster, PolyPhen-2, MaxEntScan, and CADD are discussed for their role in assessing the pathogenicity and functional implications of variants.

The report also provides a detailed explanation of the ACMG criteria, outlining the criteria used for categorizing variations into pathogenic, likely pathogenic, benign, and other classifications. Specific criteria such as PM1 for mutational hotspots, PS1 for the same amino acid change, and PVS1 for null variants are examined in detail. Additionally, guidance is provided on reporting and evaluating sequencing findings in accordance with ACMG recommendations, ensuring standardized and clinically relevant interpretation of genetic variants.

Introduction

Next-generation sequencing (NGS) has greatly improved the diagnosis potential for uncommon genetic disorders, which has transformed clinical genetics. Patients usually show a wide range of sequence changes after whole-exome or whole-genome sequencing, from 200,000 to 400,000 or 3,900,000, respectively (Lionel et al., 2018; Yang et al., 2013). It can be difficult to discern between benign and disease-causing mutations because hundreds of clinically significant variants sometimes survive despite rigorous filtering procedures. Producing clinical genetic studies of the highest caliber requires precise interpretation of these variants (Song, Duzkale, & Shen, 2018).

A crucial step in disclosing the outcomes of genetic testing in clinical settings is the clinical interpretation of DNA sequence variations. The diagnosis of genetic abnormalities in ordinary clinical practice has been made much easier by the incorporation of next-generation sequencing technology into molecular genetic testing.

Clinical laboratories around the world have adopted guidelines from the American College of Medical Genetics and Genomics (ACMG/AMP) and the Association for Molecular Pathology to standardize the clinical interpretation of sequence variants. These standards are further refined by multiple disease-specific variant curation expert panels and by the ClinGen Sequence Variant Interpretation Working Group, which has supplied detailed instructions. Nevertheless, because professional opinion is involved in the interpretation process, the classification of variations may still differ throughout laboratories even with these efforts.

Furthermore, deciphering the clinical meaning of genetic test results and interpreting clinical genetic data can be quite difficult for researchers and clinicians who are not familiar with variant interpretation processes.

Variants and Types

Genetic variants are changes in DNA that can be benign, harmful, or have unknown effects. Null variants, caused by mutations like nonsense or frameshift mutations, significantly reduce protein function. Silent variants change the DNA but not the protein due to the redundancy of the genetic code. Intronic variants occur in non-coding regions and may affect gene expression. Missense variants are single nucleotide changes that alter one amino acid in the protein, potentially affecting its function. Nonsense variants introduce a premature stop codon, leading to a non-functional protein. In-frame insertions add nucleotides while maintaining the reading frame, with effects ranging from benign to significant. Deletions remove nucleotides and can affect protein function depending on the size and location of the deletion. Understanding these variants is important for studying health and disease.

Variant Nomenclature Tools

Variant nomenclature tools are software programs or online resources that assist scientists and medical professionals in accurately naming genetic variants according to accepted standards. This ensures clear communication and data exchange between researchers. Some commonly used tools include Mutalyzer, VariantValidator, and the Human Genome Variation Society (HGVS) recommendations. Mutalyzer is a web application that validates and corrects variant names based on HGVS standards. VariantValidator focuses on validating HGVS descriptions and converting between different variant formats. The HGVS itself provides guidelines for clear and consistent variant descriptions. These tools are crucial for ensuring accurate communication of genetic variants in research and medicine.

The Human Genome Variation Society (HGVS) offers a standardized system for describing genetic variants clearly and consistently. Here are the key points:

- **Three-Level Descriptions:** Variants are described at the DNA, RNA (if applicable), and protein level, providing a comprehensive picture of the change.
- **Reference Matters:** A specific reference sequence (genome assembly or transcript) is crucial for accurate variant location and definition.
- **Standardized Symbols:** Specific symbols and abbreviations ensure concise and consistent representation of different variation types (substitutions, deletions, etc.).

- **Pinpointing the Change:** Precise location information is provided, including genomic coordinates, transcript positions, and nucleotide changes within genes or transcripts. For coding regions, changes are described relative to the start codon.
- **Protein-Level Details:** For protein changes, the amino acid change, its position in the sequence, and the affected codon are specified using standard abbreviations.
- **Complexities Covered:** Guidelines exist for describing even intricate variants involving multiple changes, indels, or structural rearrangements.

Following these HGVS recommendations promotes clear communication, data exchange, and interpretation in genetic research and clinical settings.

Integrated Database:

Integrated databases are essential for researchers, clinicians, and genetic counselors because they bring together variant information from various sources. These databases, like VarCards and VarSome, offer a centralized platform to search for variants and access data on their functional effects, disease associations, and links to external resources. This is particularly helpful because variant information is typically spread across different databases, making it hard to get a complete picture. In simpler terms, these databases act like one-stop shops for variant exploration and interpretation.

wANNOVAR is a web-based tool that simplifies access to ANNOVAR's functionalities for annotating human genetic variants. Users can upload variant files and receive results through a user-friendly interface. wANNOVAR provides various annotations, including gene information, variant frequency, and disease associations. It also offers filtering options to focus on specific variant types (e.g., non-synonymous) or rare variants. Additionally, wANNOVAR prioritizes variants based on disease information when available, aiding researchers in identifying potential disease-causing variants.

Population Database

Population genetics databases like gnomAD, BRAVO, and CMDB are valuable resources for researchers studying human genetic variation. These databases offer information on the frequency and distribution of genetic variants in different populations.

- **gnomAD:** The largest public database of genetic variations, compiled from various sequencing projects. It allows researchers to analyze variant frequency, expression levels, and disease associations.
- **BRAVO:** A user-friendly database for exploring genetic variations and allele frequencies across diverse populations. It offers detailed information on variants and allows for comparisons between populations.
- **CMDB:** A database focused on genetic variation within Chinese populations. It provides data on SNPs, indels, and structural variants, along with population representation and disease associations.

These population genetics databases play a crucial role in population genetics research, disease association studies, and interpreting the clinical significance of genetic variants. They improve our understanding of genetic diversity, ancestry, and disease risk across populations.

VARIANT DATABASE

Variant databases are valuable resources for storing and sharing genetic variant information. They collect data from various sources like sequencing projects and clinical labs, encompassing different variant types and annotations. Here's a breakdown of their key functions:

1. **Data Aggregation:** These databases combine variant data from various sources, including whole-genome sequencing, exome sequencing, and targeted studies. They catalogue different variant types like SNPs, indels, and structural variants.
2. **Annotation and Classification:** Variant databases provide detailed annotations for each variant, including its location on the genome, how often it appears in different populations, its functional effects, and potential links to diseases or traits. They also classify variants based on established guidelines to aid interpretation and assess their clinical relevance.

3. **Population Diversity:** Many variant databases include data from diverse populations, allowing researchers to explore patterns of genetic variation and ancestry across these groups. Population-specific information on allele frequencies and genetic diversity is often provided to support population genetics research and disease association studies.
4. **Clinical Relevance:** Variant databases play a crucial role in clinical genetics by providing curated interpretations of a variant's clinical significance. They leverage data from trusted sources like ClinVar and integrate predictions from computational tools to aid variant interpretation and inform clinical decision-making.
5. **Data Accessibility and Sharing:** Variant databases prioritize data accessibility and sharing within the scientific community. They offer user-friendly interfaces for searching and downloading variant data and adhere to data-sharing policies and ethical guidelines to protect participant privacy and ensure responsible data use.

Overall, variant databases are instrumental for researchers, clinicians, and genetic counselors. They function as centralized repositories of genetic variant data, supporting various applications in genetics, genomics, and personalized medicine. Some commonly used variant databases are: ClinVar, HGMD, and LOVD:

- **ClinVar:** A public archive that stores reports of human variations classified for diseases and drug responses, along with supporting evidence.
- **HGMD:** A comprehensive collection of published disease-causing mutations in human genes.
- **LOVD:** A freely available web-based software for managing and displaying DNA variants in specific gene databases.

Literature search Tools :

Literature search tools are crucial for researchers and anyone interested in scholarly articles. These tools help find relevant information from a massive collection of academic publications.

- **Search Engines:** Sophisticated algorithms are used by these tools to organize, index, and retrieve scholarly content including articles, books, conference proceedings, and patents.
- **Tools used:**
 - PubMed (from the National Institutes of Health) focuses on biomedical and life sciences literature. It offers access to millions of citations and abstracts, with links to full text articles whenever available.
 - LitVar is a specialized tool that searches biomedical literature for information relevant to specific genetic variants. It employs advanced techniques to handle different representations of the same variant, ensuring comprehensive search results. LitVar offers features like autocomplete suggestions, search result ranking by relevance or date, and filtering options to refine your search.

Computational Prediction Tools:

Computational prediction tools are valuable assets for researchers analyzing genetic variants, especially in genomics and personalized medicine. These tools use computer algorithms to predict the potential effects, disease-causing nature (pathogenicity), and clinical significance of genetic variations.

Here are some commonly used examples:

- **MutationTaster:** This tool analyzes various features around a variant, including amino acid changes, splice site modifications, and evolutionary conservation. It then classifies the variant as likely disease-causing, benign, or uncertain. MutationTaster is freely available and integrates with other tools for a more comprehensive assessment.
- **PolyPhen-2:** This tool focuses on predicting how amino acid substitutions might affect protein structure and function. It analyzes sequence and structural features like conservation across species and protein domains to

assign a score reflecting the likelihood of the variant being harmful. PolyPhen-2 is a web-based tool and is integrated into several bioinformatics platforms.

- **MaxEntScan:** This tool predicts the strength of splice sites in DNA sequences. It uses a machine learning method to analyze short sequence motifs around potential splice sites and assigns a score based on their similarity to known splice sites. Higher scores indicate stronger splice sites. MaxEntScan is available as a web tool and is also integrated into bioinformatics software packages.
- **CADD:** This tool assigns a score to genetic variations indicating their potential harmfulness. It considers various data sources like conservation and integrates them into a single score. CADD scores correlate with disease-causing variants and can help prioritize functional variants for further investigation.

Aim :

To explain the various guidelines and criteria provided by ACMG by using CFTR F508 del variant as an example

Learning Objectives:

1. Understand the significance of the CFTR F508 del variant in the context of cystic fibrosis and its prevalence in affected populations.
2. Explore the guidelines and criteria established by the American College of Medical Genetics and Genomics (ACMG) for variant interpretation.
3. Examine the specific ACMG criteria relevant to the classification of the CFTR F508 del variant, including pathogenicity assertions and supporting evidence.
4. Learn how to apply ACMG criteria such as PS1 (same amino acid change), PM2 (absent from controls), and PP3 (functional studies support) to evaluate the pathogenicity of the CFTR F508 del variant.
5. Gain insights into the clinical implications of the CFTR F508 del variant, including its role in cystic fibrosis diagnosis, prognosis, and treatment decisions.
6. Understand the importance of standardized variant interpretation guidelines in clinical practice and research, using the CFTR F508 del variant as a case study.
7. Enhance proficiency in interpreting genetic variants by critically analyzing the ACMG guidelines in the context of the CFTR F508 del variant and similar genetic mutations.
8. Recognize the challenges associated with variant interpretation and classification, particularly in complex genetic diseases like cystic fibrosis, and strategies for addressing these challenges. Rephrase and summarize the content in points

Methods :

The American College of Medical Genetics and Genomics (ACMG) developed guidelines for classifying genetic variants identified through new sequencing technologies. Here's a summary of the process:

1. **Initial Survey and Feedback:** A survey was conducted to assess preferred terminology and gather feedback on variant assessment protocols from clinical labs. This feedback informed the development of a five-tier classification system ("pathogenic," "likely pathogenic," etc.) and initial criteria for assigning variants to these tiers.
2. **Testing and Refinement:** Workgroup members tested the classification scheme on known variants and typical examples. A second survey was distributed to a wider range of labs for feedback on the proposed system. This feedback helped refine the classification criteria and weighting of different types of evidence.
3. **Workshop and Audience Participation:** A workshop was held to present the revised criteria and explore two potential scoring systems (point-based vs. qualitative weighting). Audience participation helped determine how each criterion should be weighted during variant evaluation. The workgroup opted for a qualitative approach due to limitations in assigning point values to complex genetic evidence.
4. **Literature Review:** Recommendations from other professional societies and existing variant classification guidelines were reviewed. While these resources provided valuable insights, the ACMG approach aimed to be applicable across all Mendelian genes and various lab settings.
5. **Continuous Improvement:** The ACMG acknowledges the evolving nature of genetic technologies and anticipates future refinements to the classification system. They also encourage ongoing development of more focused guidelines for specific disease groups.

Procedure

Variant Interpretation with ACMG and the University of Maryland Tool:

- **Data Input:** Upload your genetic sequence data and identify the variant(s) using the tool.
- **Classification:** Utilize the tool's features to classify the variant into one of the five ACMG categories (pathogenic, likely pathogenic, etc.).
- **Criteria Evaluation:** Assess the variant against ACMG criteria (PVS1, PS1, PM2, etc.) using the tool's functionalities.
- **Variant Annotation:** Gather information like population frequency, conservation scores, and functional impact using the tool's annotation features.
- **Clinical Context:** Consider the patient's clinical picture, and family history, and potentially perform segregation analysis for further interpretation.
- **Expert Review & Reporting:** Review the tool's results with a team of experts and generate a comprehensive report with ACMG classification, evidence, and recommendations.
- **Database Submission:** Submit the interpretation results to public databases (like ClinVar) using the tool's functionalities.
- **Continuous Learning:** Stay updated on genetic research, variant interpretation guidelines, and the tool's updates for the most accurate patient care.

Criteria for variants based on ACMG guidelines

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: - Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>) - Use caution interpreting LOF variants at the extreme 3' end of a gene - Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact - Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G→C or G→T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case-control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in trans with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see Table 6) BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing BS4 Lack of segregation in affected members of a family Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease BP2 Observed in <i>trans</i> with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in <i>cis</i> with a pathogenic variant in any inheritance pattern BP3 In-frame deletions/insertions in a repetitive region without a known function BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant. BP5 Variant found in a case with an alternate molecular basis for disease BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Observation:

Before using the variant interpretation tool, the criteria given by ACMG must be studied. So that we could know that what are the categories in which the variants are classified. Once the criteria are known we could be able to interpret the variant as

- Pathogen
- Benign
- Variance of insignificance

Using the genetic variant interpretation tool provided by the University of Maryland, on using the variant interpretation tool provided by the University of Maryland we can come to the conclusion that the given variant CFTR F508del is a pathogenic variant that causes cystic fibrosis

Interpretation:

The variant CFTR F508del comes under various criteria like

- PS3
- PP3
- PP1
- PS4
- PM3
- BS1

This gives the result of Pathogen

Functional Assays (PS3): Functional studies have demonstrated the detrimental impact of the CFTR F508del variant on protein function. This variant results in defective CFTR protein processing and trafficking, leading to impaired chloride ion transport across cell membranes, a hallmark of cystic fibrosis pathophysiology. Functional assays provide direct evidence of the variant's deleterious effects, meeting the PS3 criterion for pathogenicity.

Clinical Data (PS4): Clinical data from extensive genotype-phenotype correlation studies corroborate the association between the CFTR F508del variant and severe cystic fibrosis phenotypes, including pancreatic insufficiency, respiratory complications, and reduced life expectancy. This clinical evidence supports the pathogenicity of the variant according to the PS4 criterion

Segregation Data (PP1): Segregation analysis in families with cystic fibrosis has consistently shown the CFTR F508del variant segregating with the disease phenotype in an autosomal recessive manner. This segregation pattern supports the pathogenic role of the variant, fulfilling the PP1 criterion.

Based on these criteria the variant interpretation tools gives the result of the variant interpretation as Pathogenic

Program for Personalized and Genomic Medicine

[Programs](#) > [Program for Personalized and Genomic Medicine](#) > [Implementation Projects](#) > [Genetic Variant Interpretation Tool](#)

[Home](#)

[Leadership](#) >

[Administration](#)

[Program Members](#)

[Graduate](#)

[Programs](#)

[Shared Resources](#)

[Membership](#)

[Implementation Projects](#) >

[Research](#) >

[In the News](#) >

[Seminars](#)

[For Healthcare Providers](#) >

[Contact Us](#)

Genetic Variant Interpretation Tool

To aid our variant interpretation process, we created an openly-available online tool to efficiently classify variants based on the evidence categories outlined in the article:

[Richards, et al. *Standards and guidelines for the interpretation of sequence variants*. 2015.](#) This site displays the evidence categories and descriptions from Table 3 and Table 4 with simple checkboxes for selecting appropriate criteria. The site then incorporates the algorithm in Table 5 to automatically assign the pathogenicity or benign impact based on the selected evidence categories. Since our process often requires analyzing multiple variants per patient, we have also allowed the option of aggregating each variant into an exportable table at the foot of the website for easy documentation of the variant review process for our records. Although this tool is based on the ACMG/AMP Standards and Guidelines, it is not affiliated with ACMG, AMP, or any of the authors of the publication.

[Click here to group evidence by category](#)

Patient ID:
Variant ID:

☐ PVS1 null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease

☐ PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change
☐ PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history
☐ PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product
☐ PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls
☐ PP1 (Strong evidence) Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease

Please note that the text of the variant evidence has been pulled directly from Richards, et al. *Genet Med*. 2015 May;17(5). This site does not claim authorship of any of the variant evidence descriptions.

This tool is based on the published

Effect	Evidence	To delete
Pathogenic (Ia)	PS3, PS4, PVS1, PM3, PM4	Delete
Pathogenic (II)	PS1, PS3, PS4, PM3, BS1	Delete
VUS - conflicting evidence	PS1, PS3, PS4, PM3, BS1, BA1	Delete
Benign (I)	BA1, PVS1	Delete
VUS - not enough evidence	PVS1, BS1	Delete

Significance and Application

Importance of Variant Interpretation using ACMG Guidelines

ACMG guidelines offer a standardized method for classifying genetic variations based on their impact on health. This classification system is crucial for various aspects of clinical genetics and genomics:

- **Accurate Diagnoses & Treatment:** Doctors can use ACMG classifications to confidently diagnose genetic diseases and tailor treatment plans based on the severity of the variant (pathogenic vs. likely benign).
- **Informed Genetic Counselling:** Genetic counsellors rely on ACMG classifications to accurately assess risks and provide informed advice to individuals and families with genetic conditions. This empowers them to make informed decisions about family planning, preventative measures, and medical monitoring.
- **Research & Discovery:** ACMG classifications help researchers prioritize variants for further investigation. By systematically classifying variants based on their clinical significance, researchers can gain a deeper understanding of the genetic basis of diseases and identify novel disease-causing variants.
- **Clinical Trials & Precision Medicine:** ACMG classifications are used to select patients for clinical trials and targeted therapies. By identifying individuals with pathogenic variants for specific diseases, clinicians can enroll them in relevant trials for personalized treatment options.
- **Quality & Standardization:** ACMG guidelines ensure consistency in interpreting variants across labs and clinical settings. This improves the reliability of genetic test results and enhances the quality of patient care and research.
- **Population Screening & Public Health:** ACMG classifications can be used for population screening programs to identify individuals at high risk for genetic diseases. This allows public health initiatives to implement targeted interventions, like genetic counselling and preventative measures, to reduce disease burden and improve overall population health.

In conclusion, ACMG guidelines play a critical role in translating complex genetic data into actionable information for clinical practice and research.

Advances in systems biology modelling

Done by

Akash Anjalo



Overview

To gain a system-level understanding of biological processes, computational and mathematical models are required; but, before they can be properly used and evaluated, their limitations must be made clear. An objective evaluation of techniques is made possible by crowdsourced standards in the form of challenges. Over the past ten years, the Dialogue for Reverse Engineering Assessment and Techniques (DREAM) has arranged more than fifteen systems biology challenges. The breadth of application includes transcription factor binding, medication combinations, whole-cell models, cell-lineage reconstruction, signalling networks, transcription regulation, transcription factor binding, single-cell placement in a tissue, and cell survival.

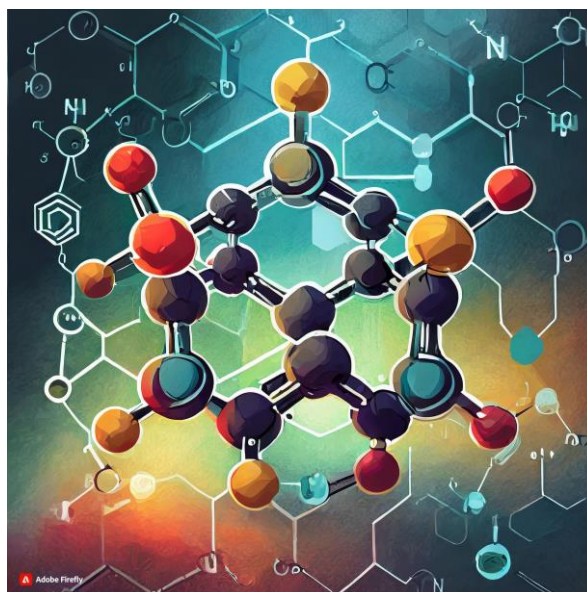
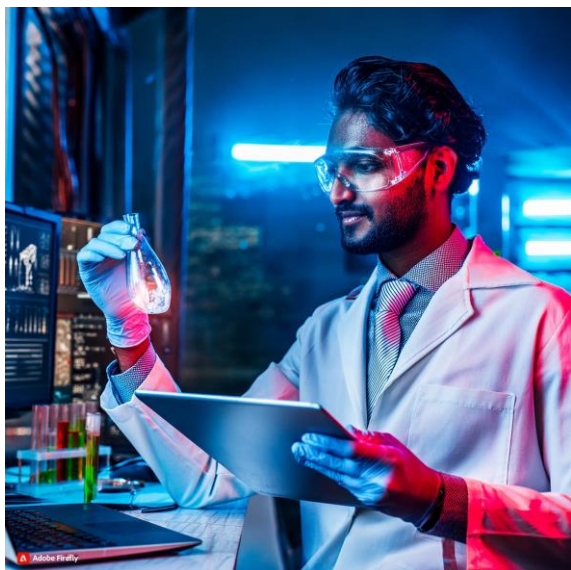
Precision medicine for cancer is fuelled by multi-omics analysis powered by artificial intelligence.



As biotechnology advances, new omics technologies emerge that provide researchers with multi-layer data from many sources, including the genome, epigenome, transcriptome, metabolome, and proteome. Many omics technologies, including bulk and single-cell omics techniques, have allowed for unprecedented size and resolution characterization of many molecular layers, resulting in a comprehensive understanding of cancer activity. Multi-omics analysis provides a systematic way to probe various molecular information at every biological layer, even if it also presents difficult challenges to solve to extract relevant insights from the quickly growing amount of multi-omics data. The improvements made feasible by artificial intelligence-based multi-omics approaches are highlighted, including response evaluation and prognosis prediction. Finally, we talk about the difficulties multi-omics analysis faces and potential future paths for the field's advancement.

Thus, efficient algorithms are required to reduce the complexity of the data while still analyzing the mysteries underlying the intricate biological processes causing cancer. Artificial intelligence has proven to be capable of complementing multi-modal data streams in the field of cancer. The simultaneous development of multi-omics technologies and artificial intelligence algorithms has allowed for speedier advancements in cancer precision medicine. The advances enabled by artificial intelligence-based multi-omics approaches are highlighted, with a focus on early cancer screening, diagnosis, response evaluation, and prognostic prediction. Finally, we talk about the difficulties multi-omics analysis faces and potential future paths for the field's advancement. Given the growing use of artificial intelligence (AI) in multi-omics analysis, we predict a paradigm change in precision medicine driven by AI-based multi-

Systems biology's use and advancement in computer-aided drug design



Thanks to scientific developments, the complex pathophysiology of diseases is now much well understood. Pharmacological design's main objective is now to reveal a drug's whole method of action and therapeutic effect. However, the traditional methods of creating medications are not up to the expectations of today. Systems biology has advanced significantly in recent years, which has paved the way for the use of several cutting-edge technologies in drug research and discovery, including proteomics, genomics, and metabolomics. Computer-aided drug design, or CADD, functions as a bridge between traditional pharmaceutical theory and modern science, thereby shortening the duration of the drug development cycle and increasing the success rate of drug design.

To clarify the general mechanism and action of pharmaceuticals, a direction and methodological foundation are provided by the application of CADD and systems biology. This paper offers resources to support the application, recommends a roadmap for future expansion, and shows the systems biology research and application in CADD from multiple perspectives.

Systems biology's challenges and future directions

Hurdles: Building precise models, integrating data from several sources, and organizing and analysing massive volumes of data are just a few of the difficulties associated with computational biology. To store, process, and analyse data, managing massive datasets calls for specialized infrastructure like cloud computing and high-performance computing. A thorough comprehension of biological processes is necessary for realistic models, in addition to the use of sophisticated mathematical and statistical techniques. The creation of standardized data formats and data integration processes is necessary for the integration of data from various sources.

Administration and synchronization: An enormous volume of biological data is being produced as a result of the development of high-throughput technologies. Maintaining and combining these many forms of data from various sources is a difficult task

Data integrity and consistency: The quality of the data that is supplied determines the accuracy and repeatability of the computational analysis. Therefore, it is crucial to establish quality control processes as well as standards for data sharing and replication.

Developing effective strategies: Analysing huge data sets is a major problem in computational biology. These algorithms need to be accurate, scalable, and flexible enough to work with a variety of data kinds.

Improvement and thorough validation of the simulation are necessary for accurate computer simulations of biological systems. Although it might be time-consuming and repetitive, this technique is necessary to guarantee the models' correctness and dependability.

Ethical considerations: Computational biology must assess the ethical consequences of its study, just like any other area that uses sensitive information. Concerns about data security, privacy, and informed consent fall under this category.

Future scope of systems biology: It is obvious that principles will be tested in both natural populations and laboratory settings employing molecular biology techniques. The structure of the biological sciences will become more like that of physics, with researchers and theorists working together to solve problems.

Conclusion



The study of systems of organisms nowadays requires computational biology. It offers an effective approach for building detailed models of biological systems as well as for analyzing and interpreting big, complicated datasets. Computational biology still has to overcome other obstacles, such as assembling and analyzing massive data sets, creating precise models, and merging data from many sources. It will take cooperation between mathematicians, computer scientists, and biologists to address these issues and continue creating state-of-the-art computational instruments.

REFERENCES

Zhang J, Yao Y, He H, Shen J. Clinical Interpretation of Sequence Variants. *Curr Protoc Hum Genet*. 2020 Jun;106(1):e98. doi: 10.1002/cphg.98. PMID: 32176464; PMCID: PMC7431429.

Variant nomenclature verification tools:

Mutalyze -Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat*. 2008 Jan;29(1):6-13. doi: 10.1002/humu.20654. PMID: 18000842.

VariantValidator -Freeman PJ, Hart RK, Gretton LJ, Brookes AJ, Dalglish R. VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Hum Mutat*. 2018 Jan;39(1):61-68. doi: 10.1002/humu.23348. Epub 2017 Oct 17. PMID: 28967166; PMCID: PMC5765404.

Human Genome Variation Society (HGVS) recommendations for sequence variant nomenclature-den Dunnen JT, Dalglish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PE. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*. 2016 Jun;37(6):564-9. doi: 10.1002/humu.22981. Epub 2016 Mar 25. PMID: 26931183.

Integrated databases:

VarSome - Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. *Bioinformatics*. 2019 Jun 1;35(11):1978-1980. doi: 10.1093/bioinformatics/bty897. PMID: 30376034; PMCID: PMC6546127.

VarCards- Li J, Shi L, Zhang K, Zhang Y, Hu S, Zhao T, Teng H, Li X, Jiang Y, Ji L, Sun Z. VarCards: an integrated genetic and clinical database for coding variants in the human genome. *Nucleic Acids Res*. 2018 Jan 4;46(D1):D1039-D1048. doi: 10.1093/nar/gkx1039. PMID: 29112736; PMCID: PMC5753295.

wANNOVAR-Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015 Oct;10(10):1556-66. doi: 10.1038/nprot.2015.105. Epub 2015 Sep 17. PMID: 26379229; PMCID: PMC4718734.

Variant databases:

ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>)

Human Gene Mutation Database (HGMD) (license required for professional version) (<http://www.hgmd.org/>),

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, ... Cooper DN (2017). The Human Gene Mutation Database: towards a comprehensive repository of

inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*, 136(6), 665–677. doi: 10.1007/s00439-017-1779-6

Leiden Open Variation Database (LOVD) (<https://www.lovd.nl/>)

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, & den Dunnen JT (2011). LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*, 32(5), 557–563. doi: 10.1002/humu.21438

Literature search tools:

PubMed (<https://pubmed.ncbi.nlm.nih.gov/>)

LitVar (<https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/LitVar/>)

Computational prediction tools:

MutationTaster (<http://mutationtaster.org/>)

Schwarz JM, Cooper DN, Schuelke M, & Seelow D (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*, 11(4), 361–362. doi: 10.1038/nmeth.2890

PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>)

Adzhubei I, Jordan DM, & Sunyaev SR (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet, Chapter 7, Unit7* 20. doi: 10.1002/0471142905.hg0720s76

MaxEntScan (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html)

Yeo G, & Burge CB (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, 11(2–3), 377–394. doi: 10.1089/1066527041410418

ACMG criteria

[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 - Genetics in Medicine \(gimjournal.org\)](http://gimjournal.org)

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PMID: 25741868; PMCID: PMC4544753.

MATLAB Certificates:



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

MATLAB Onramp

Craig L. Santos
DIRECTOR, TRAINING SERVICES

12 November 2023



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

MATLAB Programming Techniques

Craig L. Santos
DIRECTOR, TRAINING SERVICES

31 March 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Simulink Onramp


DIRECTOR, TRAINING SERVICES

26 March 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Reinforcement Learning Onramp


DIRECTOR, TRAINING SERVICES

14 December 2023



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

MATLAB for Data Processing and Visualization

A handwritten signature in black ink, reading "Craig L. Santos".

DIRECTOR, TRAINING SERVICES

20 December 2023



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Statistics Onramp

A handwritten signature in black ink, reading "Craig L. Santos".

DIRECTOR, TRAINING SERVICES

9 January 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Machine Learning Onramp



DIRECTOR, TRAINING SERVICES

15 November 2023



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Machine Learning with MATLAB



DIRECTOR, TRAINING SERVICES

3 March 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Signal Processing Onramp


DIRECTOR, TRAINING SERVICES

5 March 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Computer Vision Onramp


DIRECTOR, TRAINING SERVICES

15 December 2023



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Image Processing with MATLAB


DIRECTOR, TRAINING SERVICES

31 March 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Solving Nonlinear Equations with MATLAB


DIRECTOR, TRAINING SERVICES

1 April 2024



Course Completion Certificate

Akash Anjalo

has successfully completed **100%** of the self-paced training course

Deep Learning Onramp

DIRECTOR, TRAINING SERVICES

3 December 2023

Bioinformatics Internship 3.0 by Bversity:

Bversity
Bioinformatics
INTERNSHIP 3.0

CERTIFICATE OF COMPLETION

This Certificate is Proudly Presented To

AKASH ANJALO

for successfully completing the 5-day Bioinformatics internship 3.0
organized from 15 Dec - 19 Dec, 2023. Wishing you all the very best for
your future endeavours!

SKILLING PARTNER

OFFICIAL TRAINING PARTNER

Sudh
SUDHARSAN V
CEO, BVERSITY

