Current Trends in Biomedical Engineering & Biosciences
ISSN: 2572-1151

Juniper Publishers
key to the Researchers

Research Article
Volume 24 Issue 3 - August 2026
DOI: 10.19080/CTBEB.2026.24.556139

Curr Trends Biomedical Eng & Biosci

A Deterministic Framework for Integrated Genome Variant Interpretation - The *'GenomeVAP'*

Anusha Sunder[1] and Govindharajan Mukund[2]

[1]*Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India.*

[2]*R&D Intern, Xcode Life Sciences, Pvt. Ltd. Chennai, India.*

Submission: August 21, 2026; **Published:** August 27, 2026

***Corresponding author:** Anusha Sunder, Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India.

Abstract

High-throughput genomic sequencing generates vast amounts of data, yet the interpretation of individual-genetic variants remains hindered by the dispersion of relevant evidence across various databases. We present a modular, web-based framework (GenomeVAP) designed for deterministic evidence integration in genomic research. Unlike machine learning models that often introduce noise into genomic annotations or rely on opaque predictive thresholds, GenomeVAP utilizes a weighted, rule-based scoring methodology to synthesize evidence from primary repositories, including ClinVar,[1] dbSNP,[2] Ensembl,[3] and the GWAS Catalog.[4] We evaluate the framework's efficacy through representative batch entries of clinically significant variants, demonstrating that centralized, automated retrieval reduces manual querying time while maintaining high transparency and reproducibility. GenomeVAP is intended exclusively as a bioinformatics software framework to aid interpretation in healthcare and academic research fields17.

Keywords: Genome, Variant Interpretation; Bioinformatics; Deterministic Scoring; Evidence Integration; Genomic Research; API / Application Programming Interface; Automated Report Generation

Introduction

The rapid advancement of high-throughput genomic sequencing technologies has outpaced the development of standard interpretation pipelines.[5] For genomic researchers, the identification of a variant is only the initial step in a complex workflow.[5] The subsequent characterization of these variants requires the synthesis of heterogeneous data from siloed biological repositories, such as ClinVar for clinical significance,[1] dbSNP for nomenclature and identifiers,[2] and the GWAS Catalog for trait associations.[4] This fragmentation creates a substantial "manual search bottleneck," where researchers must query multiple interfaces to compile a single comprehensive profile for a variant of interest.

This process is inherently time-consuming and prone to human error, particularly when managing large datasets. Current computational approaches for variant interpretation frequently rely on machine learning or black-box predictive models [6-8] (A black-box predictive model is an AI or machine learning system whose internal decision-making process is hidden from users, who can only see inputs and outputs). While these methods offer speed, they often obscure the underlying evidence, making it difficult for researchers to audit the provenance of the final annotation. Furthermore, these probabilistic models can introduce latent noise, complicating the distinction between variants of interest.

To address these challenges, we propose GenomeVAP, a deterministic 'evidence integration' framework. By utilizing a weighted, rule-based approach rather than probabilistic inference, the framework provides a transparent, reproducible pipeline for retrieving and consolidating genomic data.

Materials and Methods

GenomeVAP, our deterministic 'evidence integration' framework, is designed to streamline genome data from varied definite databases into a single portal and ensure that all research annotations are directly traceable to their source data. This paper outlines GenomeVAP's system architecture, detailing the integration of its major analysis modules, and demonstrating the framework's utility through representative entries.

The following figure (Figure 1 depicts the methodology workflow of 'GenomeVAP' framework.

f1

Architecture Creation

The GenomeVAP framework is structured as a modular, web-based application utilizing a Python Flask backend,[9] chosen for its scalability and ease of integration with diverse APIs (Application Programming Interface). The system architecture is organized into three distinct layers to ensure robust data handling:

a) Orchestration Layer: Manages asynchronous requests to NCBI E-utilities [10] and EBI Ensembl endpoints.[3]

b) Normalization Middleware: Maps heterogeneous API responses to a standardized internal schema.

c) Interpretation Engine: Applies rule-based scoring (non-probabilistic).

Integration of Major Software Analysis Modules

GenomeVAP integrates several specific modules to support diverse research requirements intended for automated report generation. Such modules considered for this framework include,

a) Single Variant Analysis: The primary module accepts a unique variant identifier (e.g., rsID or chromosomal coordinate) and retrieves all associated data from registered sources.

b) CNV Analysis: Dedicated module for assessing Copy Number Variants (CNVs), enables the retrieval of overlapping gene annotations [11] and phenotypic associations of structural variations.

c) Batch Analysis: Facilitates high-throughput processing by allowing users to upload CSV or VCF files containing multiple variants, which are then queued for sequential retrieval and normalization.

d) Comparative Analysis: Provides a side-by-side interface for contrasting two or more variants, allowing researchers to visualize differences in trait associations or clinical significance scores.

e) Disease Panel Designer: Allows users to filter evidence through customizable panels (e.g., cardiovascular disease, metabolic syndrome), focusing retrieval on specific gene sets or phenotypic associations [12,13].

f) Cohort Analysis: Aggregates findings across a set of processed samples, enabling the identification of common variant profiles within specific user-defined groups.

Genomic Evidence Pooling

Retrieval from multiple resources

The following table (Table 1) elaborates the nature of data retrieved from databases / scientific resources and their purpose in GenomeVAP's framework.

Table 1: Data Retrieval from multiple scientific resources / databases

Database / Resource	Data Retrieved	Purpose in GenomeVAP
ClinVar	Clinical significance, review status, associated conditions	Clinical evidence interpretation.
dbSNP	rsID, chromosome, position, alleles.	Variant identification and genomic mapping.
Ensembl	Gene annotation, transcripts, genomic coordinates.	Functional annotation and gene mapping.
GWAS Catalog	Trait associations, p-values, study information.	Association between variants and phenotypes.
PubMed	Supporting scientific literature.	Literature evidence and publication links.

Normalization and Harmonization

The following table (Table 2) explains the normalization of integrated evidence through a few examples.

Table 2: Normalization of Integrated Evidence

Variant	Gene	Disease / Trait	Evidence Sources Retrieved	GenomeVAP Output
rs429358	APOE	Alzheimer's disease	ClinVar, GWAS, Ensembl, PubMed	Integrated interpretation report.
rs7412	APOE	Lipid metabolism / Alzheimer's	ClinVar, GWAS, Ensembl	Comparative evidence profile.
rs334	HBB	Sickle cell disease	ClinVar, Ensembl, PubMed	High clinical significance.
rs1800562	HFE	Hereditary hemochromatosis	ClinVar, GWAS, PubMed	Disease association summary
rs9939609	FTO	Obesity	GWAS, Ensembl	Trait association report.

Deterministic Rule Based Scoring

The following table (Table 3) explains the basis of 'deterministic' rule based scoring through evidence categorization from scientific resources / databases.

How to cite this article: Anusha S, Govindharajan M. A Deterministic Framework for Integrated Genome Variant Interpretation - The *'GenomeVAP'*. Curr Trends Biomedical Eng & Biosci. 2026; 24(3): 556139. DOI: 10.19080/CTBEB.2026.24.556139

Table 3: Evidence categorization and evaluation logic for deterministic rule-based scoring

Evidence Category	Source	Evaluation Logic	Contribution
Clinical Significance	ClinVar	Uses submitted clinical classification and review status.	Primary evidence.
Trait Association	GWAS Catalog	Prioritizes associations with genome-wide significant p-values.	Phenotypic evidence.
Population Frequency	Ensembl / gnomAD	Compares allele frequency across populations.	Population context.
Literature Evidence	PubMed	Links supporting publications.	Supporting evidence.
Gene Annotation	Ensembl	Maps variant to gene and transcript.	Functional context.

Determination of Framework's Utility

The following table (Table 4) shows the framework's utility by comparing conventional manual interpretations with GenomeVAP's evidence-integrated automation.

Table 4: Conventional manual variant interpretation Vs. GenomeVAP's evidence-integrated automation

Feature	Conventional Manual Workflow	GenomeVAP Workflow
Variant Identification	Search each rsID individually in dbSNP or Ensembl.	Single rsID submission through a unified interface.
Clinical Significance	Retrieved separately from ClinVar.	Automatically integrated into the report.
Trait Associations	Independently searched in the GWAS Catalog.	Retrieved and summarized alongside other evidence.
Population Frequency	Requires navigation to external population databases.	Displayed within the integrated report.
Gene Annotation	Retrieved separately from Ensembl or GeneCards.	Automatically included in the interpretation.
Literature Evidence	Manual PubMed search for supporting studies.	Direct PubMed references provided with results.
Evidence Consolidation	Researcher manually consolidates relevant information from multiple resources.	Evidence consolidated into a single report through coding and automation.
Batch Processing	Repeated manual analysis for each variant.	Multiple variants analyzed simultaneously.
Comparative Analysis	Requires manual comparison across databases or spreadsheets.	Built-in side-by-side comparative analysis module.
Disease Panel Design	Manual literature review and gene selection.	Automated panel generation based on selected disease criteria.
Cohort Analysis	Performed using external analytical workflows.	Integrated cohort analysis module.
Report Generation	Manual documentation and summarization.	Automated generation of structured interpretation reports.
Reproducibility	Depends on researcher workflow and documentation.	Standardized, deterministic workflow producing consistent outputs.
Academic Research Utility	Requires familiarity with multiple independent databases.	Single interface suitable for teaching and exploratory research.

Findings and Discussion

GenomeVAP effectively mitigates the manual search bottleneck inherent in genomic variant analysis. By integrating evidence from disparate databases into a unified, deterministic framework, it provides a reproducible alternative to opaque, probabilistic models. Its deterministic nature of the scoring system ensures that researchers can trace every calculation back to the original source data, a credible feature for scientific transparency in academic and healthcare research settings.

Comparing this automated approach to manual interpretation reveals clear advantages. Manual interpretation often requires a researcher to synthesize information across browser tabs and multiple database interfaces, a process susceptible to fatigue and inconsistent data evaluation. GenomeVAP standardizes this process, ensuring that every variant is subjected to the same rule-based scoring criteria. GenomeVAP provides a reproducible, transparent, and efficient research environment for the deterministic integration of genomic variant data. By centralizing evidence retrieval and standardizing output, the framework empowers researchers to focus on analysis rather than data acquisition.

Limitations and scope for further research

The GenomeVAP framework, however, remains subject to the inherent limitations of external API availability. Reliance on NCBI [10] and EBI [3] endpoints means that service outages or API updates may impact functionality. Most importantly, this framework is exclusively meant for improving time economy and limiting the cumbersomeness of manual effort in genome variant analysis. Hence, it does not claim clinical decision support. Future research should focus on tackling service outages, improving

How to cite this article: Anusha S, Govindharajan M. A Deterministic Framework for Integrated Genome Variant Interpretation - The *'GenomeVAP'*. Curr Trends Biomedical Eng & Biosci. 2026; 24(3): 556139. DOI: 10.19080/CTBEB.2026.24.5560139

API resilience and expanding the supported evidence sources. The GenomeVAP framework was evaluated using representative genomic variants with well-characterized biological and clinical associations to demonstrate its functionality and utility. Upon submission of a variant through either an rsID [2] or chromosomal coordinate, the framework retrieves evidence in parallel from multiple genomic resources, including ClinVar,[1] dbSNP, [2] Ensembl,[3] GWAS Catalog,[4] and PubMed.[14] The retrieved information is normalized into a unified internal schema, integrated through a deterministic rule-based evidence synthesis pipeline, and presented as a comprehensive interpretation report. The modular architecture, comprising the presentation layer, Flask-based orchestration layer,[9] evidence integration engine, and report generation module, enables seamless retrieval and consolidation of genomic annotations, clinical significance, population frequencies, trait associations, gene context[16], biological pathways [15], pharmacogenomic information, and supporting literature. In addition to single-variant interpretation, GenomeVAP supports simultaneous comparison of multiple variants through an integrated comparative analysis module, facilitating side-by-side evaluation of gene annotations, evidence confidence, clinical relevance [16,17], and prioritization metrics. The framework further extends its functionality through a Disease Panel Designer, which automatically prioritizes candidate genes by integrating genomic evidence, literature support, and deterministic scoring, thereby generating evidence-based panels suitable for research and educational applications. Together, these integrated modules demonstrate the framework's ability to centralize fragmented genomic evidence into a unified, interactive, and human-readable platform that supports efficient exploratory genomic analysis while reducing the need for repetitive manual database searches. [18-21]

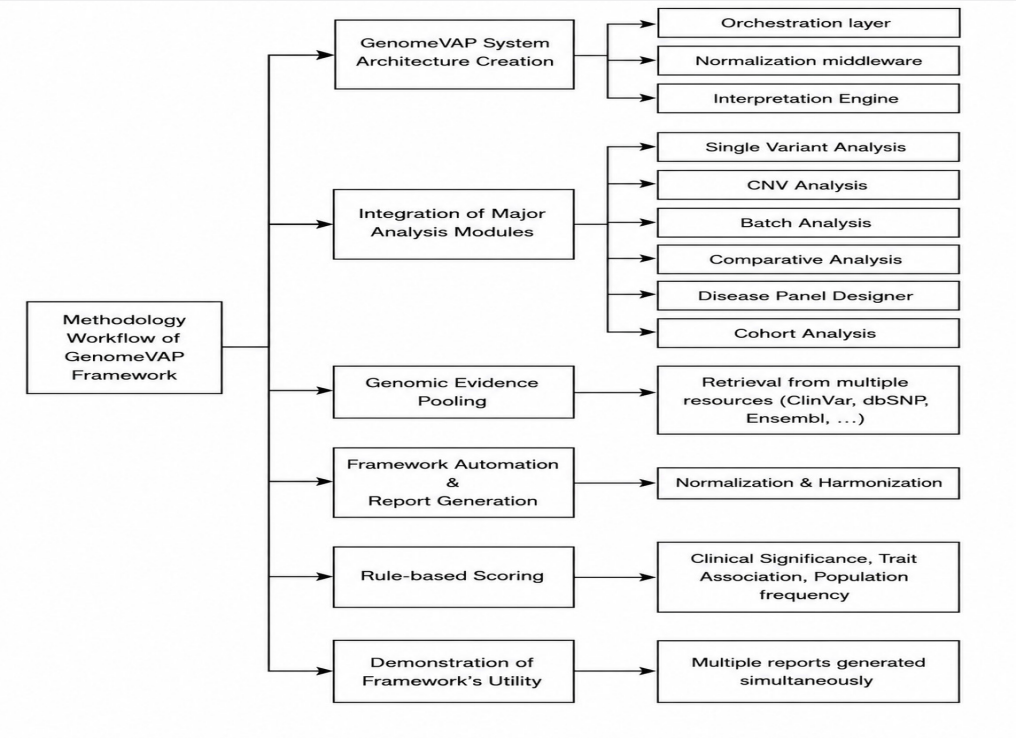

Figure 1: Methodology workflow of 'GenomeVAP' framework

References

1. Landrum MJ, Lee JM, Benson M, Garth RB, Chen C, et al. (2018) ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research 46(D1): D1062-D1067.
2. Sherry ST, Ward MH, Kholodov M, J Baker, L Phan, et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 29(1): 308-311.
3. Howe KL, Achuthan P, Allen J, Jorge AJ, Irina MA, et al. (2021) Ensembl 2021. Nucleic Acids Research 49(D1): D884-D891.
4. Buniello A, MacArthur JAL, Cerezo M, Laura WH, James H, et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies. Nucleic Acids Research 47(D1): D1005-D1012.
5. Richards S, Aziz N, Bale S, David B, Soma D, et al. (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. Genetics in Medicine 17(5): 405-424.
6. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M, et al. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research 47(D1): D886-D894.
7. Adzhubei IA, Schmidt S, Peshkin L, Vasily ER, Anna G, et al. (2010) A method and server for predicting damaging missense mutations. Nature Methods 7(4): 248-249.

How to cite this article: Anusha S, Govindharajan M. A Deterministic Framework for Integrated Genome Variant Interpretation - The *'GenomeVAP'*. Curr Trends Biomedical Eng & Biosci. 2026; 24(3): 556139. DOI: 10.19080/CTBEB.2026.24.556139

8. Pauline CNg, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research 31(13): 3812-3814.

9. Grinberg M (2018) Flask Web Development. 2nd ed. O'Reilly Media.

10. Sayers EW, Beck J, Brister JR, Devon B, James RB, et al. (2024) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 49(D1): D10-D17.

11. McLaren W, Gil L, Hunt SE, Harpreet SR, Graham RSR, et al. (2016) The Ensembl Variant Effect Predictor. Genome Biology 17(1): 122.

12. Köhler S, Gargano M, Matentzoglu N, Leigh CC, David LS, et al. (2021) The Human Phenotype Ontology in 2021. Nucleic Acids Research 49(D1): D1207-D1217.

13. Musen MA, Bean CA, Cheung KH, Patricia LW, Christopher GC, et al. (2012) The National Center for Biomedical Ontology. Journal of the American Medical Informatics Association 19(2): 190-195.

14. Canese K, Weis S (2013) PubMed: the bibliographic database. In: The NCBI Handbook. National Center for Biotechnology Information.

15. The Gene Ontology Consortium (2021) The Gene Ontology Resource: enriching a gold mine. Nucleic Acids Research 49(D1): D325-D334.

16. Li MM, Datto M, Duncavage EJ, Shashikant K, Neal IL, et al. (2017) Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. Journal of Molecular Diagnostics 19(1): 4-23.

17. Good BM, Ainscough BJ, McMichael JF, Joshua FMM, Adam CC, et al. (2020) The CIViC knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nature Genetics 49(2): 170-174.

18. Karczewski KJ, Francioli LC, Tiao G, Beryl BC, Jessica A, et al. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581: 434-443.

19. Cock PJA, Antao T, Chang JT, Brad AC, Cymon JC, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11): 1422-1423.

20. Uni Prot Consortium (2023) Uni Prot: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51(D1): D523-D531.

21. Rehm HL (2017) Evolving health care through personal genomics. Nature Reviews Genetics 18(4): 259-267.

DOI: 10.19080/CTBEB.2026.24.5560139

How to cite this article: Anusha S, Govindharajan M. A Deterministic Framework for Integrated Genome Variant Interpretation - The *'GenomeVAP'*. Curr Trends Biomedical Eng & Biosci. 2026; 24(3): 556139. DOI: 10.19080/CTBEB.2026.24.5560139