Scan and Explore!

Precision Meets Efficiency: festiVAR Streamlines Variant Prioritization with 99.93% Accuracy

festiVAR (fast estimation of variants for automated reporting), a variant prioritization algorithm developed by Strand's bioinformaticians, enables faster and more accurate identification of rare genetic disorders by automating critical steps in the diagnostic process and integrating into the whole exome sequencing (WES) workflow.

festiVAR Advantage: Refining WES Workflow Efficiency

- festiVAR cuts down the number of variants needing review from ~50k to just ~50, thereby streamlining the diagnostic process significantly.
- Leveraging LLMs, the tool improves the correlation between genetic data and clinical indications, ensuring more accurate identification of relevant gene variants, further reducing the number of variants to be assessed to ~25.
- Compliant with ACMG guidelines, it automates the variant classification process, greatly reducing the time spent on manual literature searches and speeding up the pathogenicity assessment, leading to the final reporting of typically 1-2 variants.
- Implementation of festiVAR has led to identifying the correct variants within the top 25 genes in 99.93% of cases, significantly improving productivity and time efficiency in handling exomes.

3 Key Outcomes—High Accuracy, Improved Productivity, Faster Results

festiVAR integrates into our diagnostic process, displaying ranked gene lists with phenotypic data for quick reference, which enables efficient and accurate genotype-phenotype correlations. The system ensures that users benefit from faster interpretations, reduced turnaround times, and fewer manual checks, leading to quicker and more accurate diagnostic reports that enhance patient care outcomes.

High Accuracy and Improved Productivity:

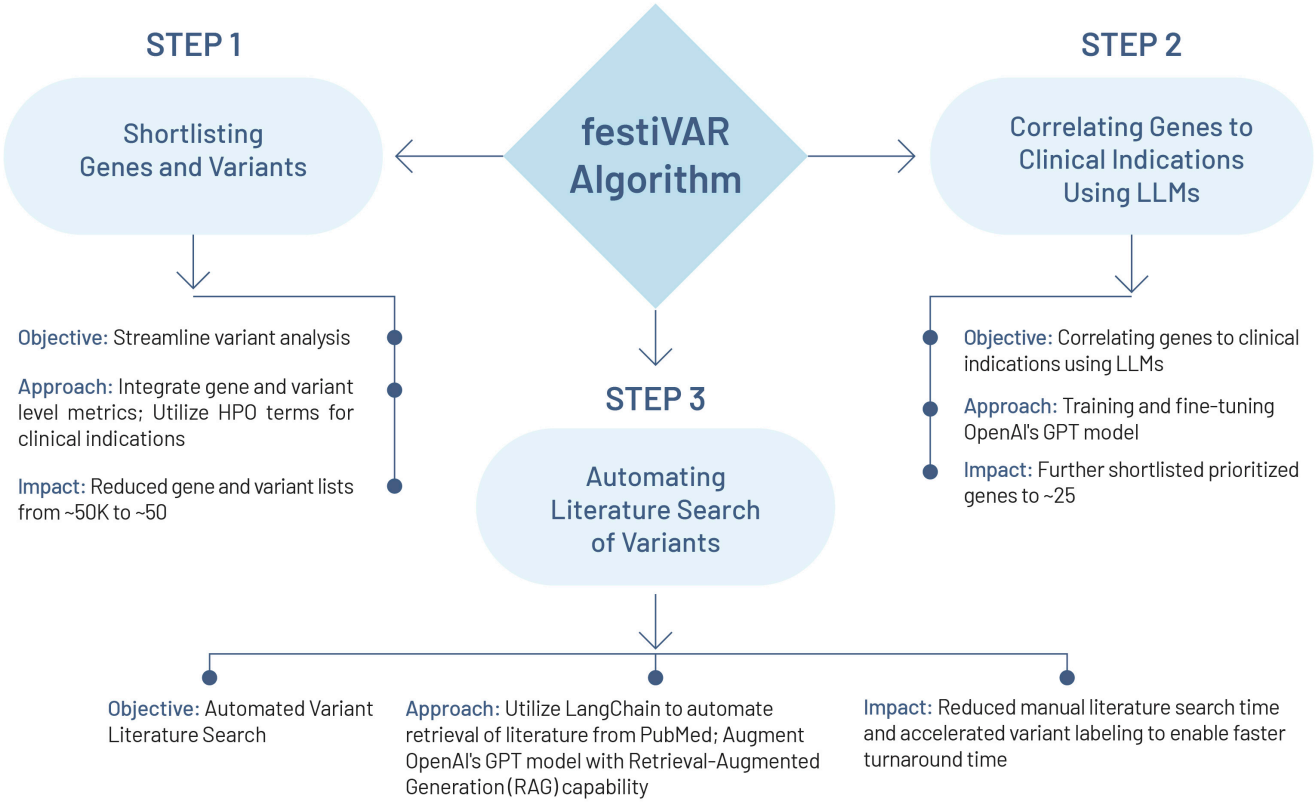
- From a pool of 996 cases that were analyzed in our CAP-accredited laboratory using an algorithm that on average shortlisted ~60 genes per case, the festiVAR system identified the reported variant within the top 25 genes in 99.93% of cases—a substantial improvement over the previous algorithm.

Time Efficiency:

- The introduction of LLMs has significantly reduced the time needed for variant evaluation, from 6 hours to under 2 hours, allowing our team to process hundreds of exomes monthly. Specifically, over 500 cases are processed through festiVar on a monthly basis.

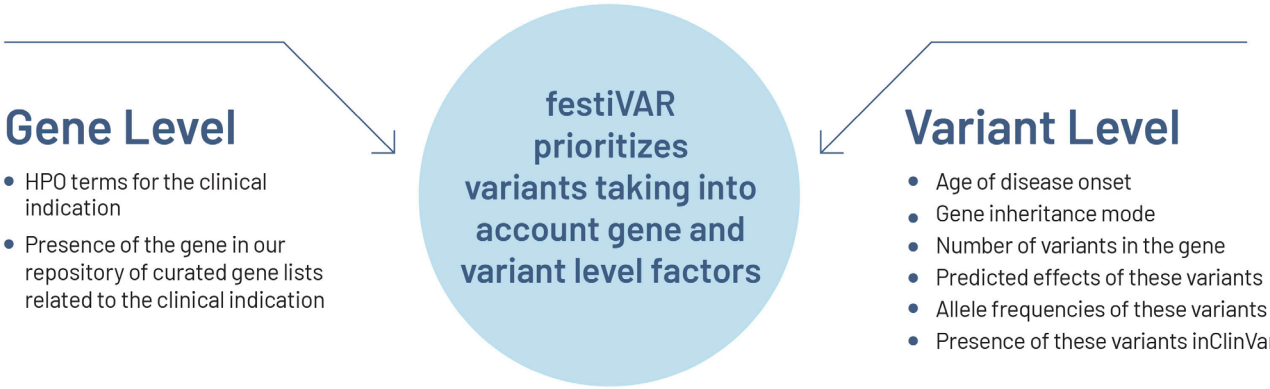
Inside festiVAR—Building Speed and Efficiency

Our three-step automation process is described below:



1. Gene and Variant Shortlisting

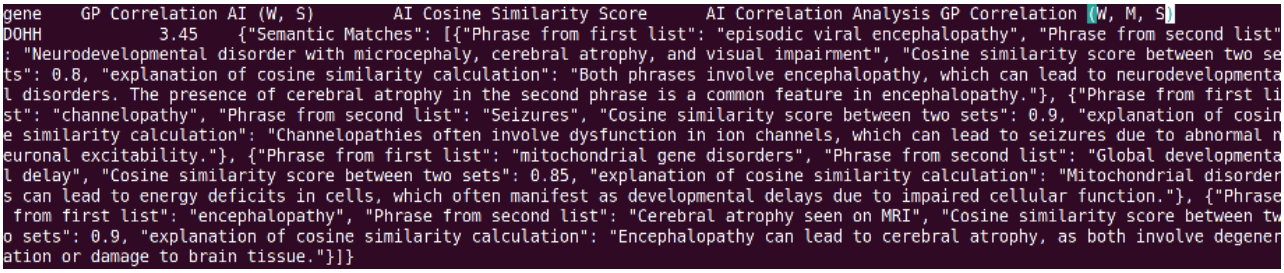
- Reduce the volume of variant analysis to manageable levels by incorporating gene-level and variant-level metrics and Human Phenotype Ontology (HPO) terms from clinical indications.



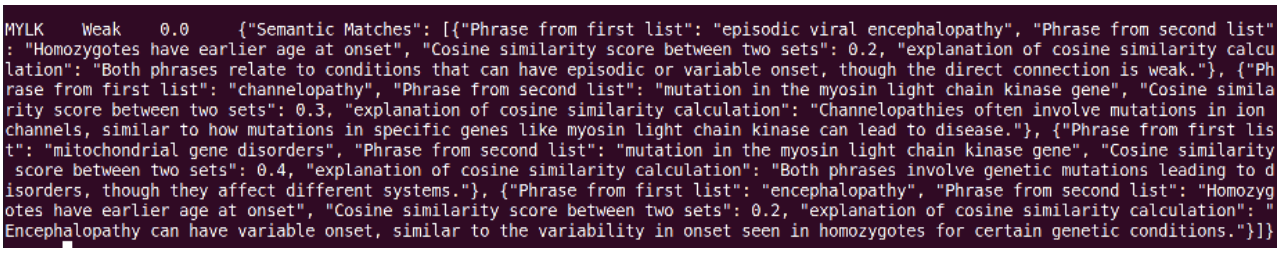
- Adjust for the variable number of genes associated with different HPO terms through a normalization scheme and by implementing a scoring system that prioritizes unique gene associations while de-emphasizing common terms.

2. Using LLMs to Correlate Clinical Indications

- Refine the gene list by correlating with clinical symptoms and indications, through the training and fine-tuning of OpenAI's GPT 4o model using a manually curated dataset of gene-phenotype correlations.
- In order to improve the accuracy of matching genes with clinical indications and minimize false positives, we implemented a cosine similarity scoring system.
    - A customized script was developed to filter out repeated phrases and apply a cosine similarity score cutoff.
    - This led to a 30% reduction in the number of genes requiring further verification of their association with clinical indications.
    - Additionally, each gene analysis includes AI analysis that is automatically generated and explains the rationale behind the similarity scores, facilitating faster validation.
    - This approach not only improves accuracy but also greatly reduces the time required for interpretation. The implementation of festiVAR in our pipeline has significantly reduced manual literature search time and accelerated variant labeling, decreasing the turnaround time from 6 hours to under 2 hours.



An example of the rationale and summary generated for a strong case

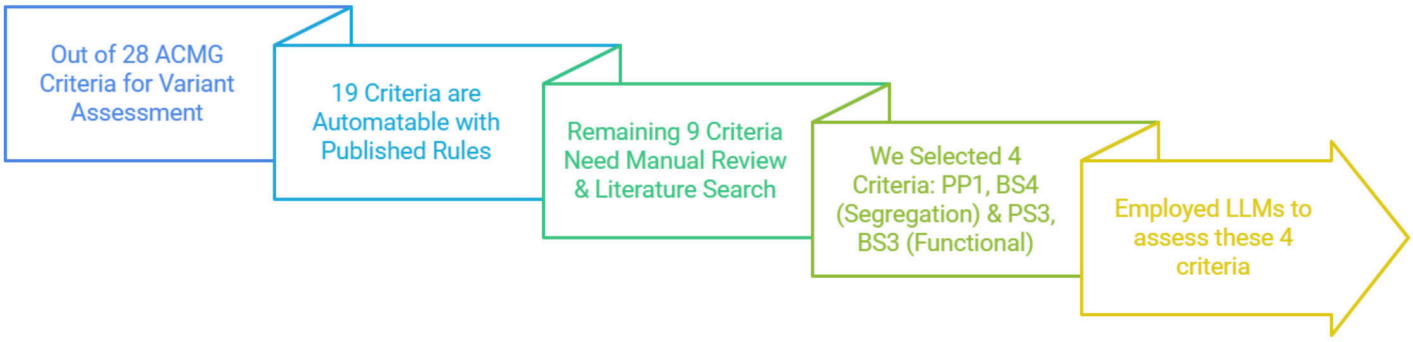


An example of the rationale and summary generated for a weak case

- A frequent challenge in WES-based clinical reporting of hereditary disorders is classifying variants of uncertain significance (VUSs), which can be time-consuming and delay report generation. Our experience indicates that cases involving multiple medium-ranked genes that exhibit weak correlations with clinical symptoms are typically associated with a higher occurrence of VUSs.

3. Automated Variant Literature Search

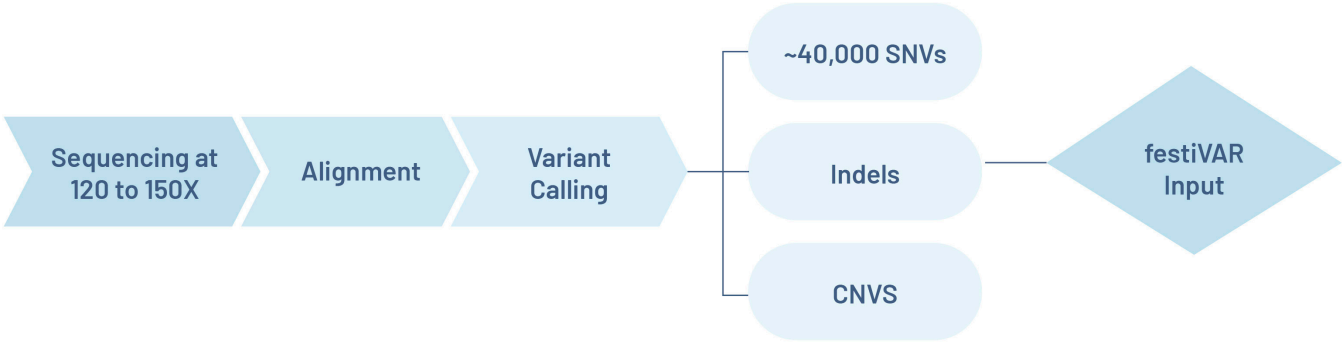
- Streamline variant classification as per ACMG guidelines by automating literature searches using LangChain to retrieve relevant studies from PubMed and PubMed Central, followed by augmenting OpenAI's GPT model with Retrieval-Augmented Generation (RAG) capabilities.



- This automation reduces manual search time for researchers and addresses 19 of the 28 ACMG criteria, greatly speeding up the pathogenicity assessment process.

What Sets festiVAR Apart:

- festiVAR analyzes around 40,000 SNVs, small InDels, and CNVs, obtained from sequencing at a 120-150x depth.



- It then prioritizes variants based on gene and variant-level factors, reducing the volume of variant analysis to manageable levels.
- Benchmarking against a previous variant interpretation algorithm, festiVAR prioritized all reported variants within the top 25 genes with an accuracy of 99.93% across approximately 1000 cases.
- festiVAR was augmented with an LLM-based approach, fine-tuning the GPT 4o with clinical notes and 1,418 manually assessed gene-phenotype correlations to achieve an accuracy of 98.43%.
- Lastly, we have integrated a LangChain application to automate scientific literature searches and implemented an assessment interface.

In conclusion, the festiVAR tool, now in production with a user interface, allows our team to efficiently process hundreds of exomes each month. This makes it an essential asset for any diagnostics company specializing in inherited disorders, using WES/WGS to identify causative genes and variants for clinical indications.

Omics CRO

Curation

15 years of experience curating variants, genes, pathways and diseases for clinical reporting and pharma/biotech custom solutions

~50
Molecular Biologists

Bioinformatics and Software

22 years of experience providing bioinformatics solutions to global instrument, diagnostic and pharma companies

~220
SW Engineers, Bioinformaticians

Omics Assays

11 years of experience with sequencing-based diagnostics across oncology and genetics, at our CAP lab in India

~90
Lab Scientists, Clin. Res. Scientists

> We were very impressed with the quality of work and timeliness; you're definitely our go-to for bioinformatics consulting
>
> - Director, Bioinformatics, Illumina

> We were immensely impressed by Strand's ability to rapidly recruit a substantially sized clinical cohort of cancer patients, and to design and run a complex liquid biopsy panel on samples drawn from the cohort, all in roughly a year's time.
>
> - Dr. Nishant Agarwal
> Chief of Otolaryngology-Head and Neck surgery and director of Head and Neck Surgical Oncology, University of Chicago.

> We have been using the StrandOmics pipeline to analyze and generate a report for our clinical cancer panels for over three years now. i would highly recommend using it to analyze data generated from clinical cancer NGS panels and the outputted clinical report provided after analysis.
>
> - Senior Scientist/ Medical laboratory director for NY State, Prim Bio Research Institiute

25+ YEARS OF EXPERIENCE

80,000+ Genetic Tests Reported

500+ Projects Executed for Genomics Majors Globally

Presence in **20+** Countries

illumina
Illumina Connected Analytics (ICA) Training and Onboarding

aws PARTNER

strand

7th Floor, MSR North Tower, #144, Outer Ring Road, Nagavara, Bengaluru - 560045

+91 9980448044 hello@strandls.com strandls.com