BMC Bioinformatics

SOFTWARE

Open Access

ACVI-Med, an open source variant interpretation tool for medical genomics

Moritz Stadler[1], Sabina Gansberger[1], Liliane Borik-Heil[1], Johannes Griss[1], Ichiro Okamoto[1], Wolfgang P. Weninger[1] and Matthias Wielscher[1*]

*Correspondence:
Matthias Wielscher
matthias.wielscher@meduniwien.ac.at
[1]Department of Dermatology, Medical University of Vienna, 1090 Vienna, Austria

Abstract

Background The adoption of whole genome and exome sequencing (WGS/WES) is expanding rapidly, yet variant interpretation remains a critical bottleneck, limiting diagnostic yield and delaying precision care. We present ACVI-Med (Accelerated Candidate Variant Interpretation in Medicine), an open-source, on-premise platform for comprehensive analysis of genomic data.

Results ACVI-Med integrates phenotype-aware filtering, ACMG-guided variant classification, and secondary findings detection into a unified, containerized system built for reproducibility and deployment across diverse institutional environments. Designed for accessibility, ACVI-Med enables users without extensive bioinformatics expertise to interpret variants from VCF (Variant Call Format) files through a flexible interface supporting phenotype-driven filtering, ACMG-guided variant classification, and integration of population, clinical, and predictive annotations. The software incorporates modules for secondary findings detection, aligned with ACMG Secondary Findings v3.2 recommendations, extending its utility beyond primary variant analysis. ACVI-Med's customizable filtering logic, real-time access to curated knowledge bases, and interactive visualization tools support transparent and reproducible genomic interpretation.

Conclusion ACVI-Med is an open-source software for genomic variant interpretation. Unlike commercial solutions, ACVI-Med is cost-effective, containerized, and can be institutionally self-hosted, reducing barriers to adoption in research settings.

Keywords ACMG guidelines, Variant interpretation, Medical genomics, Open-source software, Secondary findings, Pharmacogenomics

Background

Medical genomics significantly impacts all areas of medicine, both in clinical practice and research. While many medical professionals can easily suspect a genetic component in a patient's disorder, pinpointing the exact mutation often remains challenging. Genomics England, who evaluated the overall percentage of successfully solved patient cases, found that around 25–40% of rare disease cases could be definitively diagnosed, using the most advanced sequencing technologies and clinical data integration. One of

the major contributing factors to the number of unsolved cases is the annotation bottleneck, which describes the difficulty in translating raw genomic data into clinically actionable insights [1]. While sequencing has become faster and more affordable, the interpretation of millions of variants per genome remains time-consuming and often requires specialized expertise. Several efforts have been made to resolve the annotation bottleneck, including variant interpretation and secondary findings frameworks by the ACMG (American College of Medical Genetics) [2–4] and international data-sharing standards like those developed by GA4GH (Global Alliance for Genomics and Health) [5]. Tools such as the Genomics England PanelApp [6] and high-throughput genome analysis pipelines like Illumina DRAGEN [7] demonstrate how comprehensive, scalable genome interpretation is increasingly feasible in both research and clinical settings.

Variant interpretation pipelines typically involve several stages, including annotation with tools such as Ensembl's Variant Effect Predictor (VEP) [8], ANNOVAR [9], or SnpEff [10], followed by filtering and prioritization using clinical databases like ClinVar [11], gnomAD [12], and OMIM [13]. These tools form the backbone of most academic workflows, yet they require considerable bioinformatics expertise and manual integration of multiple software components. To simplify this process, several web-based and commercial platforms - such as VarSome [14], Congenica or Euformatics - offer integrated variant analysis environments. These systems are closed-source, relatively expensive, and reliant on third party infrastructure, which can raise data privacy and regulatory concerns, particularly in settings with real patient data. In parallel, open-source solutions like Exomiser [15], seqr [16] and the Molecular Tumor Board Portal (MTBP) [17] have emerged to support variant prioritization in research environments. Additionally pathogenicity prediction and variant prioritization servers, such as MutationTaster [18], focus on estimating the functional impact of individual variants and are commonly used as components within larger analysis workflows. While these tools have substantially advanced rare disease genomics, they typically address specific stages of the interpretation process, such as variant-level pathogenicity prediction or phenotype-driven prioritization, rather than providing fully standardized, transparent ACMG-based [2] variant classification. Many require nontrivial setup and continuous maintenance, making usage challenging for smaller institutions without dedicated bioinformatics support. Beyond accessibility, a major obstacle in medical genomics is the lack of transparent and reproducible pipelines for variant classification [19]. Most cloud-based systems do not provide insight into the exact rules or thresholds applied when computing pathogenicity tiers, complicating quality assurance and regulatory compliance [20].

In consequence, there remains a need for an open, easily deployable system that combines ACMG-compliant classification, phenotype-aware filtering, and secondary findings detection within a secure, on-premise infrastructure (see Table 1).

We present an open-source and easy to use, yet powerful software solution designed for academic institutions and independent researchers. This application enables a wide range of genetic analyses on WGS or WES (Whole Genome Sequencing) data. It offers comprehensive search capabilities, ACMG-compliant variant tiering, and includes modules for pharmacogenomics and secondary findings.

Table 1 A comparison with existing tools. Placing ACVI-Med in the context of other established open-source applications

Primary focus	ACVI-Med	Exomiser	seqr	MTBP
	Case-level variant interpretation	**Phenotype-driven variant prioritization**	**Collaborative variant analysis**	**Collaborative variant analysis**
Guided classification	Automated ACMG classification, transparent rule evaluation	Automated ACMG classification	Primarily manual	Automated ESCAT-ranked actionability
Phenotype integration	HPO- and panel-based filtering	HPO-base filtering	Phenotype annotation and filtering	Clinical/pathological staging & disease hierarchy
Deployment model	On-premise/HPC environments	On-premise/HPC environments	Cloud-hosted (Broad)/On-premise (Open-source)	Cloud-hosted/ On-premise (Open-source)
Target users	Researchers without dedicated bioinformatics teams	Bioinformatics-supported research groups	Large collaborative research projects	Oncology medical teams and Tumor Boards
Additional modules	Integrated PGx & ACMG SF v3.2	Manual (via gene lists)	Manual (via gene lists)	Automated trial matching & germline screening

Implementation

System architecture

ACVI-Med is a containerized client-server software platform with a React JS frontend and a Java Spring backend. Users access the application via their web browser. Genomic data in the VCF format is imported via a custom parser, processed, and stored in a central PostgreSQL [21] database, while user data, metadata about imported samples and permissions are stored in a separate database. Each service runs in an isolated Docker container and is managed collectively via Docker Compose, allowing simple deployment on standard Linux systems and scalability across institutional environments. The main backend service processes requests, checks permissions, and performs database queries. User authentication, access control, and sample-specific permissions are managed centrally, ensuring secure and auditable handling of sensitive genomic data. Additionally, the backend handles periodically recurring tasks and automatically performs requests to external APIs such as Genomics England PanelApp [6], HPO (Human Phenotype Ontology) [22] or the ClinGen Criteria Specification (CSpec) Registry [23]. These external APIs are queried daily at 02:00 a.m., and the retrieved results are cached locally to ensure that the data required for further processing remains up to date. The frontend application provides a UI to interact with the backend by gathering input data and displaying filtering results. The communication between these services occurs via a REST API. ACMG-guided variant classification tiers [2] were implemented on a tier to tier basis following ACMG/AMP recommendations. These rules are automatically evaluated to enable standardized and reproducible variant interpretation.

VCF importer

The importer automatically detects and structures variant annotations, including predictive scores and allele frequencies, which facilitates downstream filtering and classification. The system supports multiple isoforms per variant and can process VCF [24] files containing differing numbers and types of INFO and consequence fields. During import the application traverses the VCF file twice. During the first traversal the importer

determines the number, names, formats and datatypes of the input file's INFO- and consequence fields, allowing dynamic inference of the appropriate database schema. After completion a custom database table is created, consisting of default and interfered columns. The second traversal inserts the data into the PostgreSQL database [21]. After insertion, commonly queried fields such as allele frequency, impact, and gene symbol are automatically indexed. A multicolumn index combining these fields further accelerates ACMG-based tiering and complex filter combinations. In benchmarking tests, this reduced query times on whole-genome datasets (~ 20 million variants) to a few seconds.

ClinVar integration and precomputation

To accelerate variant classification, ACVI-Med incorporates precomputed ClinVar datasets containing variant identifiers, clinical significance, associated phenotypes, and protein-level annotations. These are compressed and stored locally in four structured files that can be efficiently queried during tier evaluation, substantially improving ACMG classification performance.

A detailed description of the annotation workflow, the implementation of ACMG-guided variant classification, and further technical background is provided in the Supplementary Methods.

Results

We developed a software tool optimized for the analysis of whole genome and whole exome sequencing data called ACVI-Med (Accelerated Candidate Variant Interpretation in Medicine). This tool allows medical professionals with little or no particular experience in the field of genetics to analyze their samples and identify causative mutations for observed phenotypes. The tool is designed to process standard VCF (Variant Call Format) [24] files and provides a structured framework for variant interpretation. Our software aids medical professionals in prioritizing, filtering, and conducting agnostic inspections of variants, as well as in sharing their findings with colleagues. Figure 1 provides an overview of the ACVI-Med workflow and its main functional modules.

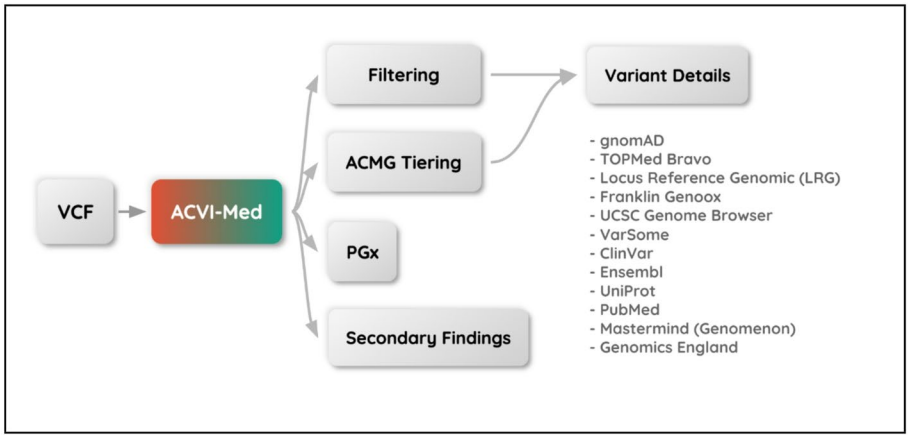

Fig. 1 Workflow overview. Starting from an Ensembl VEP [8] (Variant Effect Predictor) annotated VCF file, our software enables variant filtering and ACMG-compliant tiering [2] to identify candidate variants. Each variant can be individually inspected with links to a wide range of external data sources. In addition to variant prioritization, the tool includes modules for ACMG-recommended secondary findings and pharmacogenomics (PGx) analysis

Variant explorer and filtering

One of the core functions of ACVI-Med is filtering the patient's sequenced variants by gene, panel, genotype or a range of other variant attributes. This is used to identify pathogenic mutations, allowing users to quickly sift through the vast amount of variants and identify those of importance to the phenotype. Additionally, users can sort the data by a number of variant attributes (such as pathogenicity prediction scores or allele frequency) in ascending or descending order. The filtering page was designed to allow users to freely filter variants without constraining them to a fixed number of logical operations or a rigid structure, using an intuitive graphical user interface. Researchers may place conjunctions (AND, OR) between expressions and set parentheses around statements. This allows for a straightforward definition of a filter, without being limited to preset selection criteria. Building a filter consists of four steps, which the user is guided through via a graphical interface:

1. Selecting a number of genes or panels to be included in the search results. These genes and their associated panels are automatically retrieved from the Genomics England Panel App API on a daily basis [6].
2. Filtering for variant attributes such as protein impact, pathogenicity scores (SIFT [25], PolyPhen [26], DANN [27], GERP [28] among others) and allele frequencies as included in our annotation pipeline which is available as part of the software.
3. Filtering for a genotype of the patient or genotype combinations if the patient's relatives are included in the analysis (trio). This allows for e.g. the identification of recessive mutations where patients exhibit a genotype of 1/1 while parents are only carriers with 0/1.
4. Selecting columns to sort the result by.

Once the filter is applied, the number of variants displayed is reduced to those matching the filter. Clicking on a search result redirects the user to the variant's summary page containing all associated information.

Variant summary page

The summary page contains all information associated with the variant that was added through our annotation process as well as hyperlinks specifically generated for each variant, pointing to associated information in external databases (Fig. 2). This combination of locally annotated information and links to external sources should enable the user to evaluate the pathogenicity of the variant in each particular case. The links are generated from variant identifiers, gene symbols, protein ids and other data stored with the variant. Linked databases include gnomAd [12], TopMed Bravo [29], LRG [30], Franklin, UCSC [31], Varsome [14], ClinVar [11], CCDS [32], Ensembl [33], Uniprot/SwissProt [34], UniParc [35], PubMed, Mastermind [36] and Omim [13]. On the very top of the summary page a brief overview of the most important annotations of each variant - such as chromosome and position, reference allele, alternate allele and the filter field indicating the quality of the variant - as well as the fields GT (genotype), GQ (genotype quality), DP (depth, number of reads in a specific locus) and AD (allelic depth) are displayed for each person included in the sample. A polar area chart is displayed, showing the predicted effect on the gene product as determined by the available pathogenicity

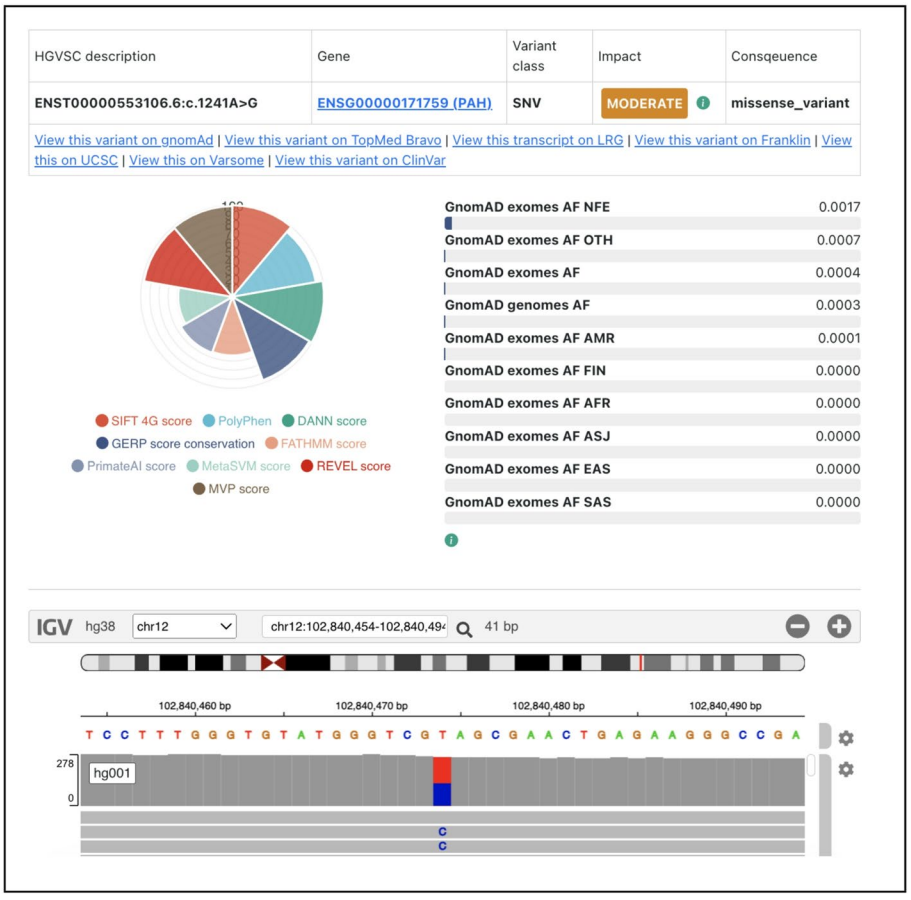

Fig. 2 Variant summary page. A screenshot of the variant summary page. Besides basic information about the variant, we display various pathogenicity prediction scores, population statistics and - when available - IGV [44] data

prediction scores including SIFT 4G [37], PolyPhen [26], DANN [27], GERP [28], FATHMM [38], PrimateAI [39], MetaSVM [40], REVEL [41], MVP [41] which should allow for a quick first evaluation of the variant before starting a more in-depth analysis. Pathogenicity prediction scores are normalized to a 0–100% scale, with 100% indicating the most damaging effect of the variant. A bar graph shows the allele frequencies of the variant in different populations, sorted from most to least frequent. If a CRAM (Compressed Reference-oriented Alignment Map) [42] or BAM (Binary Alignment Map) [43] file is provided, the Integrative Genomics Viewer (IGV) [44, 45] is displayed on the variant overview page, to allow inspection of the actual raw data leading to the variant call. All remaining INFO and consequence fields of the variant as stored in the VCF file are displayed in a predefined order, hierarchically grouped. Alongside each variant attribute a description of the selected property and - if available - a link to the original publication describing the property is given. Variant attributes such as scores or percentage values are displayed as bar charts. In case the variant or its gene is listed in ClinVar [11] or Genomics England PanelApp [6] the data available from both sources is displayed in separate information boxes.

ACMG tier-based filtering "Quick-Tiering"

To provide researchers with a simple way to perform a standardized screening of a patient's genome, an additional search function with a subsequent tiering of variants was implemented. Tiering variants according to the guidelines provided by the American College of Medical Genetics (ACMG) assigns each variant one or more ACMG tiers which can either support or oppose a variant's claim to pathogenicity [2].

The ACMG-based tiering process starts with the selection of one or more Genomics England Panel App panels and/or HPO (Human Phenotype Ontology) [22] terms that in combination best describe the condition of the patient in the analyzed sample. This tiering requires at least one HPO term or PanelApp panel. Our software converts each HPO term into a number of loci by searching the ClinVar database for the term's id and gathering all loci associated with it. It is important to note that for many HPO terms no association has been recorded in the ClinVar database, it is therefore advisable to choose multiple terms and use them in combination with Genomics England panels. The combination of HPO terms and Genomics England panels allows for the creation of a query used to select variants for tiering and further analysis. Besides terms and panels, queries can contain a variant allele frequency threshold of 10%, which decreases overall processing time, while simultaneously avoiding eliminating possibly pathological variants. Each variant included in the result set is evaluated based on the tiering rules established by ACMG [2]. This process results in a number of tiers for each variant in the initial result set (Fig. 3). To improve transparency, our tool provides detailed descriptions of each assigned tier. This description includes the list of conditions that have been met by the variant which led to the tier being selected and features the exact values that were used as a basis to determine if the tier applies. Out of the total 16 ACMG tiers supporting pathogenicity, our application implements 11, and out of the 12 tiers

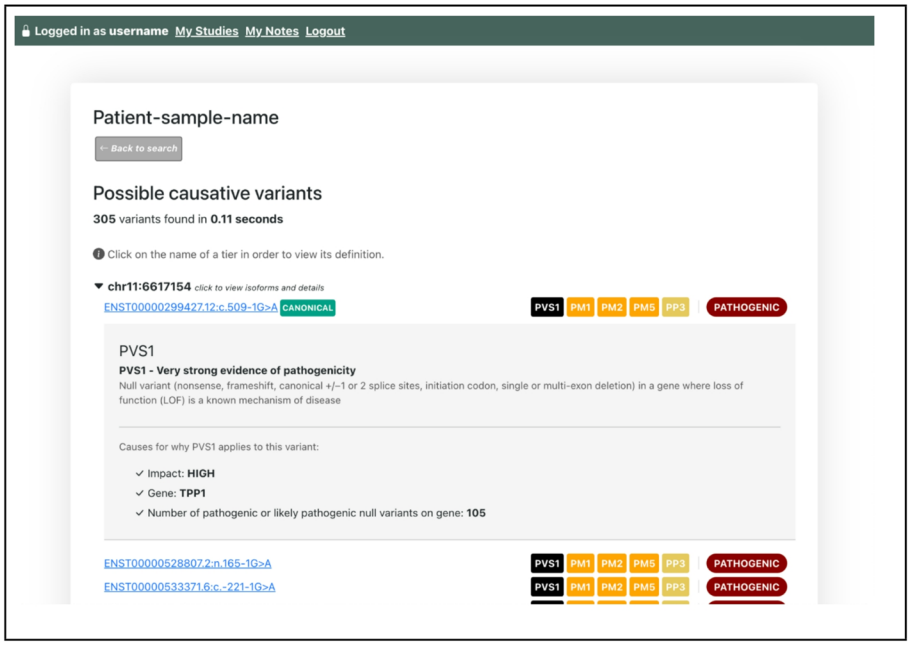

Fig. 3 ACMG-based tiering. A screenshot of the ACMG-based tiering results [2]. Each variant matching the phenotype in the search query is shown with the corresponding evidence tiers (e.g., 'PVS1', 'PM1', 'PM2') and the resulting classification (e.g., 'Pathogenic'). The reasoning for each tier and the final classification is shown separately to help users verify their accuracy

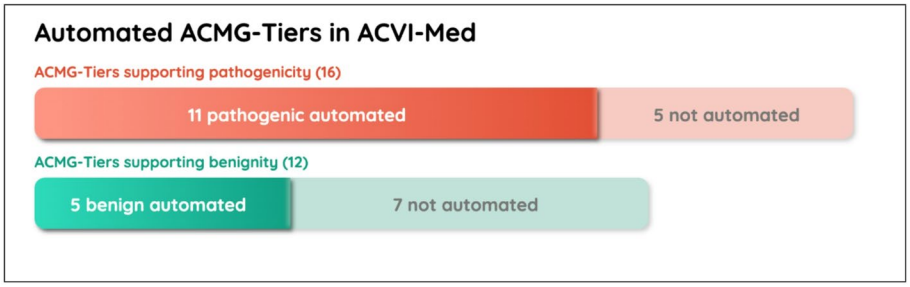

Fig. 4 Implemented ACMG-Tiers. Out of the 16 tiers supporting pathogenicity our application implements 11. Out of the 12 tiers supporting benignity, our application implements 5

supporting benignity, it implements 5 (Fig. 4). In the final step of the ACMG-based tiering process the tiered variants are sorted by ACMG tiers starting with the most severe variants. Using these tiers the variants are classified as either "pathogenic", "likely pathogenic", "uncertain significance", "likely benign" and "benign" according to the rules recommended by ACMG [2]. This approach simplifies identification of causative mutations by researchers, since the most damaging variants are displayed on the top of the results list and can easily be investigated further.

For variants located in genes under the purview of a ClinGen Variant Curation Expert Panel (VCEP) [21], the interface identifies the availability of specialized rules by cross-referencing the data from the ClinGen Criteria Specification (CSpec) Registry. While baseline ACMG criteria are automated, this feature allows users to manually select gene-specific evidence codes and strength modifications. These parameters are evaluated via the CSpec Reasoner SVI API [23], ensuring the final classification conforms to expert-level interpretations that supersede the general Richards et al. 2015 guidelines.

Additional capabilities

Secondary findings

In a 2013 paper the American College of Medical Genetics published a consensus paper in which they acknowledge that with the greater availability of WES and WGS the question of how to interpret incidental findings becomes a frequent practical concern [4]. Their working group established a list of conditions, genes and variants where findings should always be reported, no matter the original indication for ordering the test. Their list focuses on disorders that have an established treatment, a high penetrance and tend to be asymptomatic for a long time [4]. In order to comply with these recommendations, we added a secondary findings module based strictly on the ACMG SF v3.2 reporting guidelines [3]. The implementation of this feature requires the user to enter whether the patient is an adult or a child and if the patient is male (single X-chromosome) or female (multiple X-chromosomes) which is required to accurately follow ACMG recommendations on secondary findings. Once this information is entered, the software tiers all variants located on genes in the ACMG SF v3.2 list. Depending on the individual recommendations from said list only pathogenic or both pathogenic and likely pathogenic variants are reported in the final result which shows the number of identified variants per gene as well as detailed information on the identified variants. The ACMG-curated secondary findings list includes prominent genes such as BRCA1, BRCA2 and APC - that oftentimes were not the target of the primary investigation but nonetheless cause

potentially treatable dispositions, hoping to strike the right balance between over-reporting and missing critical diagnoses. Having knowledge about the patient's carrier status of these actionable variants allows the researcher to potentially start preventive treatment early on and facilitate a better overall outcome for the patient.

Pharmacogenomics

We additionally implemented a pharmacogenomics module which allows users to predict potential drug responses based on the patient's genetic makeup. The pharmacogenomics module in ACVI-Med integrates data from PharmGKB [46] by directly mapping identified variants to clinical annotations via their rsID. This approach focuses on individual variant-drug associations rather than translating combinations into star (*) alleles. The module reports findings based on the presence of specific variants detected in the VCF file, and it does not infer or report on missing variants that would be required for full haplotype reconstruction. For each identified interaction, the module provides detailed information, including the gene involved, the specific variant, and the predicted effect on drug metabolism or efficacy. This information is presented in a user-friendly format, allowing researchers to quickly assess the potential impact of genetic factors on drug therapy. By incorporating pharmacogenomics into the ACVI-Med tool, we aim to enhance the utility of genomic data analyzed.

Conclusion

ACVI-Med provides a user-friendly tool for variant interpretation, secondary findings identification, and pharmacogenomics analysis in genomics. ACVI-Med's accessibility is a key strength, allowing medical professionals with limited genetics experience to analyze genomic data using any VCF file. The tool's integration of ACMG-conform tiering and flexible filtering options enhances the efficiency of identifying causative mutations. Additionally, the secondary findings and pharmacogenomics modules extend its utility, offering insights into relevant mutations and drug response predictions, respectively.

One of the main advantages of ACVI-Med is its cost-effectiveness and open-source nature, which makes it an attractive alternative to commercial solutions like Varsome [14], Euformatics, and Congenica. Unlike these commercial tools, ACVI-Med is on-premise and fully containerized, allowing small institutions to effectively analyze samples. However, this also means that ACVI-Med lacks the dedicated support and extensive community features that commercial solutions can offer. Users need to rely on their own hardware and expertise to troubleshoot issues, which could be a limitation for some institutions.

Furthermore, we want to address the limitations of ACVI-Med's ACMG-based tiering. While we implemented the majority of the ACMG tiers, not all can be computed in an automated way. Given that ACMG only provides guidelines on the implementation of tiers, some details are open to interpretation. While we have tried to stay true to the exact definition, sometimes adaptations were necessary to ensure automated processing is possible. Lastly, it is important to note that since not all tiers were implemented, the final classification may be skewed. In order to ensure full transparency, we provide a list of all ACMG-tiers and their respective implementations or adaptations in our Supplementary Methods. Further limiting factors of ACVI-Med's effectiveness are the completeness and accuracy of external databases, out of which especially ClinVar

and Genomics England Panel App play an important role. Another critical consideration is the privacy and security of patient data. Genomic data, particularly BAM [43] files, can be identifiable and have to be handled with utmost care. While ACVI-Med provides access protection to safeguard this information, the institution hosting the application needs to ensure their infrastructure's security. Even rare single variants might make patients identifiable which warrants careful consideration of researchers when sharing results. One benefit of ACVI-Med in this context is that the analysis can remain entirely in-house.

We believe ACVI-Med offers a valuable solution for genomic data analysis, addressing the annotation bottleneck and providing a cost-effective alternative to commercial tools. Bridging the gap between high-end prioritization tools like Exomiser [15] and large-scale platforms like seqr [16], ACVI-Med offers a unique, transparent ACMG-based workflow (Table 1) for the research community. By empowering academic institutions and independent researchers, ACVI-Med facilitates the integration of genomics into research. Future improvements could focus on optimizing data processing times and expanding compatibility with new genomic data formats.

Although ACVI-Med is currently restricted to research use due to regulatory considerations, we hope it will inspire further collaboration between researchers, clinicians and developers, driving innovation in genomic medicine and accelerating the transition of genomics into everyday clinical care. By lowering barriers to advanced genomic analysis, we envision ACVI-Med playing a role in democratizing access to precision medicine and ensuring that cutting-edge genomic insights are within reach for institutions of all sizes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-026-06414-2.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
M.S. contributed to conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing (original draft and review & editing). S.G. contributed to validation and writing (review & editing). L.B. contributed to validation and writing (review & editing). J.G. contributed to formal analysis, validation, and writing (review & editing). I.O. contributed to validation and writing (review & editing). W.W. contributed to validation and writing (review & editing). M.W. contributed to conceptualization, data curation, formal analysis, funding acquisition, project administration, supervision, validation, and writing (review & editing).

Funding
This work received funds from the Vienna Science and Technology Fund (ME/CFS Call 2024 "Understanding ME/CFS", PI Wielscher, Grant agreement No. ME-CFS24-002). Some computational resources were provided in part by the Austrian Scientific Cluster (ASC).

Data availability
The software and all scripts used for analysis in this manuscript are openly available and published under the MIT license at: https://github.com/moritzstadler/acvi-med. Example datasets used to demonstrate the functionality of the software can be accessed through the Genome in a Bottle project of the National Institute of Standards and Technology (NIST): https://www.nist.gov/programs-projects/genome-bottle. Project name: ACVI-Med (Accelerated Candidate Variant Interpretation in Medicine). Project home page: https://github.com/moritzstadler/acvi-med. Operating systems: Platform independent, containerized. Installation on Linux is recommended. Programming language: Java (Spring-framework), Typescript (React.JS-framework). Other requirements: Docker. License: MIT. Any restrictions to use by non-academics: No restrictions. Demo page: https://derma-genomics.meduniwien.ac.at/.

Declarations

Ethics approval and consent to participate
Not applicable, no clinical trial number.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 22 October 2025 / Accepted: 24 February 2026
Published online: 04 March 2026

References

1. Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. N Engl J Med. 2021;385:1868–80. https://doi.org/10.1056/NEJMoa2035790.
2. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24. https://doi.org/10.1038/gim.2015.30.
3. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, et al. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2023;25:100866. https://doi.org/10.1016/j.gim.2023.100866.
4. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med Off J Am Coll Med Genet. 2013;15:565–74. https://doi.org/10.1038/gim.2013.73.
5. Rehm HL, Page AJH, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: international policies and standards for data sharing across genomic research and healthcare. Cell Genomics. 2021;1:100029. https://doi.org/10.1016/j.xgen.2021.100029
6. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51:1560–5. https://doi.org/10.1038/s41588-019-0528-2.
7. Behera S, Catreux S, Rossi M, Truong S, Huang Z, Ruehle M, et al. Comprehensive and accurate genome analysis at scale using DRAGEN accelerated algorithms. BioRxiv Prepr Serv Biol. 2024. https://doi.org/10.1101/2024.01.02.573821.:2024.01.02.573821.
8. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122. https://doi.org/10.1186/s13059-016-0974-4.
9. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–164. https://doi.org/10.1093/nar/gkq603.
10. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w. iso-2; iso-3 Fly (Austin). 2012;6:80–92. https://doi.org/10.4161/fly.19695.
11. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–7. https://doi.org/10.1093/nar/gkx1153.
12. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43. https://doi.org/10.1038/s41586-020-2308-7.
13. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33:514–7. https://doi.org/10.1093/nar/gki033. Database issue:D.
14. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35:1978–80. https://doi.org/10.1093/bioinformatics/bty897.
15. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10:2004–15. https://doi.org/10.1038/nprot.2015.124.
16. Pais LS, Snow H, Weisburd B, Zhang S, Baxter SM, DiTroia S, O'Heir E, England E, Chao KR, Lemire G, Osei-Owusu I. Seqr: a web-based analysis and collaboration tool for rare disease genomics. Hum Mutat. 2022;43(6):698–707. . https://doi.org/10.1002/humu.24366
17. Tamborero D, Dienstmann R, Rachid MH, Boekel J, Lopez-Fernandez A, Jonsson M, et al. The molecular tumor board portal supports clinical decisions and automated reporting for precision oncology. Nat Cancer. 2022;3:251–61. https://doi.org/10.1038/s43018-022-00332-x.
18. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7:575–6. https://doi.org/10.1038/nmeth0810-575.
19. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol. 2013;9:e1003285. https://doi.org/10.1371/journal.pcbi.1003285.
20. Public Population Project in Genomics and Society (P3G) International Steering Committee, International Cancer Genome Consortium (ICGC) Ethics and Policy Committee, Dove ES, Joly Y, Tassé A-M, Knoppers BM. Genomic cloud computing: legal and ethical points to consider. Eur J Hum Genet. 2015;23:1271–1278. https://doi.org/10.1038/ejhg.2014.196
21. PostgreSQL. Global Development Group. PostgreSQL.
22. Gargano MA, Matentzoglu N, Coleman B, Addo-Lartey EB, Anagnostopoulos AV, Anderton J, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. Nucleic Acids Res. 2024;52:D1333–46. https://doi.org/10.1093/nar/gkad1005.
23. Rivera-Muñoz EA, Milko LV, Harrison SM, Azzariti DR, Kurtz CL, Lee K, et al. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. Hum Mutat. 2018;39:1614–22. https://doi.org/10.1002/humu.23645.
24. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinforma Oxf Engl. 2011;27:2156–8. https://doi.org/10.1093/bioinformatics/btr330.

25. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–4. https://doi.org/10.1093/nar/gkg509.
26. Adzhubei I, Jordan DM, Sunyaev SR Predicting functional effect of human missense mutations using PolyPhen-2. Curr protocols Hum Genet. 2013;76(1):7–20. . https://doi.org/10.1002/0471142905.hg0720s76.
27. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinforma Oxf Engl. 2015;31:761–3. https://doi.org/10.1093/bioinformatics/btu703.
28. Cooper GM, Stone EA, Asimenos G, Comparative Sequencing NISC, Program, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15:901–13. https://doi.org/10.1101/gr.3577405.
29. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021;590:290–9. https://doi.org/10.1038/s41586-021-03205-y.
30. Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, et al. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med. 2010;2:24. https://doi.org/10.1186/gm145.
31. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006. https://doi.org/10.1101/gr.229102.
32. Pujar S, O'Leary NA, Farrell CM, Loveland JE, Mudge JM, Wallin C, et al. Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding regions supported by expert curation. Nucleic Acids Res. 2018;46:D221–8. https://doi.org/10.1093/nar/gkx1031.
33. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48:D682–8. https://doi.org/10.1093/nar/gkz966.
34. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004. https://doi.org/10.1093/nar/gkh131. 32 Database issue:D115-119.
35. Leinonen R, Diez FG, Binns D, Fleischmann W, Lopez R, Apweiler R. UniProt archive. Bioinforma Oxf Engl. 2004;20:3236–7. https://doi.org/10.1093/bioinformatics/bth191.
36. Chunn LM, Nefcy DC, Scouten RW, Tarpey RP, Chauhan G, Lim MS, et al. Mastermind: a comprehensive genomic association search engine for empirical evidence curation and genetic variant interpretation. Front Genet. 2020;11:577152. https://doi.org/10.3389/fgene.2020.577152.
37. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11:1–9. https://doi.org/10.1038/nprot.2015.123.
38. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013;34:57–65. https://doi.org/10.1002/humu.22225.
39. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018;50:1161–70. https://doi.org/10.1038/s41588-018-0167-z.
40. Kim S, Jhong J-H, Lee J, Koo J-Y. Meta-analytic support vector machine for integrating multiple omics data. BioData Min. 2017;10:2. https://doi.org/10.1186/s13040-017-0126-8.
41. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99:877–85. https://doi.org/10.1016/j.ajhg.2016.08.016.
42. Cochrane G, Alako B, Amid C, Bower L, Cerdeño-Tárraga A, Cleland I, et al. Facing growth in the European Nucleotide Archive. Nucleic Acids Res. 2012;41:D30–5. https://doi.org/10.1093/nar/gks1175.
43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
44. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinforma Oxf Engl. 2023;39:btac830. https://doi.org/10.1093/bioinformatics/btac830.
45. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92. https://doi.org/10.1093/bib/bbs017.
46. Thorn CF, Klein TE, Altman RB. PharmGKB: The Pharmacogenomics Knowledge Base. In: Innocenti F, Van Schaik RHN, editors. Pharmacogenomics. Totowa, NJ: Humana; 2013. pp. 311–20. https://doi.org/10.1007/978-1-62703-435-7_20.

Publisher's note