

UC Davis

UC Davis Previously Published Works

Title

ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines.

Permalink

<https://escholarship.org/uc/item/0xq3j4jp>

Journal

Genome Medicine, 14(1)

Authors

Preston, Christine

Wright, Matt

Madhavrao, Rao

et al.

Publication Date

2022-01-18

DOI

10.1186/s13073-021-01004-8

Peer reviewed

SOFTWARE

Open Access



ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines

Christine G. Preston^{1†}, Matt W. Wright^{1†}, Rao Madhavrao¹, Steven M. Harrison², Jennifer L. Goldstein³, Xi Luo⁴, Hannah Wand⁵, Bryan Wulf¹, Gloria Cheung¹, Mark E. Mandell¹, Howard Tong¹, Shaung Cheng¹, Michael A. Iacocca¹, Arturo Lopez Pineda⁶, Alice B. Popejoy⁶, Karen Dalton⁷, Jimmy Zhen⁷, Selina S. Dwight⁸, Lawrence Babb², Marina DiStefano², Julianne M. O'Daniel³, Kristy Lee³, Erin R. Riggs⁹, Diane B. Zastrow¹⁰, Jessica L. Mester¹¹, Deborah I. Ritter⁴, Ronak Y. Patel¹², Sai Lakshmi Subramanian¹², Aleksander Milosavljevic¹², Jonathan S. Berg³, Heidi L. Rehm^{2,13}, Sharon E. Plon^{4,12}, J. Michael Cherry¹⁴, Carlos D. Bustamante^{6,14}, Helio A. Costa^{1,6*} and on behalf of the Clinical Genome Resource (ClinGen)

Abstract

Background: Identification of clinically significant genetic alterations involved in human disease has been dramatically accelerated by developments in next-generation sequencing technologies. However, the infrastructure and accessible comprehensive curation tools necessary for analyzing an individual patient genome and interpreting genetic variants to inform healthcare management have been lacking.

Results: Here we present the ClinGen Variant Curation Interface (VCI), a global open-source variant classification platform for supporting the application of evidence criteria and classification of variants based on the ACMG/AMP variant classification guidelines. The VCI is among a suite of tools developed by the NIH-funded Clinical Genome Resource (ClinGen) Consortium and supports an FDA-recognized human variant curation process. Essential to this is the ability to enable collaboration and peer review across ClinGen Expert Panels supporting users in comprehensively identifying, annotating, and sharing relevant evidence while making variant pathogenicity assertions. To facilitate evidence-based improvements in human variant classification, the VCI is publicly available to the genomics community. Navigation workflows support users providing guidance to comprehensively apply the ACMG/AMP evidence criteria and document provenance for asserting variant classifications.

* Correspondence: hcosta@stanford.edu

†Christine G. Preston and Matt W. Wright contributed equally to this work.

¹Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, MSOB x313, Stanford, CA 94305, USA

⁶Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: The VCI offers a central platform for clinical variant classification that fills a gap in the learning healthcare system, facilitates widespread adoption of standards for clinical curation, and is available at <https://curation.clinicalgenome.org>

Keywords: Variant curation, Precision medicine, Clinical genetics, Clinical Genome Resource Consortium

Background

The application of genomics to precision medicine holds great promise for the implementation of tailored diagnostics, optimized patient care management, and personalized therapies in healthcare. The past decade has seen the development of technological and computational innovations to bring both DNA-sequencing methodologies and bioinformatic algorithms into routine standard-of-care for diagnostic medical genomics. While there has been a broad consensus in terms of bioinformatics best practices, quality control metrics, and community adoption of variant calling and classification standards, substantial variability remains among variant curation tools and data sharing by health care providers, clinical diagnostic laboratories, and researchers.

Clinical interpretation of genomic sequencing data requires both the standardization of variant classification guidelines, as well as consistency in the workflow and evidence considered when determining the relationship between a variant and a disease phenotype. In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) released guidelines for the interpretation of germline genetic variants [1]. These germline variant curation guidelines have been broadly adopted by clinical genetic testing laboratories globally [2]. Additionally, the National Institutes of Health (NIH)-funded Clinical Genome Resource (ClinGen) Consortium [3] has further developed refined and standardized evaluation criteria of sequence variant pathogenicity [4–16]. Despite these efforts, the uniform adoption and application of these frameworks have proven challenging without robust computational infrastructure and curation software to consistently guide biocurators through these complex germline variant curation guidelines.

Here we present the ClinGen Variant Curation Interface (VCI), which is a comprehensive germline variant classification platform designed to support both individual and group classification in accordance with the ACMG/AMP germline classification guidelines. The VCI is intended to be a publicly available variant curation tool which programmatically guides users through a standard process for variant evidence classification and application of ACMG/AMP guidelines in a controlled workflow to enforce rigor and quality in germline variant classification (Fig. 1). The VCI aims to serve as a central platform for clinical variant classification that fills a gap

in the learning healthcare system and facilitates the widespread adoption of standards for clinical curation.

Implementation

The VCI curation platform has been developed to facilitate the Federal Drug Administration (FDA)-recognized ClinGen variant classification process, support transparent evidence review, and provide timely dissemination to the genomics community. Users can curate individually or communally in groups known as affiliations. The VCI programmatically displays relevant data types from external sources (Table 1) and displays evidence identified by other VCI users in an organized user interface enabling an environment to document ACMG/AMP criteria codes.

The core elements of the VCI data model are shown in Fig. 2. The full data model is stored in JavaScript Object Notation (JSON) format with references to data elements. The data model is centered upon a variant classification, with attributes consisting of data and context related to asserting the variant's pathogenicity. The classification model is based on the combined variant, disease, and mode of inheritance data models. Each variant is evaluated by biocurators against specific evidence types which are reflected in the VCI's data models (e.g., population data, experimental data, computational data), as well as literature-based evidence which can be manually added by a biocurator and related to any of the other evidence types. Biocurator-selected evaluations of the evidence criteria provide a computed variant pathogenicity, which can be manually overridden by biocurators based on expert opinion consistent with the ACMG/AMP guidelines.

The classification model is designed to support the ClinGen workflow of variant curation, which is an iterative and manual process, with biocurators making criteria evaluations for the assessment of genetic variant pathogenicity, and expert panels considering and approving final pathogenicity classifications. The platform models this process in two ways, first by marking all curations with the most recent status, progressing from "In-progress" (interpretations where the evidence is still being evaluated), "Provisional" (interpretations completed by the primary biocurator but awaiting expert approval) and "Approved" (interpretations which have been fully reviewed and classified). Should a revision to the evidence evaluation need to be made to an

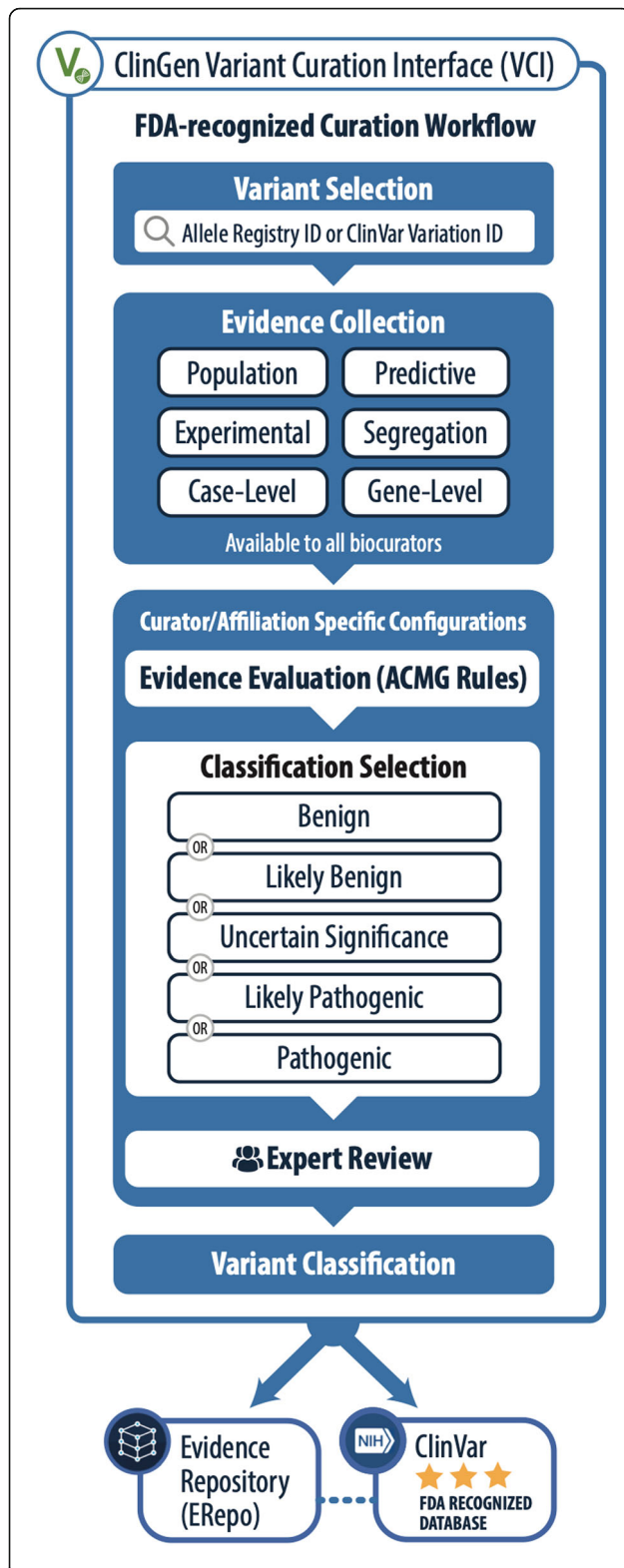


Fig. 1 ClinGen FDA-recognized variant curation process and VCI. Overview of the ClinGen variant curation process using the VCI, an FDA-recognized workflow. Biocurators select a variant and evaluate evidence that falls into six categories. VCI viewers may view all evidence available for any variant using the VCI. The VCI supports users in making a final pathogenicity classification keeping with the ACMG/AMP guidelines. ClinGen expert panels then disseminate their variant classifications through two community resources: the Evidence Repository (ERepo) and ClinVar

“Approved” interpretation the VCI will save those changes under the status “New-Provisional,” which will require a new approval. Approved classifications by ClinGen Variant Curation Expert Panels (VCEPs) may be sent to the ClinGen Evidence Repository (ERepo; <https://erepo.clinicalgenome.org/>) [26]. The ERepo is intended to provide access to variant level evidence used and applied by VCEPs in the classification of variants. Upon submission to the ERepo, approved classification will have the additional “Published” status appended. Secondly, the classification model only stores references to its related models (e.g., variant, disease, evidence, evidence criteria), which also store the most recent information. When a classification has a status of “Provisional,” “Approved,” or “Published,” the snapshot model is used to store an instance (point in time) containing all the related data. This allows changes to the related data to be identified, while previous data, which may have been used to make a criteria evaluation, are still preserved in the snapshot. Snapshots published to ERepo and the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) [27] undergo a transformation to the ClinGen Interpretation model (<http://dataexchange.clinicalgenome.org/interpretation/index.html>) [28], which aligns with a related community model; the Monarch SEPIO Framework (<https://github.com/monarch-initiative/SEPIO-ontology/wiki/SEPIO-Overview>) [29]. The SEPIO framework was chosen as it provides an ontology-based modeling framework which supports the scientific assertions and provides a structure for the evidence and provenance supporting those assertions.

The VCI is accessed through a web browser where users can perform curation activities including review of imported evidence, entry of evidence gathered from published and unpublished sources, ACMG/AMP criteria application, pathogenicity evaluation, and classification review and approval. The classified variants from ClinGen-approved VCEPs are then shared with the ERepo and ClinVar to enable peer review and public access.

The software for the VCI is freely and openly available in perpetuity via publicly accessible web pages and two publicly available GitHub repositories, one for the 1.0 legacy code (<https://github.com/ClinGen/clincoded/>) [30], and one for the current 2.0 codebase (<https://github.com/>

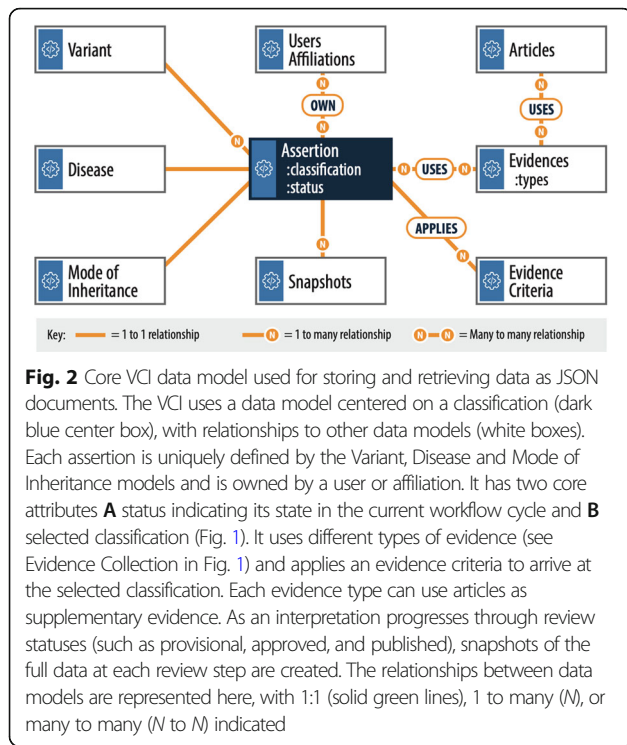
Table 1 VCI displayed information type and source

Information type	Displayed data	Data source
Basic Information	<ul style="list-style-type: none"> • Variant ID • HGVS terms 	ClinGen Allele Registry [17]
	<ul style="list-style-type: none"> • ClinVar Variation ID • ClinVar Overall interpretation • ClinVar Submitted interpretations • ClinVar Primary transcript • RefSeq transcripts • dbSNP variant ID • Entrez Gene ID 	NCBI E-utilities [18]
	<ul style="list-style-type: none"> • RefSeq transcripts • Ensembl transcripts • Molecular consequences 	Ensembl VEP [19]
	<ul style="list-style-type: none"> • Monarch Disease Ontology (Mondo) human disease term(s) • Phenotypic abnormality term(s) 	Ontology Lookup Service [20] Human Phenotype Ontology (HPO) [21]
Population	<ul style="list-style-type: none"> • Allele frequencies <ul style="list-style-type: none"> ◦ gnomAD ◦ ExAC ◦ Exome Sequencing Project 	MyVariant.info [22]
	<ul style="list-style-type: none"> • Allele frequencies <ul style="list-style-type: none"> ◦ PAGE Study 	GGV Browser [23]
	<ul style="list-style-type: none"> • Allele frequencies <ul style="list-style-type: none"> ◦ 1000 Genomes 	Ensembl VEP [19]
Variant Type	<ul style="list-style-type: none"> • <i>In silico</i> predictor scores <ul style="list-style-type: none"> ◦ REVEL ◦ SIFT • PolyPhen2 • LRT • MutationTaster • MutationAssessor <ul style="list-style-type: none"> ◦ FATHMM ◦ PROVEAN ◦ MetaSVM ◦ MetalR ◦ CADD ◦ FATHMM-MKL ◦ fitCons • Conservation analysis scores <ul style="list-style-type: none"> ◦ phyloP100way ◦ phyloP30way ◦ phastCons100way ◦ phastCons30way ◦ GERP++ ◦ SiPhy 	MyVariant.info [22]
Experimental	<ul style="list-style-type: none"> • Experimental functional data 	ClinGen Functional Data Repository (FDRepo)
Gene-Centric	<ul style="list-style-type: none"> • Gene symbol 	HGNC [24]
	<ul style="list-style-type: none"> • ExAC constraint scores • UniProt protein ID • GeneCards gene 	MyGene.info [25]

[ClinGen/gene-and-variant-curation-tools](https://curation.clinicalgenome.org)) (Table 2). The VCI website (<https://curation.clinicalgenome.org>) [31] is developed as a common interface for both the VCI and the related Gene Curation Interface (GCI), which is used to evaluate the strength of evidence that variation in a particular gene causes a particular disease. These two tools, VCI and GCI, use the same platform and share components such as a user database and classification data. User access to the VCI and GCI is available by authenticated

login. Login permissions are required to document the provenance of evidence added to the interfaces.

Users access the VCI web application via the browser and execute the workflow tasks needed to perform variant curation. The current deployment, VCI v2.0, utilizes cloud-based web development best practices and a “serverless architecture,” which is a cloud development approach where all application resource management and scaling needs are automatically determined and handled



by the cloud services. All the components essential for the application including authentication, gateways to receive and respond to browser requests, microservices, and storage are provided by Amazon Web Services (AWS). This scalable and robust architecture is based on several AWS serverless components shown in Fig. 3, the role of key components is further described here. The Application Programming Interface (API) Gateway handles tens of thousands of requests per second and provides automatic schema validation of data, ensuring data integrity in the VCI. Lambda spawns microservices to store or retrieve data, managing and scaling the computing resources required by the VCI. DynamoDB is a flexible, document-based database that provides constant load-independent performance, supporting the VCI in long-term goals of scaling to large numbers of variant classifications and providing bulk variant curation support, while Simple Storage Service (S3) stores the large VCI

database. Additionally, Cognito is used for user management, and Amplify for web content integration with backend microservices. The user interfaces are created using standard JavaScript programming (ReactJS), and they obtain information from the database via an API using a standard JSON format. For comparison, this continuous integration and deployment provide greater reliability and cost-savings relative to the initial deployment of the VCI v1.0 built following a classical three-tier architecture with a web-frontend component (ReactJS), backend business logic layer (Python and Pyramid), a split persistence layer containing the state and metadata database (PostgreSQL), and search indexes (AWS Elasticsearch).

Service components include an external resource manager, which is responsible for obtaining data from external sources (Table 1) that feed into the variant, gene, disease, population, predictive, functional, and gene data models.

Finally, ClinGen provides supplemental resources for VCI users including general information about biocuration, summary videos outlining the concepts and methods behind biocuration, and links to biocuration resources such as ClinGen’s documented standard operating procedures (<https://clinicalgenome.org/curation-activities/variant-pathogenicity/>) [32].

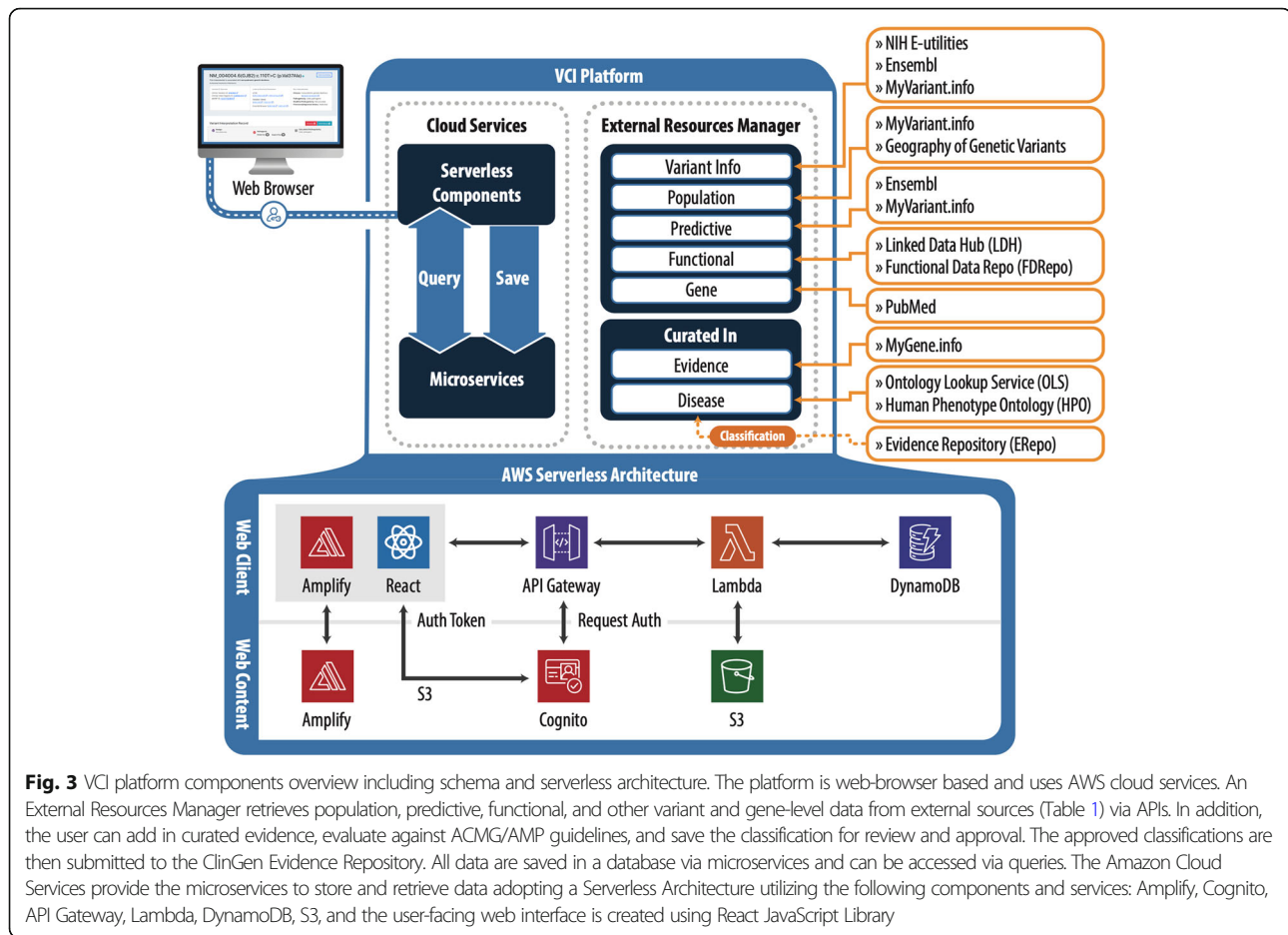
Development process

The VCI software and product development teams worked alongside the ClinGen Variant Curation Interface Task Team to develop the initial platform. This product was designed through a user engagement process and VCI v1.0 was launched for use in September 2016, with new features developed and released monthly. The completely re-architected and updated VCI v2.0 platform was launched in December 2020 and is the current production version.

VCI development continues to evolve with the input of core members of the ClinGen variant curation community that meet twice per month with the VCI development team. This group includes members of the ClinGen Sequence Variant Interpretation (SVI) Working Group, which provides guidance on how to interpret, refine, and standardize the ACMG/AMP guidelines [5–7,

Table 2 VCI software components

Component	Location	Description
Front-end	gci-vci-react	Contains all front-end code used for user-interface development
Back-end	gci-vci-serverless	All back-end code including controllers, and database access objects
Database models	gci-vci-serverless/src/models/	This is a set of models that are used for validation of the document data posted to the dynamodb database
API development	Gci-vci-api	Added in August 2021 to provide API support for GCI/VCI data
Messaging	gci-vci-kafka-to-lambda gci-vci-serverless/src/helpers/message_helpers.py	The messaging component to exchange data with other ClinGen tools



11, 16], and members of ClinGen’s VCEPs. Additional guidance for VCI development comes from the ClinGen Data Access, Protection and Confidentiality (DAPC) Working Group, which reviews tools and data practices in the ClinGen curation ecosystem to ensure that software development efforts are informed by updated data sharing policies. Detailed best practice recommendations for biocurator use of the VCI and associated resources for variant classification are provided in the ClinGen Variant Curation Standard Operating Procedure.

Variant identification and evidence

It is possible to define a variant in several different ways. This promiscuity arises because of the availability of multiple transcripts and genomic reference sequences and various ways to describe insertion/deletions. As a result, unambiguous identification of variants is critical to the downstream usability of the curated data. The VCI identifies a variant by either a ClinVar Variation ID [33] or a ClinGen Allele Registry ID [17]. ClinVar IDs are assigned to each set of submitted variants, generally resulting in a single variant being associated with a single ID, ClinVar does support two subclasses of IDs,

allowing IDs for variants being directly interpreted and those being interpreted in the context of a set of variants. ClinGen’s allele registry provides a globally unique “canonical” variant identifier (CAid) on demand for variants. This enables aggregation of variant information from different sources [17]. The variant is associated with a Mondo Disease Ontology [34] term and a Human Phenotype Ontology [21] mode of inheritance term by the biocurator or VCEP. Recognizing that a variant can be curated for more than one potential disease, each user is restricted to one variant classification record per variant. Within the interfaces each variant is titled based on a hierarchy of naming conventions, preferentially using a title based on the Matched Annotation from NCBI and EMBL-EBI (MANE) Select [35] transcript when available (Fig. 4A).

The VCI aggregates and displays multiple types of evidence about a variant, separated into six tabs structured by data type, providing a rich and structured evidence gathering experience to biocurators, while supporting variant classification in accordance with the ACMG/AMP guidelines. This promotes consistency in terms of the evidence evaluated, application of the ACMG/AMP

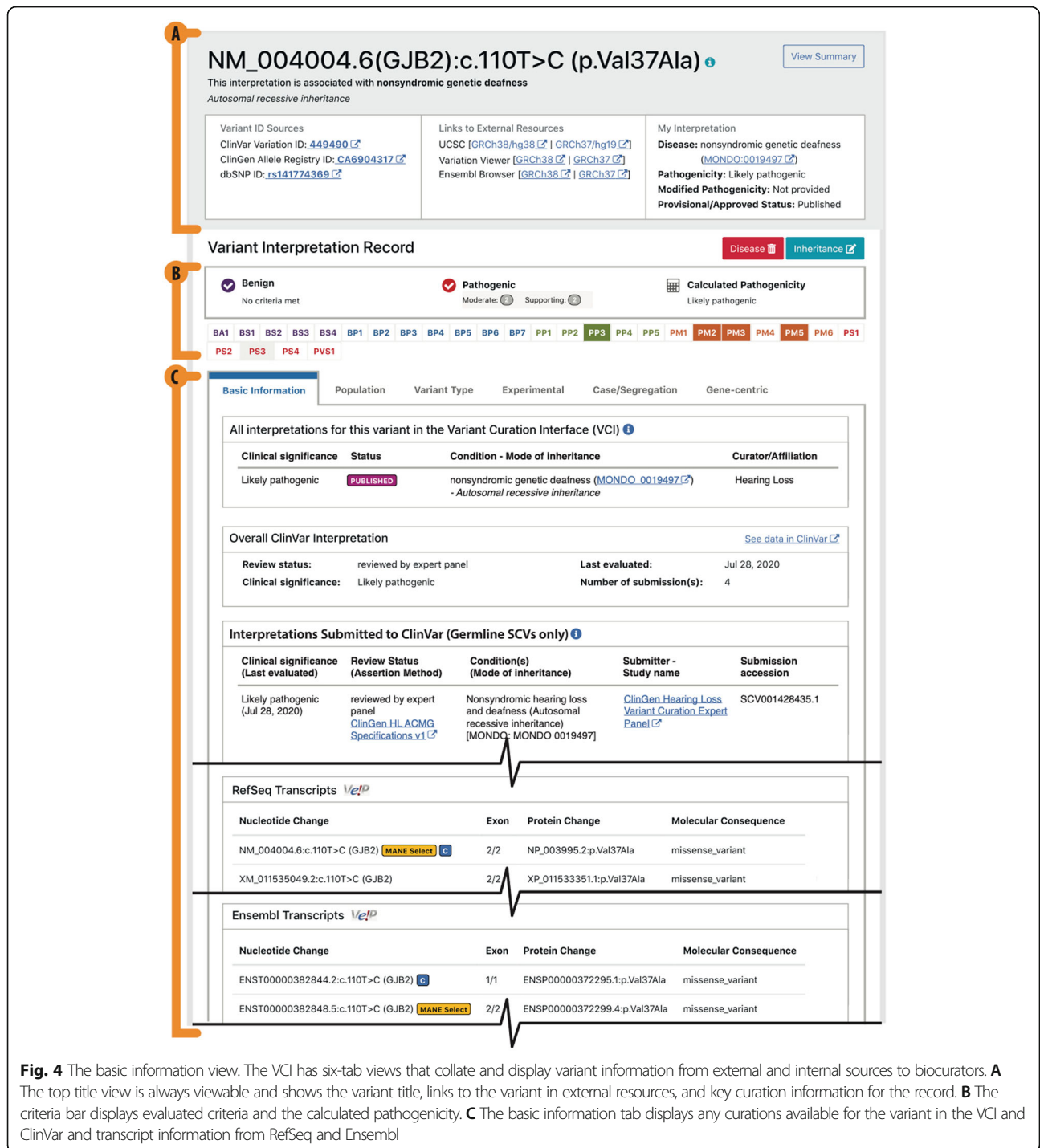


Fig. 4 The basic information view. The VCI has six-tab views that collate and display variant information from external and internal sources to biocurators. **A** The top title view is always viewable and shows the variant title, links to the variant in external resources, and key curation information for the record. **B** The criteria bar displays evaluated criteria and the calculated pathogenicity. **C** The basic information tab displays any curations available for the variant in the VCI and ClinVar and transcript information from RefSeq and Ensembl

criteria, and pathogenicity calculations. In keeping with the ClinGen goal to support appropriate community data sharing, all evidence added by users is viewable by any other VCI user. While all evidence is viewable by all users, a user’s evidence evaluations and pathogenicity calculations remain private until the classification record has been finalized (set to “Approved” status), at which

point other users can view, but not edit the final classification in the VCI.

Automated evidence

The VCI programmatically retrieves and displays many different types of evidence for each variant (Table 1). This includes the many possible variant nomenclatures

on different transcripts and human genome builds, population frequency data, in silico prediction scores, conservation data, gene and protein resources, and all classifications and submissions for the variant currently present in ClinVar or the VCI. When evidence is unavailable for direct display via API, dynamic links to external information sources are embedded within the relevant evidence tab. Relevant external information sources are identified in conjunction with core members of the ClinGen variant curation community described above.

Manually curated evidence and structured data capture

Users can manually add information relevant to the variant being classified from published articles for any evidence type into the relevant sections in the VCI. Additionally, structured data capture is supported for published functional data and for published and unpublished case and segregation evidence. Such structured data inputs ensure curation consistency, as they organize and accurately define the information so that it can easily be retrieved, facilitate searching of the captured data, and enable downstream data processing applications such as data mining and machine learning. Two examples of structured data capture within the VCI are outlined below (Functional data capture and Case level data capture).

Functional data capture

The ACMG/AMP guidelines [1] require the assessment of well-established in vivo or in vitro functional studies showing “no damaging effect” (BS3) or “supportive of damaging effect” (PS3) on protein function or splicing. We have developed a structured framework with the narrative of (1) method, (2) material, and (3) effect (with or without a quantifier), using standardized terminology from ontologies, for users to define the functional data they have derived from published articles in a consistent and reproducible way. We provide users with a standardized template for capturing these structured data which they can then submit to ClinGen’s Functional Data Repository (FDRRepo [<https://ldh.genome.network/fdr/ui/>]) [36]; subsequently, these granular functional data for each variant are viewable in the “Experimental” tab in the VCI. Future enhancements will include updating the structured narrative and data fields in accordance with new standards [7] and augmenting capacity for bulk annotations to be imported from literature annotations or databases of functional evidence including the increasing availability of data from multiplex assays of variant effect [37, 38].

Case level data capture

Case level and segregation level data are critical to pathogenicity evaluations using the ACMG/AMP

guidelines. However, individual case observations have the potential to be linked to individual patient identities if enough ancillary information is also included about the case. It is for this reason that the case segregation tab of the VCI prompts users to remember the Terms of Use for this tool, which include prohibitions on entering protected health information (PHI) or other sensitive information that could possibly identify an individual data subject and entering only the minimum necessary information to resolve a particular case. To further protect data subjects from the possibility of re-identification (which is also strictly prohibited for users of the VCI as stated in the Terms of Use), individual case-level data are not made publicly available through ClinGen tools except in aggregate. When entering in case observations or pedigree segregation evidence, users are directed to a form that has separate fields to capture each distinct case-level observation or co-segregation event. Then, individual counts for each distinct data type are summed together and analyzed in aggregate along with the same information from other evidence sources (Fig. 5B).

Curation workflow

ClinGen variant curation through the VCI enables the use of the nomenclature, criteria codes, and rules defined in the joint 2015 ACMG/AMP guidelines on variant classification [1]. Embedded flexibility is designed to allow biocurators to incorporate modifications and additional guidance produced by ClinGen’s Sequence Variant Interpretation WG as well as disease specifications from ClinGen VCEPs following the FDA-recognized validation process. To further aid this process, the VCI allows groups of users to curate variants as a single entity, known as an “affiliation.” VCI affiliations are often ClinGen VCEPs [39, 40]; however, any group of users who wish to curate variants together (e.g., a clinical or research laboratory) may form an affiliation. Once a VCI user initiates a classification, it belongs to that individual or affiliation and can only be edited by them.

The ACMG/AMP guidelines provide a set of criteria to be considered when classifying a variant. The VCI is designed to help users evaluate the applicability of these criteria in an efficient and structured way (Fig. 1). The VCI groups criteria by evidence types: displaying both the relevant criteria and any related evidence. These groups are (1) Population (known variant allele frequencies), (2) Variant Type (predicted impact of the variant on the gene product), (3) Experimental (functional assay data), and (4) Case/Segregation (relevant observations of the variant). The VCI also groups together gene-focused resource links, and basic information, displaying ClinVar and VCI curations for the variant as well as the molecular consequence of the variant on all known transcripts (Fig. 4C and Table 1).

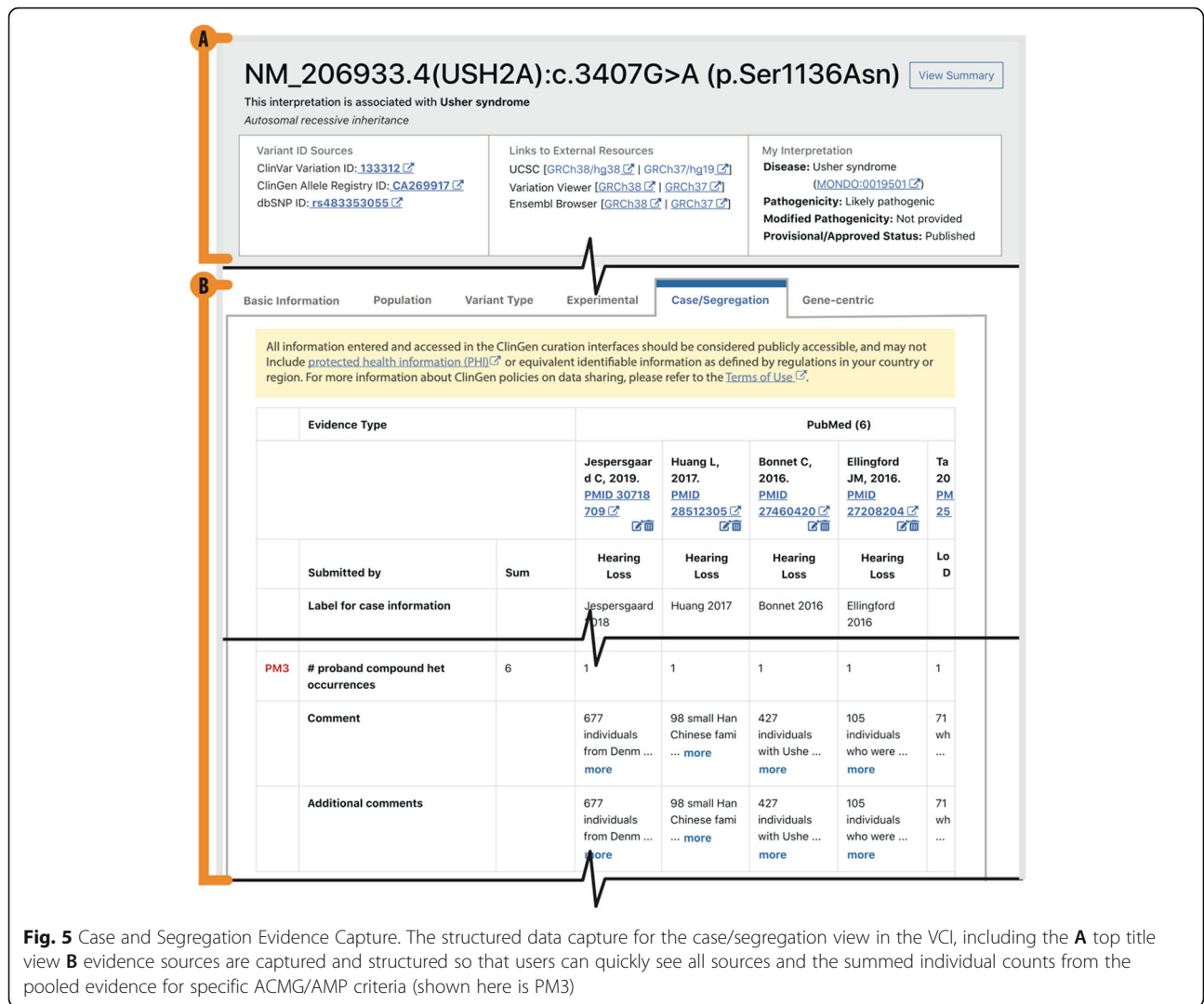


Fig. 5 Case and Segregation Evidence Capture. The structured data capture for the case/segregation view in the VCI, including the **A** top title view **B** evidence sources are captured and structured so that users can quickly see all sources and the summed individual counts from the pooled evidence for specific ACMG/AMP criteria (shown here is PM3)

For each ACMG/AMP criteria, users are provided a description of the guideline in the VCI. The user can view, add, and evaluate the relevant evidence and then set their criteria evaluation and write an explanation. As some criteria are applicable at different pathogenicity strengths, users can choose the appropriate strength from a pulldown list containing only the appropriate strength options for that criterion. For instance, the available options provided for evaluating PP1 are Not Evaluated, Met, Not Met, PP1_Moderate, and PP1_Strong. As a user saves their evaluations, a criteria bar (Fig. 4B) in the header of the interface keeps track of their progress by indicating which criteria have been “Met” (solid color background with white criteria code), “Not Met” (gray background with colored criteria code) or remain “Not Evaluated” (white background with colored criteria code). If a user scrolls over individual criteria codes in this bar, they will see a description for each criterion and they can click on individual criteria

codes to link to the pertinent section in the VCI. Additionally, a progress bar shows the number of criteria met according to the strength of the evaluation and whether they are “Benign” or “Pathogenic” and automatically calculates the pathogenicity each time an evaluation is saved or updated. This auto-classification is based on the default guidelines for weighing and combining the Pathogenic and Benign evaluated ACMG/AMP criteria as laid out in Richards et al. [1]. At any time, a user can view an “Evaluation Summary” that summarizes all their evaluated evidence. If a criterion code is not evaluated, then it would not be considered in the calculation of a predicted classification. Once a bio-curator is satisfied, they have reviewed all pertinent evidence and evaluated all relevant criteria; they can save their classification as “Provisional”. This generates a PDF version of the “Evaluation Summary” that can be distributed among the VCEP membership domain experts to aid in their review process. Upon satisfactory completion

of the review process, a final classification can be saved as “Approved,” at which point the “Evaluation Summary” can now be viewed by all VCI users.

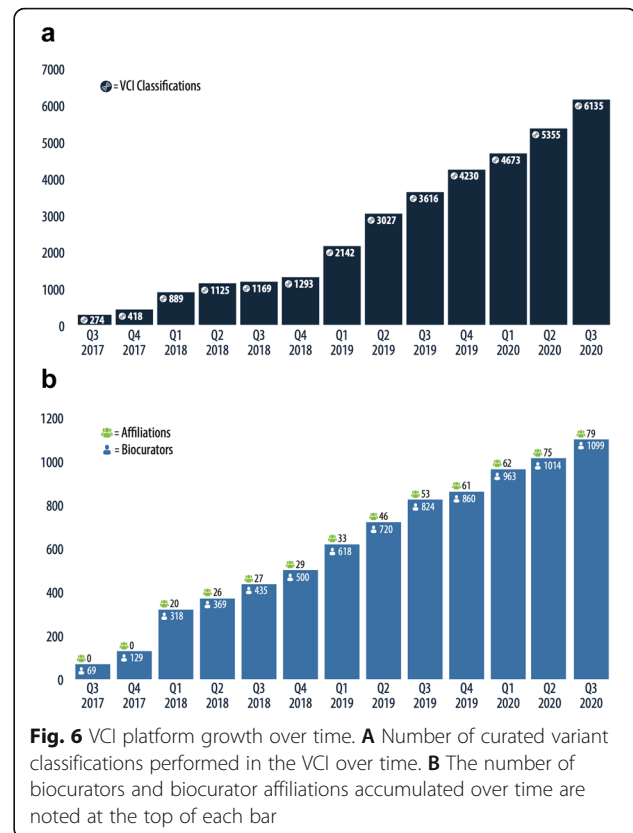
FDA recognition and data dissemination

The VCI generates an output file of the final variant pathogenicity classification in an auto-generated format compatible with ClinVar submission specifications. This is intended to facilitate timely dissemination of variant classifications to the genomics community via ClinVar and is a requirement for ClinGen VCEPs. The ultimate goal is to support fully automated, API-based ClinVar submission through the VCI once ClinVar provides support for API-based submission. Once submitted, a “submission to ClinVar” (SCV) identifier is obtained and can be viewed in the VCI variant record.

The ClinGen variant curation process was recognized by the FDA in December 2018 [41] and is followed by all ClinGen VCEPs. The evidence curation data and pathogenicity classifications generated within the VCI by ClinGen VCEPs are therefore considered to be valid scientific evidence that can be used to streamline the test development and validation processes. As such, additional steps and requirements apply to the information specifically generated through the ClinGen VCEP variant curation and classification process. Specifically, all the evidence that has been curated and evaluated, along with provenance should be made publicly available and easily accessible. With this in mind, the VCI saves all evidence that is evaluated by its users. In addition, upon final approval of a classification from a ClinGen VCEP, the VCI facilitates data flow to the ClinGen Evidence Repository (ERepo), where the finalized classifications and associated evidence evaluation is published. Importantly, the VCEP generated variant record in the ERepo includes comments for specific codes enabling in-depth and transparent data for peer review (Fig. 1). These publicly accessible displays of the final ClinGen VCEP variant classifications are accessed via the ERepo API at <https://erepo.clinicalgenome.org/evrepo/> [26].

Current status and users

VCI currently has over 1100 registered users, two thirds of whom are members of ClinGen VCEPs (Fig. 6). The VCI is publicly accessible (with registration) for variant curation. When curating together as an affiliation, all members can view and edit all information added by anyone in that affiliation. This popular feature is currently used by 79 registered affiliations, most of which represent official ClinGen VCEPs but also include other affiliations of ClinGen members (e.g., institution-specific biocurator teams) and groups unrelated to ClinGen (e.g., clinical and research laboratories).



Results and discussion

Here we present the development of a genetic variant biocuration platform for health care providers, researchers, and the medical genetics community to determine which gene variants are causal for a disease. The VCI supports the FDA-recognized ClinGen variant curation process and combines clinical, genetic, population, and functional evidence with expert review to classify variants into ACMG/AMP 2015 variant classification guideline categories [1]. Primary features of the VCI include the ability to (1) curate individually or in groups, (2) associate pertinent evidence with variant classifications, (3) allow users to assess evidence per variant curation disease/gene-specific protocols, (4) enable users to save provisional records, (5) support an expert review process of curated evidence, and (6) automatically publish classifications and underlying criteria assessments to the ERepo.

Future VCI improvements will focus on enhancing scale, workflow, throughput, and support ongoing compliance with FDA recognition of the ClinGen Variant Curation Expert Panels through the FDA Human Variant Database program. The current VCI v2.0 platform has over 6300 variant classifications in different curation stages, and our modernized architecture is able to scale to support over 1 million future classifications. We plan to enhance workflow usability and curation efficiency by

making the platform more proactive with (1) task management (supporting assigning of variant classification records to users) and action items (alerts to users), (2) support for bulk variant curation workflows, (3) automatically bring in additional variant evidence data and monitor major data changes via a streaming service so curations can be updated as needed, and (4) provide customized curation experiences based on VCEP specifications. To ensure compliance with FDA requirements for the auditability of ClinGen variant curation process, the VCI database maintains a complete audit trail of all saved curation actions. We will further support FDA compliance with (1) additional traceability, (2) permanent archiving, (3) regular knowledge updating through literature and database monitoring, and (4) update alerts provided to curation teams.

Conclusions

The VCI provides needed software infrastructure and a comprehensive curation platform necessary for supporting variant classification, a critical step in the use of genomics in medicine. This global open-source platform aids individual biocurators and teams of collaborating biocurators in performing the complex task of variant curation in an efficient workflow to enforce rigor and quality in variant classification ultimately contributing to scientific advancement and informing health care management.

Acknowledgements

We would like to thank all the members of the Clinical Genome Resource consortium, especially the core members attending bimonthly VCI feedback meetings for their continual feedback.

Authors' contributions

All authors contributed to the project design. C.G.P., M.W.W., R.M., and H.A.C. drafted the initial version of the manuscript. The authors contributed to and approved the final version of the manuscript.

Funding

This research was funded by grants from the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (U41HG009649, U01HG007436, U41HG006834, U01HG007434, U41HG009650, 2U24HG009649-05).

Availability of data and materials

Project name: ClinGen Variant Curation Interface
 Project home page: <https://curation.clinicalgenome.org> [31]
 Operating System: Platform independent
 Programming languages: JavaScript, Python
 Other requirements: none
 License: MIT Open-Source
 All code for the VCI including front-end, back-end, database schemas, analysis pipelines, and user interfaces are freely available under MIT Open-Source licenses via the GitHub repositories (The VCI 1.0 code is available at, <https://github.com/ClinGen/clincoded/> [30], while the VCI 2.0 code can be found at <https://github.com/ClinGen/gene-and-variant-curation-tools> [42]. Extensive documentation for the usage of the VCI, including links to video tutorials, detailed explanations of all major features, and screenshots are available at <https://github.com/ClinGen/clincoded/wiki/VCI-Curation-Help> [43]. Additional training modules on ClinGen Variant Curation, including the ClinGen Standard Operating Procedure for Variant Curation, and links to

publications detailing standards and recommendations for using the ACMG/AMP criteria can be found at <https://clinicalgenome.org/curation-activities/variant-pathogenicity/training-materials/> [32].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

S.E.P. is a member of the Baylor Genetics Scientific Advisory Panel. A.M. is an employee of Baylor College of Medicine (BCM) and performs integration consulting services for BCM-developed software including Genboree through IP Genesis Inc. C.D.B. is on the scientific advisory boards (SAB) of AncestryDNA, Arc Bio LLC, Etalon DX, Liberty Biosecurity, and Personalis. C.D.B. is on the board of EdenRoc Sciences LLC. C.D.B. is also a founder and SAB chair of ARCBio. J.L.M. is an employee of GeneDx/BioReference Laboratories, Inc./OPKO Health and has a salary as the only disclosure. None of these entities played a role in the design, execution, interpretation, or presentation of this study. The remaining authors declare that they have no competing interests.

Author details

¹Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, MSOB x313, Stanford, CA 94305, USA. ²Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ³Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. ⁴Department of Pediatrics/Hematology-Oncology, Baylor College of Medicine, Houston, TX 77030, USA. ⁵Center for Inherited Cardiovascular Disease, Stanford Health Care, Stanford, CA 94305, USA. ⁶Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁷Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁸Grace Science LLC, Menlo Park, CA 94025, USA. ⁹Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA 17837, USA. ¹⁰Sutter Health, Mountain View, CA 94040, USA. ¹¹GeneDx Inc., Gaithersburg, MD 20877, USA. ¹²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. ¹³Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA. ¹⁴Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA.

Received: 13 February 2021 Accepted: 12 November 2021

Published online: 18 January 2022

References

- Richards S, 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(5):405–24. <https://doi.org/10.1038/gim.2015.30>.
- Niehaus A, Azzariti DR. A survey assessing adoption of the ACMG-AMP guidelines for interpreting sequence variants and identification of areas for continued improvement. *Genet Med*. 2019;21(8):1699–701. <https://doi.org/10.1038/s41436-018-0432-7>.
- Rehm HL, Berg JS. ClinGen—the clinical genome resource. *N Engl J Med*. 2015;372(23):2235–42. <https://doi.org/10.1056/NEJMs1406261>.
- Amendola LM, Jarvik GP. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research consortium. *Am J Hum Genet*. 2016;98(6):1067–76. <https://doi.org/10.1016/j.ajhg.2016.03.024>.
- Abou Tayoun AN, Pesaran T. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018;39(11):1517–24. <https://doi.org/10.1002/humu.23626>.
- Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med*. 2018;20:1687–8.
- Briich SE, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12:3.

8. Luo X, Feurstein S. ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv.* 2019; 3(20):2962–79. <https://doi.org/10.1182/bloodadvances.2019000644>.
9. Lee K, Krempely K. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum Mutat.* 2018;39(11):1553–68. <https://doi.org/10.1002/humu.23650>.
10. Oza AM, Di Stefano MT. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat.* 2018;39(11): 1593–613. <https://doi.org/10.1002/humu.23630>.
11. Ghosh R, Harrison SM. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat.* 2018;39(11):1525–30. <https://doi.org/10.1002/humu.23642>.
12. Mester JL, Ghosh R. Gene-specific criteria for PTEN variant curation: recommendations from the ClinGen PTEN Expert Panel. *Hum Mutat.* 2018; 39(11):1581–92. <https://doi.org/10.1002/humu.23636>.
13. Zastrow DB, Baudet H. Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): the ClinGen IEM Working Group and the Phenylalanine Hydroxylase Gene. *Human Mutat.* 2018;39(11):1569–80. <https://doi.org/10.1002/humu.23649>.
14. Gelb BD, et al. ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genet Med.* 2018;20(11):1334–45. <https://doi.org/10.1038/gim.2018.3>.
15. Kelly MA, et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med.* 2018;20(3):351–9. <https://doi.org/10.1038/gim.2017.218>.
16. Tavtigian SV, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med.* 2018;20(9):1054–60. <https://doi.org/10.1038/gim.2017.210>.
17. Pawliczek P, et al. ClinGen Allele Registry links information about genetic variants. *Hum Mutat.* 2018;39:1690–701.
18. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2018;46:D8–D13.
19. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol.* 2016;17(1):122. <https://doi.org/10.1186/s13059-016-0974-4>.
20. Madeira F, Park Y, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47(W1):W636–41. <https://doi.org/10.1093/nar/gkz268>.
21. Köhler S, Carmody L. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* 2019;47(D1):D1018–27. <https://doi.org/10.1093/nar/gky1105>.
22. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. *Genome Biol.* 2016;17(1):91. <https://doi.org/10.1186/s13059-016-0953-9>.
23. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics.* 2017;33:594–5.
24. Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* 2019;47(D1): D786–92. <https://doi.org/10.1093/nar/gky930>.
25. Wu C, Macleod I, Su AI. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.* 2013;41:D561–5.
26. ClinGen Evidence Repository. <https://erepo.clinicalgenome.org/>. Accessed 08 Nov 2021.
27. ClinVar Database. <https://www.ncbi.nlm.nih.gov/clinvar/>. Accessed 08 Nov 2021.
28. ClinGen Data Exchange Interpretation Model. <http://dataexchange.clinicalgenome.org/interpretation/index.html>. Accessed 08 Nov 2021.
29. Monarch SEPIO Framework. <https://github.com/monarch-initiative/SEPIO-ontology/wiki/SEPIO-Overview>. Accessed 08 Nov 2021.
30. Variant Curation Interface GitHub Repository (v1). <https://github.com/ClinGen/clincode>. Accessed 08 Nov 2021.
31. Variant Curation Interface Website. <https://curation.clinicalgenome.org/>. Accessed 08 Nov 2021.
32. ClinGen Biocuration Standard Operating Procedures. <https://clinicalgenome.org/curation-activities/variant-pathogenicity/>. Accessed 08 Nov 2021.
33. Landrum MJ, Lee JM. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(D1): D980–5. <https://doi.org/10.1093/nar/gkt1113>.
34. Shefchek KA, Harris NL. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* 2020;48(D1):D704–15. <https://doi.org/10.1093/nar/gkz997>.
35. Matched Annotation from NCBI and EMBL-EBI (MANE). <https://www.ncbi.nlm.nih.gov/refseq/MANE/>.
36. ClinGen Functional Data Repository. <https://ldh.genome.network/fdr/ui/>. Accessed 08 Nov 2021.
37. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, et al. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol.* 2019;20(1):223. <https://doi.org/10.1186/s13059-019-1845-6>.
38. Gelman H, et al. Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. *Genome Med.* 2019;11(1):85. <https://doi.org/10.1186/s13073-019-0698-7>.
39. Harrison SM, Biesecker LG, Rehm HL. Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. *Curr Protoc Hum Genet.* 2019; 103(1):e93. <https://doi.org/10.1002/cphg.93>.
40. Rivera-Muñoz EA, 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.
41. Office of the Commissioner. FDA takes new action to advance the development of reliable and beneficial genetic tests that can improve patient care. <https://www.fda.gov/news-events/press-announcements/fda-takes-new-action-advance-development-reliable-and-beneficial-genetic-tests-can-improve-patient> (2018).
42. Variant Curation Interface GitHub Repository (v2). <https://github.com/ClinGen/gene-and-variant-curation-tools>. Accessed 08 Nov 2021.
43. Variant Curation Interface GitHub Documentation and Tutorials. <https://github.com/ClinGen/clincode/wiki/VCI-Curation-Help>. Accessed 08 Nov 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

