Building the ClinGen Pathogenicity Calculator Version 2.0 by leveraging ClinGen API Microservices.

K. Riehle[1], A. R. Jackson[1], N. Shah[1], A. A. Zuniga[1], T. Farris[1], A. Kodochygov[1], K. Yu[1], A. Mihajlović;[2], B. Jevtić;[2], M. Milinkov[2], N. Vuč;inić;[2], D. Jevtić;[2], N. Martinović;[2], M. Wright[3], C. Preston[3], M. Mandell[3], G. Cheung[3], D. Ritter[1], J. Arce[1], V. Dang[1], G. Noble[1], S. Harrison[4], A. Byrne[5], D. Azzariti[5], E. Riggs[6], M. DiStefano[5], L. Babb[5], K. Ferriter[5], E. McCormick[7], X. Gai[8], L. Shen[8], M. Falk[7], A. Wagner[9], J. Arbesfeld[9], M. Cline[10], A. F. Rubin[11], O. Griffith[12], M. Griffith[12], T. Klein[3], S. Plon[1], A. Milosavljevic[1], Clinical Genome Resource

[1]Baylor Coll. of Med., Houston, TX, [2]Persida Inc., Brooklyn, NY, [3]Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Ambry Genetics, Salt Lake City, UT, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Geisinger, Huntsville, AL, [7]Children's Hosp. of Philadelphia, Philadelphia, PA, [8]CHILDREN S Hosp. OF LOS ANGELES, Los Angeles, CA,, [9]Nationwide Children's Hosp., Columbus, OH, [10]UC Santa Cruz, Santa Cruz, CA, [11]WEHI, Melbourne, Australia, [12]Washington Univ. Sch. of Med., St Louis

Introduction

The Clinical Genome Resource (ClinGen) suite of microservices provides a platform to build applications within and outside of ClinGen by leveraging Application Programming Interfaces (APIs). The ClinGen APIs and messaging queues have been employed within ClinGen to support variant curation, which includes data aggregation, curation, and dissemination of published pathogenicity assertions. Version 1.0 of the Pathogenicity Calculator was aimed at non-ClinGen users, supported ACMG/AMP v3.0 2015 guidelines, and was utilized as a classic siloed application. Version 2.0 is upgraded to support an early draft of the ACMG/AMP/CAP/ClinGen SVC v4.0 2023 guidelines and leverages ClinGen API microservices, including the ClinGen Allele Registry, ClinGen Linked Data Hub, and ClinGen Criteria Specification Registry. In addition to the utility to the end-users, Version 2.0 thus provides a new model for how interfaces and workflows within and outside of ClinGen may leverage the ClinGen Resource.

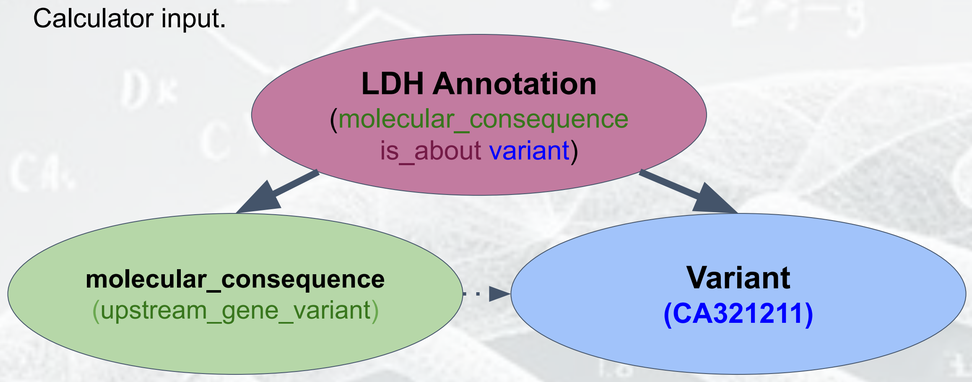
Leveraging Microservices

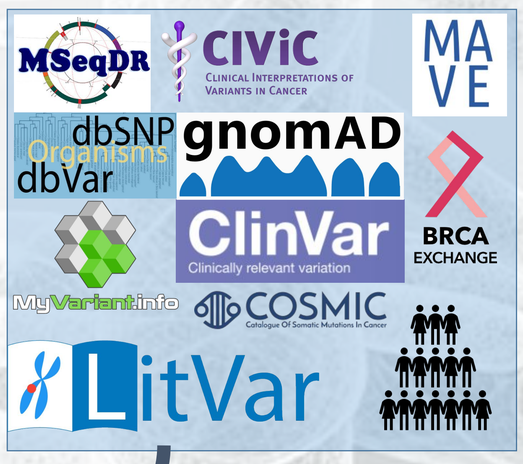
ClinGen API microservices provide for ample flexibility for designing pipelines and workflows to leverage specific functionalities of the ClinGen Data Platform suite of services, which are not confined to ClinGen applications. Here we describe the design of a newly architected pathogenicity calculator that transitions away from a siloed approach towards leveraging individual microservices accessible via API. Individually, APIs are leveraged towards populating variants (via API PUT) in the Allele Registry, supporting evidence in the Linked Data Hub, and guidelines in the Criteria Specification Editor. For the newly re-architected ClinGen Pathogenicity Calculator, APIs are invoked (via API GET / POST) on creation of a new interpretation to access canonical alleles in the Allele Registry and aggregated variant level supporting evidence data in the Linked Data Hub. Upon selecting a guideline for the variant of interest from available Criteria Specifications (suggested from condition, mode of inheritance, and variant-gene association), users are guided in filling out met evidence codes from collected supporting evidence. This flexible design approach encourages re-use from additional applications and provides good footing for future developments in this field.

ClinGen Linked Data Hub

https://ldh.clinicalgenome.org

Supporting evidence is obtained and aggregated in the form of excerpts linked to a given a canonical allele identifier (CA ID) from the Allele Registry in the Linked Data Hub, which facilitates open sharing of linked data from external sources and automated calculation of some evidence codes as Calculator input.

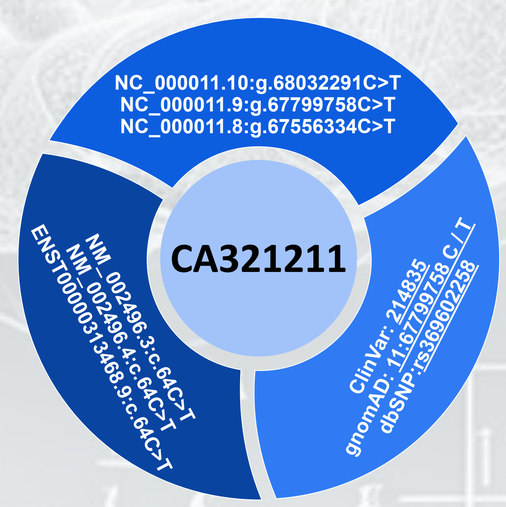




ClinGen Allele Registry

https://reg.clinicalgenome.org

Calculator 2.0 users first identify a variant of interest, uniquely identified by a stable canonical allele identifier (CA ID) obtained from free, open, on-demand registration within the Allele Registry. The Allele Registry contains over 2.57B registered variants contributed from users and external databases (e.g. ClinVar, gnomAD, dbSNP, etc.). We actively support both a user interface for individual and small batches of variants and an API for bulk query and registration (100 - 100,000 variants / second) (http://tinyurl.com/car-guide).

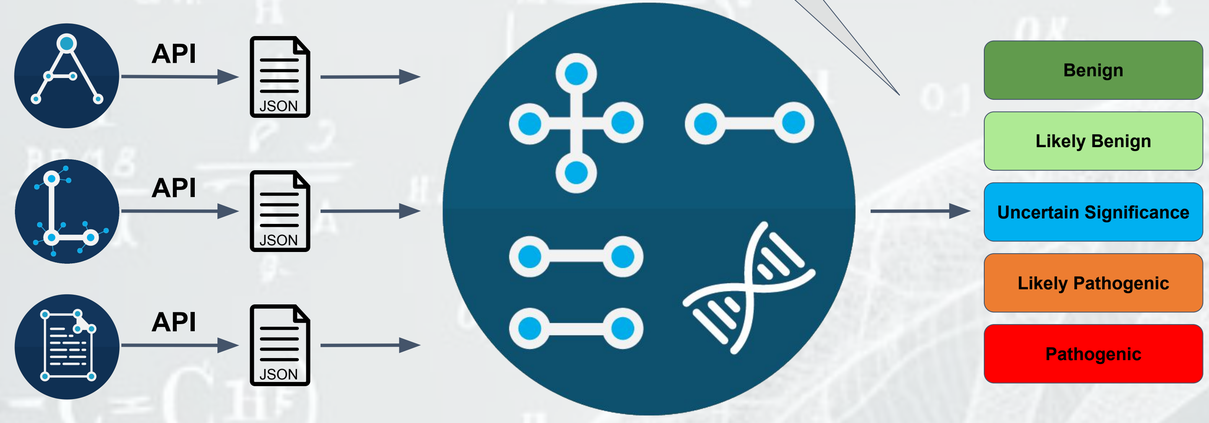


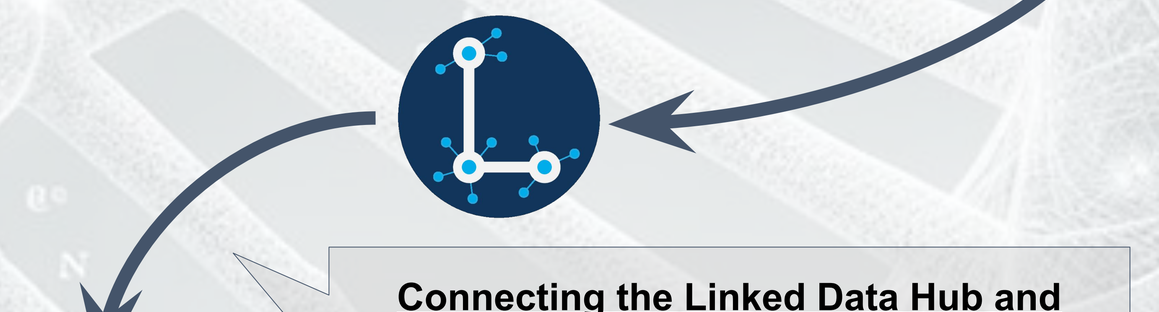
Connecting the Linked Data Hub and Criteria Specification Registry

- Connecting evidence properties in speciations with supporting evidence from linked data allows for:
  - Transparent audit trail history from versioned documents
  - Automated (suggested) calculation of individual codes

Pathogenicity Calculation

- Calculation is computed using the CSpec Reasoner Engine which provides consistent and repeatable calculations
- Reasoner Engine is API accessible and accepts evidence and specification as input





ClinGen Criteria Specification Registry

https://cspec.clinicalgenome.org

ClinGen variant curation expert panel (VCEP) specifications are created, edited, and approved following the ClinGen expert panel process within the Criteria Specification Editor and are publicly-accessible via the Criteria Specification Registry. Criteria specifications are gene and disease specific structured data developed from a combination of empirical analysis and expert recommendations. Calculator 2.0 can use the tailored VCEP-specific gene-condition criteria specifications from the Criteria Specification Registry with a flexible user interface to solicit the required user input and calculate variant pathogenicity scores.

Conclusion

https://calculator.clinicalgenome.org

In conclusion, Version 2.0 of the ClinGen Pathogenicity Calculator will exemplify the power of ClinGen API microservices to serve as a platform for rapid development of new applications within and beyond ClinGen. This microservice-oriented model for interface development is particularly relevant for established variant curation and interpretation workflows outside of ClinGen, as they may tap into ClinGen API microservices to accelerate development, leverage ClinGen knowledge within their workflows, and adopt the latest professional guidelines. We anticipate release of the newly re-designed calculator supporting ACMG/AMP v3.0 2015 guidelines to be released as a beta version by the end of 2023. Please utilize the above URL for updates or email Kevin Riehle (riehle@bcm.edu).

Acknowledgement: This work is supported by the Grants ***[U41HG006834], [U41HG009649], [U41HG009650]*** from the National Human Genome Research Institute (NHGRI).