# HHS Public Access

# ClinGen Advancing Genomic Data-Sharing Standards as a GA4GH Driver Project

**Lena Dolman**[*][1,2], **Angela Page**[2,3], **Lawrence Babb**[4], **Robert R. Freimuth**[2,5], **Harindra Arachchi**[2,3,6], **Chris Bizon**[7], **Matthew Brush**[2,8], **Marc Fiume**[1,2,9], **Melissa Haendel**[2,8,10], **David Hansen**[2,11], **Aleksandar Milosavljevic**[12], **Ronak Y. Patel**[12], **Piotr Pawliczek**[12], **Andrew D. Yates**[2,13], and **Heidi L. Rehm**[2,3,14]

[1]Ontario Institute for Cancer Research, Toronto, Ontario, Canada

[2]Global Alliance for Genomics and Health (GA4GH), Toronto, Ontario, Canada

[3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[4]Sunquest Information Systems, Boston, Massachusetts, USA

[5]Center for Individualized Medicine, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA 55905

[6]Center for Mendelian Genomics, The Broad Institute of MIT and Harvard, Cambridge, MA

[7]Renaissance Computing Institute, University of North Carolina at Chapel Hill

[8]Oregon Clinical & Translational Research Institute, Oregon Health & Science University, Portland, OR, USA

[9]DNAstack, Toronto, Ontario, Canada

[10]Linus Pauling Institute, Oregon State University, Corvallis OR, USA

[11]Australian e-Health Research Centre, CSIRO, UQ Health Sciences Building, Herston, Qld, Australia

[12]Baylor College of Medicine, Houston, USA

[13]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

[14]Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA

## Abstract

The Clinical Genome Resource (ClinGen)'s work to develop a knowledge base to support the understanding of genes and variants for use in precision medicine and research depends on robust, broadly-applicable, and adaptable technical standards for sharing data and information. To forward these goals, ClinGen has joined with the Global Alliance for Genomics and Health (GA4GH) to support the development of open, freely-available technical standards and regulatory frameworks for secure and responsible sharing of genomic and health-related data. In its capacity as one of 15

Corresponding Author: Heidi L. Rehm: hrehm@bwh.harvard.edu.

inaugural GA4GH "Driver Projects", ClinGen is providing input on the key standards needs of the global genomics community, and has committed to participating on GA4GH Work Streams to support the development of: i) a standard model for computer-readable variant representation; ii) a data model for linking variant data to annotations; iii) a specification to enable sharing of genomic variant knowledge and associated clinical interpretations, and; iv) a set of best practices for use of phenotype and disease ontologies. ClinGen's participation as a GA4GH Driver Project will provide a robust environment to test drive emerging genomic knowledge sharing standards and prove their utility among the community, while accelerating the construction of the ClinGen evidence base.

## Keywords

standards; variant representation; variant annotation; phenotype ontology; data sharing; genomic knowledge

---

The Clinical Genome Resource (ClinGen) is building a central knowledge base for understanding the clinical relevance of genes and variants for use in precision medicine and research [Rehm, 2015]. This includes the curation of genes for disease validity, dosage sensitivity and actionability as well as the curation of variants for pathogenicity. Curated variants are shared in the National Center for Biotechnology Information's ClinVar data archive and curated genes are available on the ClinGen website. These resources depend on robust technical standards for sharing data and information that are broadly applicable to a variety of use cases and are adaptable across a diversity of countries and systems, including both clinical and non-clinical settings. ClinGen is working to (i) standardize the clinical annotation and interpretation of genomic variants, (ii) enable clinicians, researchers, and patients to share evidence including genomic and phenotypic data, and (iii) provide unrestricted access to its knowledge base for direct use as well as integration into EHRs and other resources. As part of this effort, ClinGen has joined with the Global Alliance for Genomics and Health (GA4GH; www.ga4gh.org), an international, nonprofit alliance that is catalyzing the creation of technical standards and regulatory frameworks to enable responsible, voluntary, and secure sharing of genomic and health-related data across institutional and national boundaries.

Formed in 2013 to accelerate the potential of research and medicine to advance human health [Page, 2016], the GA4GH membership brings together over 500 leading organizations as well as individual contributors working in healthcare, research, patient advocacy, life science, and information technology, from across more than 70 countries. In October 2017, GA4GH launched a new phase ("GA4GH Connect", [https://www.ga4gh.org/docs/GA4GH-Connect-A-5-year-Strategic-Plan.pdf]) that depends on the expertise of real-world clinical and research projects to establish priorities and needs within the community. These real-world "Driver Projects" provide contexts for international genomic data sharing by: (i) establishing priorities for tool development, (ii) contributing to the creation of technical standards, policies, and other deliverables, and (iii) implementing GA4GH standards into real-world use in order to provide feedback and demonstrate the value of genomic data sharing to the broader community.

Previously, ClinGen and GA4GH have worked together on developing guidelines for sharing pediatric genomic data [Freidman, 2018; Rahimzadeh, 2018] and variant-level information with ClinVar [Azzariti, 2018], and developing consent resources for clinical genomic data sharing [Riggs, 2018]. ClinGen is also a key participating group within the BRCA Challenge, one of four early demonstration projects that helped launch and demonstrate the value of GA4GH. The BRCA Challenge launched the BRCA Exchange (http://brcaexchange.org/) which brings together all publically accessible variant resources on *BRCA1* and *BRCA2* including ClinVar as the primary source of interpreted *BRCA1* and *BRCA2* variants. Following this established history of mutual collaboration, ClinGen was invited to serve as one of the 15 inaugural GA4GH Driver Projects, alongside other leading research and clinical initiatives in North America, Europe, and Australia. In this role, ClinGen is contributing to the development of standards for discovering, accessing, storing, and analyzing genomic and related-health data that will be used by projects across the globe, including national precision medicine initiatives, such as the U.S.-based *All of Us* Research Program [Collins, 2015] and Genomics England [Marx, 2015], both of which are also participating as GA4GH Driver Projects. It will also play a leadership role in the representation of genomic knowledge for use in the accurate interpretation of genomic data.

In February 2018, GA4GH released a Strategic Roadmap [https://www.ga4gh.org/howwework/strategic-roadmap.html] based on the input and guidance received from Driver Projects regarding immediate, key needs for enabling data sharing. The Roadmap describes 28 deliverables that will be produced by eight GA4GH Work Streams over the next one to three years; all of the deliverables build upon the *Framework for Responsible Sharing of Genomic and Health-Related Data* [Knoppers, 2014] and will be made freely and openly available on https://ga4gh.org. ClinGen and the other GA4GH Driver Projects will work together to support the development of these deliverables and to ensure their relevance for use in real-world projects, as representatives of the broader genomics community. ClinGen has committed to participating in multiple sub-groups across four of the Work Streams, with key contributions listed below. In particular, ClinGen will help to:

1.

   **Create a standard model for computer readable variant representation.**

   Genomic variants are described with many naming conventions, making it difficult to unambiguously define a variant and ensure the accurate use of associated knowledge. ClinGen is leveraging prior work done by its Data Modeling Work Group (including experience in developing the ClinGen Allele Registry (https://reg.clinicalgenome.org)), contributing examples, and providing domain expertise, to inform efforts within GA4GH to develop a data model for unambiguous representation of variants. This work began within GA4GH as the Variant Modeling Consortium (VMC) [https://github.com/ga4gh/vmc] and is now continued through the Variant Representation subgroup of the Genomic Knowledge Standards Work Stream (GKS). GKS includes representatives from other organizations, including HL7 Fast Healthcare Interoperable Resources (FHIR) [http://hl7.org/fhir/] and Human Genome Variation Society (HGVS), ensuring that its standards meet the needs of the clinical and genomics communities and are compatible with HGVS standards that are widely used to contextualize genetic

variation. Notably, both VMC and the ClinGen Allele Registry, as transmission formats, have the potential to adapt to each other fluidly, and the ClinGen Allele Registry is working with the GA4GH group to define a pilot project to implement the GKS Variant Representation specification. A 0.1 release of VMC has been released and already proposes a language and nomenclature for describing variation.

2.

### Develop a data model for linking variant data to annotations.

Standardized variant annotation and interpretation is central to ClinGen's mission and is an area in which the consortium has considerable expertise. ClinGen and the Monarch Initiative [Mungall, 2016] are working with the GA4GH GKS Work Stream to develop a common data model to guide the linkage of variant evidence to clinical interpretations with a standard format. This includes support for applying current professional interpretation standards (e.g., ACMG/AMP [Richards, 2015]) in a computable manner that can be validated, as well as documenting the associated disease and inheritance pattern.

3.

### Develop a network for sharing knowledge about genomic variants and associated clinical interpretations.

Sharing curated genomic knowledge with databases like ClinVar is a high priority for the genomics community. Building off of work in the GKS and Clinical & Phenotypic Data Capture (CPDC) Work Streams, the Discovery Work Stream will develop standards for sharing variant classifications and supporting evidence. The effort will standardize technical descriptions of a variant and its attributes (e.g., clinical significance) to streamline the electronic submission of clinically relevant information to genomic knowledge bases like ClinVar. The data models will build off of GA4GH standards developed by the GKS and CPDC Work Streams in the areas of variant annotation and phenotyping, and will be implemented within ClinGen's knowledge bases, as well as disseminated to the broader community for widespread use. Facilitating knowledge exchange between disparate sources will enable the development of integrative and comprehensive applications helping to inform clinicians of the consequences and impacts of genomic variant events.

4.

### Establish recommended phenotype & disease ontologies and best practices for their use in genomic medicine.

Interpretation of genes and variants and their possible role in a patient's disease requires associating genes and variants to diseases and phenotypic features. The GA4GH CPDC Work Stream is developing standards, best practices, and benchmarking for the use of ontologies and clinical terminologies to capture clinical phenotype information and gene-disease associations for use in genomic medicine and to ensure data captured clinically can be used in genomic research. CPDC will also develop standards and best practices for how clinical phenotype information can be exchanged between clinical information systems and with

research, through using the emerging HL7 FHIR) and Phenopackets [http://phenopackets.org/] standards. ClinGen is implementing these standardized disease and phenotype ontologies into its gene-disease curation efforts as well as incorporating and testing developed phenotyping standards in its data capture approaches, including through its GenomeConnect patient registry [Kirkpatrick, 2015].

In summary, ClinGen is a critical Driver Project for GA4GH, providing a robust environment to test drive genomic knowledge sharing standards and prove their utility in the sharing of evidence and knowledge among the community, as well as applying that knowledge to clinical care and research. In exchange, ClinGen can more quickly and consistently build its evidence base by working with GA4GH to disseminate and instantiate the collaboratively built standards through global involvement and engagement.

## Acknowledgments

## References

Azzariti DR, Riggs ER, Niehaus A, Rodriguez LL, Ramos EM, Kattman B, … Rehm HL (2018). Points to consider for sharing variant-level information from clinical genetic testing with ClinVar. Molecular Case Studies,4(1). 10.1101/mcs.a002345

Collins FS, & Varmus H (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793–795. 10.1056/NEJMp1500523 [PubMed: 25635347]

Friedman JM, Bombard Y, Cornel MC, Fernandez CV, Junker AK, Plon SE, … Knoppers BM (2018). Genome-wide sequencing in acutely ill infants: Genomic medicine's critical application? Genetics in Medicine 10.1038/s41436-018-0055-z

Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, … & Faucett WA (2015). GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to

improve genomic knowledge. Human mutation, 36(10), 974–978. 10.1002/humu.22838 [PubMed: 26178529]

Knoppers BM (2014). Framework for responsible sharing of genomic and health-related data. The HUGO journal,8(1), 3 10.1186/s11568-014-0003-1 [PubMed: 27090251]

Marx V (2015). The DNA of a nation. Nature, 524, 503–505. 10.1038/524503a [PubMed: 26310768]

Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, … & Foster E (2016). The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic acids research, 45(D1), D712–D722. 10.1093/nar/gkw1128 [PubMed: 27899636]

Page A, Baker D, Bobrow M, Boycott K, Burn J, Chanock S, … Hudson TJ (2016). A federated ecosystem for sharing genomic, clinical data. Global Alliance for Genomics and Health. Science, 352(6291), 1278–1280. 10.1126/science.aaf6162 [PubMed: 27284183]

Rahimzadeh V, Schickhardt C, Knoppers BM, Sénécal K, Vears DF, Fernandez CV, … Friedman JM (2018). Key Implications of Data Sharing in Pediatric Genomics. JAMA Pediatrics,172(5), 476 10.1001/jamapediatrics.2017.5500 [PubMed: 29554172]

Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, … Watson MS (2015). ClinGen — The Clinical Genome Resource. New England Journal of Medicine,372(23), 2235–2242. 10.1056/nejmsr1406261 [PubMed: 26014595]

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, … & Voelkerding K (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine, 17(5), 405 10.1016/j.jmoldx.2016.10.002 [PubMed: 25741868]

Riggs ER, Azzariti DR, Niehaus A, Goehringer SR, Ramos EM, Rodriguez LL, … Martin CL (2018). Development of a consent resource for genomic data sharing in the clinical setting. Genetics in Medicine 10.1038/s41436-018-0017-5