Creating An Open-source Gene Curation Database From The Gene Curation Coalition

Marina T. DiStefano[1], Scott Goehringer[1], Joanna Amberger[2], Christina Austin-Tse[3, 4], Marie Balzotti[5], Jonathan S. Berg[1], Ewan Birney[6], Ozge Ceyhan-Birsoy [4], Carol Bocchini[2], Elspeth Bruford[7], Alison Coffey[8], Heather Collins[9,10], Fiona Cunningham[6], Paul Denny[7], Helen Firth[11], David Fitzpatrick[11], Rebecca E. Foulger[12], Oleg Gerasimenko[12], Jenny Goldstein[1], Ada Hamosh[2], Matthew Hurles[11], Sarah E. Leigh[12], Ivone U.S. Leong[12], Antoine Marmignon[13], Christa L. Martin[1], Kalotina Machini [15], Ellen M. McDonagh[12], Annie Olry[13], Kelly Radtke[14], Erin M. Ramos[15], Erin Rooney Riggs[1], Charlotte Rodwell[13], Antonio Martin Rueda[12], Catherine Snow[12], Zornitza Stark[16], Jackie Tahiliani[17], James S. Ware[18,19], Eleanor Williams[12], Caroline Wright[11], and Heidi L. Rehm[1]

[1]Clinical Genome Resource (ClinGen), USA, [2]Online Mendelian Inheritance in Man (OMIM), Johns Hopkins University School of Medicine, USA, [3]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA, [4]Partners Healthcare Laboratory for Molecular Medicine, Cambridge, MA, USA, [5]Myriad Women's Health, San Francisco, CA, USA, [6]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK , [7]HUGO Gene Nomenclature Committee (HGNC), European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [8]Illumina Clinical Services Laboratory, Illumina Inc., San Diego, CA, USA, [9]Genetics Home Reference, National Library of Medicine, Bethesda, MD, USA, [10]ICF, 9300 Lee Highway, Fairfax, VA, 22031 USA,[11]Transforming Genetic Medicine Initiative, Wellcome Genome Trust, England, UK Deciphering Developmental Disorders/Gene2Phenotype, England, UK, [12]Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square London, EC1M 6BQ, UK, [13]Inserm,US14-Orphanet, France, [14]AmbryGenetics, Aliso Viejo, CA, USA, [15]National Human Genome Research Institute, National Institutes of Health, USA, [16]Australian Genomics, Australia, [17]Invitae, San Francisco, CA, USA, [18]National Heart & Lung Institute & MRC London Institute of Medical Sciences, Imperial College London, London, UK, [19]Royal Brompton & Harefield Hospitals NHS Trust, London, UK

GenCC
The Gene Curation Coalition

Background

Several groups and resources provide information that pertains to the validity of gene-disease relationships; however, the standards and terminologies to define the evidence base for a gene's role in disease are still evolving and the community is in need of trusted and harmonized sources that define the level of evidence for a gene's role in disease. To tackle this issue, the **Gene Curation Coalition (GenCC)** was formed.

The Gene Curation Coalition brings together groups engaged in the evaluation of gene-disease validity with a willingness to share data publicly, to develop consistent terminology for gene curation activities and to facilitate the consistent assessment of genes that have been reported in association with disease.

The goals of the GenCC are as follows:

- Clarify the overlap between gene curation efforts
- Understand the aims, processes, information used, classification systems, and users of the different curation efforts
- Develop consistent terminology for validity assessment as well as inheritance, allelic requirement, and mechanism of disease
- Collaborate on gene curation projects

More information, news, and updates about the GenCC can be found on our newly launched website: **www.TheGenCC.org**

GenCC Member Groups

Ambry Genetics
A CLIA-certified fee-for-service clinical testing laboratory

ClinGen
National Institutes of Health (NIH)-funded resource dedicated to building a central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.

DECIPHER
Database of genomiC varIation and Phenotype in Humans using Ensembl Resources is an interactive web-based database which incorporates a suite of tools designed to aid the interpretation of genomic variants. It is widely used to support clinical practice in the UK NHS, and an international network of clinicians and researchers working with predominantly rare diseases.

Genomics England's PanelApp
A knowledgebase of curated gene panels which crowdsources expert reviews for gene-disease validity assessment. The gene panels are utilised by Genomics England's genome interpretation services, support a consensus in gene content for the NHSE Genomic Medicine Service, as well as worldwide for omics analysis.

Illumina Clinical Services Laboratory
Offers the TruGenome Undiagnosed Disease Test, a clinical whole-genome sequencing test for patients with a suspected rare and undiagnosed genetic disease. The lab also supports clinical programs such as the iHope Program, which donates clinical genome sequencing tests to help find answers for children facing these types of diseases

Invitae
A CLIA-certified fee-for-service clinical testing laboratory

Myriad Women's Health
A CLIA-certified clinical testing laboratory that provides genetic screening and support for women and their families

OMIM
Online Mendelian Inheritance in Man is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily

Orphanet
Orphanet (www.orpha.net) is a knowledge base on rare diseases and orphan drugs, bridging the fields of healthcare and research. Orphanet, a network of 38 countries, aims to increase knowledge on rare diseases so as to improve the diagnosis, care, and treatment of rare diseases. Orphanet provides a medical terminology dedicated to rare diseases, the Orphanet nomenclature of rare diseases (ORPHA code) used in healthcare and research in Europe: it is annotated with curated scientific data, including rare disease-related genes

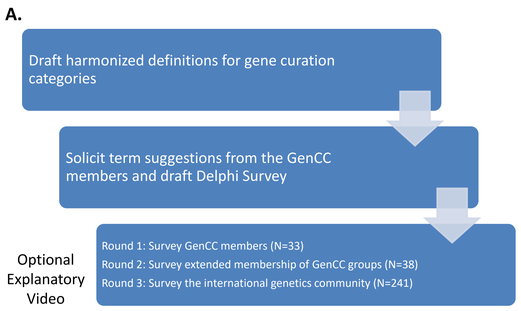
Partners Laboratory for Molecular Medicine
A CLIA-certified molecular diagnostic laboratory operated by Partners HealthCare Personalized Medicine. The LMM is led by a group of Harvard Medical School-affiliated faculty, geneticists, clinicians, and researchers from Brigham and Women's Hospital and Massachusetts General Hospital, Partners' founding members. Our mission is to bridge the gap between research and clinical medicine.

PanelApp Australia
Managed by the Australian Genomics Health Alliance (Australian Genomics) and is used by Australian diagnostic laboratories, clinicians and researchers to establish and maintain consensus virtual gene panels for use in genomic analysis.

Transforming Genomic Medicine Initiative
(TGMI) aims to improve the quality and efficiency of clinical reporting from genomic sequence information. It is a collaboration between researchers at the University of Edinburgh, EMBL European Bioinformatics Institute, University of Cambridge, University of Exeter, Imperial College London, University of Manchester, the Broad Institute, and the Wellcome Sanger Institute, and is funded by the Wellcome Trust. TGMI are building resources to improve the speed accuracy, sensitivity and precision of information to support clinical genome interpretation

Steering Committee Members	
Heidi Rehm, GenCC Chair	ClinGen
Marina DiStefano, GenCC Coordinator	ClinGen
Christina Austin-Tse	Laboratory for Molecular Medicine (LMM)
Marie Balzotti	Myriad Women's Health
Alison Coffey	Illumina
Heather Collins	Genetics Home Reference
Fiona Cunningham	EBI
Helen Firth	DECIPHER
Ada Hamosh	OMIM
Ellen McDonagh	Genomics England PanelApp
Annie Olry	Orphanet
Kelly Radtke	Ambry
Zornitza Stark	PanelApp Australia
Jackie Tahiliani	Invitae
James Ware	Gene2Phenotype (G2P) from TGMI

Clinical Validity Term Delphi Survey



B.

Overall preferred Terms/Set
Definitive
Strong
Moderate
Limited
Disputed Evidence
Refuted Evidence
Animal Model Only
No Known Disease Relationship

Figure 1: A. GenCC Clinical Validity Term Delphi Survey Process. To harmonize terms describing gene-disease validity, the GenCC used a Delphi method to survey both members of our GenCC organizations and the international genetics community. This survey was conducted in three rounds. A short introductory video describing the survey was provided to all survey takers.
B. Standardized Clinical Validity Term Set. Terms that were agreed upon are "Definitive, Strong, Moderate, Limited, Disputed Evidence, Refuted Evidence, Animal Model Only, and No Known Disease Relationship". GenCC groups will now use or map to relevant harmonized terms in their curation work.

Validity Data Summary Mockup

NOTE: This is a MOCKUP and does not display real curation results from GenCC Members

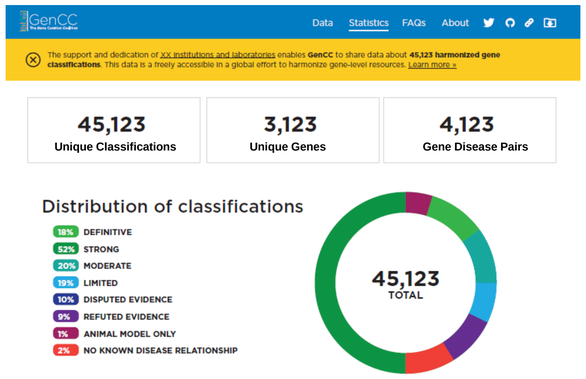


Figure 2: Mockup of Summary Data, Users will see the number of unique classifications (a classification is attached to a gene, disease, a mode of inheritance, and a GenCC member group), unique genes, and unique diseases. Clinical validities are differentiated by color.

Validity Gene List Mockup

NOTE: This is a MOCKUP and does not display real curation results from GenCC Members

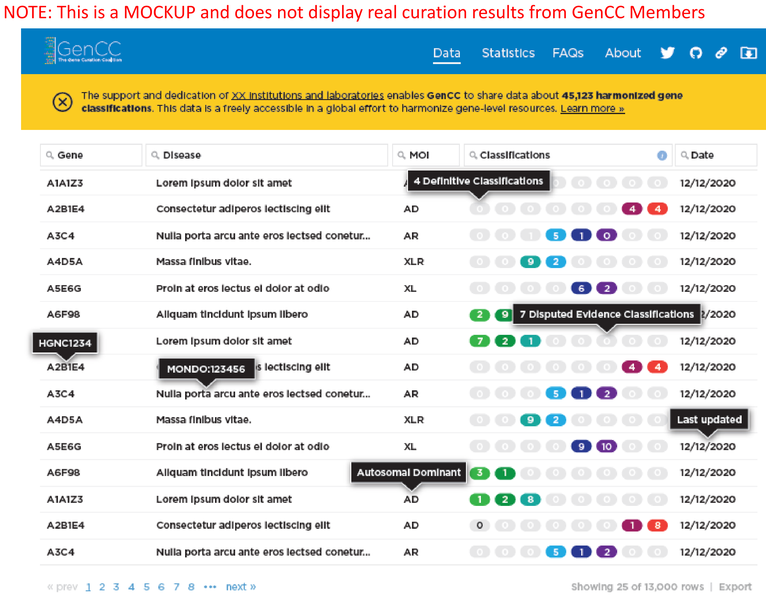


Figure 3: Mockup of the gene list. Curated genes will be displayed in a list format with the HGNC gene ID, the disease from the Monarch Disease Ontology (MONDO), the Mode of Inheritance (MOI), and the date of the most recent validity classification. Curations will be grouped and tallied by clinical validity with differing classifications for the same gene disease pair being displayed as different colors. The list will can be searched, filtered, and downloaded.

Validity Data from GenCC Members

GenCC is creating a harmonized display for the following validity data:

- **~25,000** OMIM entries
- **▪872** classifications from ClinGen
- **▪>12,000** gene disease relationships from Genomics England's PanelApp
- **~7400** curations from Orphanet
- **5095** gene disease pairs curated by Ambry
- **~14,000** gene disease pairs curated by Invitae
- **~4000** gene disease relationships from the PanelApp Australia

Classification Page Mockup

NOTE: This is a MOCKUP and does not display real curation results from GenCC Members

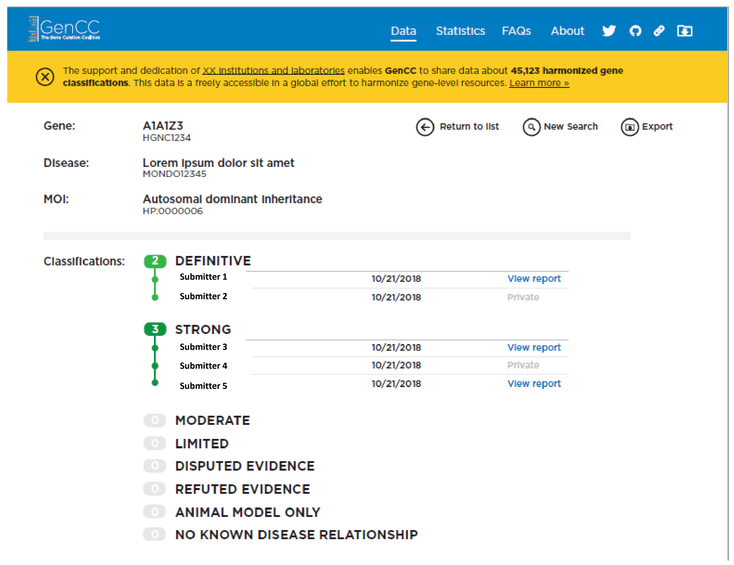


Figure 4: Classification Page Mockup. Each page will display the gene, disease (MONDO), and mode of inheritance (MOI). Classifications will be sorted by clinical validity. Each submitter, the date curated, and a link to public evidence will be included. Each page can be exported.

Conclusions and Future Directions

- The Gene Curation Coalition (GenCC) was formed to standardize nomenclature surrounding gene curation
- A Delphi survey was completed to standardize clinical validity terms and all GenCC members will adopt them
- An ongoing project is working with all members to publicly display thousands of gene curation results

References

1) Strande NT, Rooney Riggs E, et al. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the Clinical Genome Resource. bioRxiv. doi: https://doi.org/10.1101/111039
2) Farwell hagman KD, Shinde DN, Mroske C, et al. Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. Genet Med. 2017;19(2):224-235.
3) Ceyhan-birsoy O, Machini K, Lebo MS, et al. A curated gene list for reporting results of newborn genomic sequencing. Genet Med. 2017;19(7):809-818.
4) Machini K, Ceyhan-birsoy O, Azzariti DR, et al. Analyzing and Reanalyzing the Genome: Findings from the MedSeq Project. Am J Hum Genet. 2019;105(1):177-188.

This work was supported in part by the following sources: ClinGen (NIH/NHGRI Grants: U41HG006834 (Rehm), U41HG009649 (Montine), U41HG009650 (Berg)), TGMI/G2P: Wellcome Strategic Award 200990/Z/16/Z