Curating the Clinical Genome Meeting 2019

Poster Presentation
Abstracts

Curating the Exome of hypomagnesemia with secondary hypocalcemia patient

M. Kamran Azim

Department of Biosciences, Mohammad Ali Jinnah University, Karachi, Pakistan

Hypomagnesemia with secondary hypocalcemia is a rare autosomal-recessive disorder characterized by intense hypomagnesemia associated with hypocalcemia (HSH). Mutations in the *TRPM6* gene, encoding the epithelial Mg^{2+} channel TRPM6, have been proven to be the molecular cause of this disease. This study identified causal mutations in patients of hypomagnesemia. Biochemical analyses indicated the diagnosis of HSH due to primary gastrointestinal loss of magnesium. Whole exome sequencing of the trio (i.e. proband and both parents) was carried out with mean coverage of $>150\times$. ANNOVAR was used to annotate functional consequences of genetic variation from exome sequencing data. After variant filtering and annotation, a number of single nucleotide variants (SNVs) and a novel 2 bp deletion at exon26:c.4402_4403delCT in *TRPM6* gene were identified. This deletion which resulted in a novel frameshift mutation in exon 26 of this gene was confirmed by Sanger sequencing. In conclusion, among several candidate genes, present trio exome sequencing study identified a novel homozygous frame shift mutation in TRPM6 gene of HSH patient. However, it should be noted that exome sequencing does not cover large genomic rearrangement such as copy number variations (CNVs).

Medical Action Ontology (MAxO)

Leigh C. Carmody, Xingmin A. Zhang, Nicole A. Vasilevsky, Chris J. Mungall, Nico Matentzoglu, Peter N. Robinson

A standardized, controlled vocabulary allows medical actions to be described in an unambiguous fashion in medical publications and databases. The Medical Action Ontology (MAxO) is being developed to provide a structured vocabulary for medical procedures, interventions, therapies, and treatments for rare diseases. The ontology will capture medical actions used to treat rare diseases from published manuscripts and will be utilized in new computational algorithms that exploit semantic similarity between related medical actions, phenotypic abnormalities as coded in the Human Phenotype Ontology (HPO), and related genetic mutations using Exomiser to suggest new medical actions for rare diseases. Annotations to publications about rare disease clinical management will be added to the HPO resource. This ontology coupled with computer algorithms does not intend to replace physician's advice, guidelines of medical societies, or textbooks, but rather to complement these sources of information. It provides a convenient method for quickly extracting difficult to find information about treatments for rare diseases. Currently, MAxO contains approximately 300 terms and will be expanded via more data mining from sources such as GeneReviews, PubMed queries, and Rare Disease Network. MAxO is open source and will become freely available. A preliminary version is available at https://github.com/monarch-initiative/MAxO.

Use of ACMG Criteria in a cWGS Test for Rare and Undiagnosed Disease.

Anjana Chandrasekhar, Krista Bluske, Nicole J. Burns, Alison J. Coffey, Alka Malhotra, David R. Bentley, Ryan J. Taft, Denise L. Perry

Illumina Clinical Services Laboratory, Illumina Inc., San Diego, CA, USA.

The Illumina TruGenome Undiagnosed Disease™ clinical whole genome sequencing (cWGS) test is intended to provide information to physicians to aid in the diagnosis of patients with suspected rare and undiagnosed genetic disease. Variants are filtered based on frequency, phenotype and inheritance before selection for curation and classification in accordance with American College of Medical Genetics and Genomics (ACMG) guidelines.

A retrospective analysis of 151 variants revealed that certain ACMG criteria were used more frequently than others based on the nature of the cWGS test. Absence of the variant from population databases (PM2) was the most often used criterion, applied to 98 variants (65%). Loss of function criterion (PVS1) was used 35 times (23%), either at full strength or at reduced strength. The de novo inheritance criterion (PS2) was used for 26 variants (17%), as our analyses typically include parental samples when available. Several criteria were seldom used due to the rare nature of the variants identified in our test, including identification of a variant at the same nucleotide (PS1) or amino acid position (PM5) as an established pathogenic variant. Criteria used to classify benign variants were also very rarely applied to curated variants, as the conditions met by these criteria were utilized upstream in the analysis to rule out variants that are unlikely to result in disease. No ACMG criteria were applied to 23 (15%) of the variants, resulting in a classification of VUS.

To help with consistent application of criteria across the interpretation team, we have developed internal supplementary guidelines that add granularity to ACMG criteria. For example, our internal guidelines allow application of criteria at different strengths depending on the evidence in the literature, including number of affected individuals (PS4) and quality and biological relevance of functional data (PS3).

In summary, our experience highlights the need for a broader community discussion about the application of ACMG criteria in a clinical genome or exome setting, and the potential need to develop deeper guidance on the application of certain criteria to ensure consistency across the community.

The Illumina Clinical Services Laboratory (ICSL)'s use of reactive gene curation to support reporting for a clinical whole genome sequencing test

Amanda R. Clause, Julie P. Taylor, Erin Thorpe, Alicia Scocchia, David R. Bentley, Ryan J. Taft, Denise L. Perry, Alison J. Coffey

Illumina Clinical Services Laboratory, Illumina, Inc.

Gene curation is essential to variant interpretation. A variant's contribution to a disease cannot be determined without first evaluating the evidence for the role of the gene in the disease. The strength of the gene-disease association (GDA) then determines how a variant in that gene is reported. ICSL uses the ClinGen Gene-Disease Validity framework to support reporting for the TruGenome Undiagnosed Disease clinical whole genome sequencing (cWGS) test for patients with suspected rare and undiagnosed genetic disease (RUGD). GDAs identified through case analysis are reactively curated within the 5-day interpretation and reporting window.

To date, we have curated ~300 GDAs through our RUGD workflow. 216 were not previously curated by ClinGen or BabySeq and are novel contributions to the GDA knowledge base. 226 GDAs were classified as definitive, strong, or moderate, qualifying likely pathogenic and pathogenic variants to be reported as positive findings within our report structure.

Approximately 50% of cases in which a potentially causative variant was identified required gene curation. GDAs previously classified as strong or definitive by ClinGen or BabySeq were not re-curated. Only six genes had variants reported more than twice in unrelated cases, reflecting the diversity of genes and diseases encountered in the RUGD population and in a cWGS test. 193 GDAs fall outside the scope of ClinGen Gene Curation Expert Panels, highlighting the need for a panel for rare disease.

56 GDAs were classified as limited. Strong phenotypic overlap or experimental support for the GDA led to variants in 13 of these genes being reported as a VUS in a GUS for the disease. Examples include the AQP11 gene, with strong experimental support for an association with polycystic kidney disease, and the TBC1D32 gene, with moderate experimental support and a single case report with strong phenotypic overlap with the proband. Genes with a limited disease association are submitted to GeneMatcher to identify additional cases to support the GDA, possibly leading to novel disease associations, such as AMMERC1 with growth, bone and heart alterations (Moyses-Oliveira et al. 2017).

Rigorous evaluation of the strength of a GDA using the ClinGen framework can be successfully incorporated to support variant interpretation in a fast-paced, reactive reporting environment and provides a foundation for up-to-date and consistent clinical reporting decisions. All curated GDAs will be shared through ClinGen.

BRCA Exchange: Data Integration for BRCA Variant Interpretation

Authors: Melissa Cline[1], Faisal Alquaddomi[2], Samantha Baxter[3], Amy Coffin[1], Rob Currie[1], Zachary Fischmann[1], Mary Goldman[1], Marc Zimmerman[2], Rachel Liao[3], Gunnar Råtsch[2]

Author affiliations:

1. UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
2. ETH Zurich, Zurich, Switzerland
3. Broad Institute, Cambridge, MA, USA

Abstract

The rate of discovery of new genetic variants has outpaced our ability to interpret them. Even in the well-characterized BRCA genes, upwards of 40% of the variants currently in ClinVar are Variants of Uncertain Significance (VUS). Further, more than 25% of the publicly-known variants are not yet in ClinVar. The BRCA Exchange (https://brcaexchange.org), a driver project of the Global Alliance for Genomics and Health, seeks to accelerate BRCA variant interpretation through data integration. BRCA Exchange currently contains the largest set of publicly-available BRCA variants anywhere, with variants from ClinVar, Leiden Open Variant Database (LOVD), population databases and other sources. These variants are displayed with rich data annotations including functional scores, expert-based pathogenicity prediction and extensive literature references, and are integrated with analysis tools to allow visualization of the variant in its transcript and protein structure context. Since its inception in 2016, BRCA Exchange has supported the ENIGMA Consortium in increasing the number of expert-interpreted BRCA variants by about seven-fold. BRCA Exchange is a work product of the BRCA Challenge, an international consortium of expert clinicians, diagnosticians, researchers and database providers, all with a common goal of advancing our understanding of BRCA1 and BRCA2 variation and providing an exemplar for additional genes.

Evaluating the consequences of abnormal gene dosage

Jannine D. Cody[1,2,*], Jonathan A. Gelfond[3]
[1] Department of Pediatrics, The Chromosome 18 Clinical Research Center , University of Texas Health Science Center at San Antonio, San Antonio, Texas, 78229, USA [2]The Chromosome 18 Registry and Research Society, San Antonio, Texas, 78229, USA [3]Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, Texas, 78229, USA ,

Genomic copy number variation is a characteristic of the human genome but is also a major contributor to congenital disability and disease risk. This makes it imperative to understand which regions of the genome can withstand copy number change without effect, which predispose to disease and which cause outright disability. Numerous studies have created bioinformatic tools for predicating which genes aredosage sensitive on a genome wide scale. However, these studies have not had sufficiently robust data with which to test the validity of their predictions. We have created such a data set by individually curating the data for each gene on chromosome 18 with regard to their potential effects when present in an abnormal dosage. We utilized data from the human genetic disease literature, animal models, the Database of Genomic Variants as well as our own extensive longitudinal dataset on over 650 individuals with chromosome 18 copy number changes. We created a novel clinical gene dosage classification: (1) no clinically adverse effect (2) a risk factor (3) a conditional factor (4) a low penetrance effect (5) causal of an abnormal phenotype and (6) an embryonic lethal. We evaluated nine bioinformatic copy number sensitivity prediction tools with our classifications of the chromosome 18 genes. While some of the nine bioinformatic predictors were statistically associated with our robustly curated clinical classifications (Rudefer et al., 2016, ExAC pLI Spearman's $r=0.22$, $p=0.001$), we found little predictive accuracy of these predictions which explained from <1% to 5% of the variation of clinical dosage sensitivity.

A. Danos, K. Krysiak, A. Wagner, S. Kiwala, J. McMichael, A. Coffman, E. Barnell, L. Sheta, S. Pema, N. Spies, L. Kujan, M. Griffith, O. Griffith

As next generation sequencing techniques continue to be integrated into clinical practice, the clinical interpretation of tumor variants becomes a bottleneck for treatment as it requires engaging a large volume of information. One approach to the bottleneck problem is to utilize paid curation, housed behind a paywall. A different approach is to create a fully free and open resource by means of public curation. The Clinical Interpretations of Variants in Cancer (CIViC, www.civicdb.org) database employs the later method via a publicly curated and expertly moderated collection of cancer variants where the clinical annotations are stored in a structured manner, with a flexible public API for easy programmatic access, as well as a web interface with advanced search capability for users. No login is required for full access to the data, while a free account is required for curation. The CIViC model emphasizes transparency, as the provenance of all variant annotation from submission, revision, through acceptance of the evidence is maintained and viewable on the web interface. The interface also allows curators as well as organizations to track and see summary statistics of all of their contributions and activity. To date, CIViC has a community of over 150 contributors and 16,000 clinical and research users worldwide.

CIViC continually adds features, updates, and developments to the data model based on user feedback and internal discussion. The fundamental unit of knowledge in CIViC is the evidence item (EID), and all EIDs are associated with a PubMed ID. CIViC now also accepts ASCO abstracts as an EID evidence source. EIDs in CIViC are of predictive (therapeutic), prognostic, diagnostic or predisposing type. In order to capture functional data, a functional evidence type has been introduced, which is structured to annotate gain of function, loss of function and neomorphic variant annotations. EIDs are used to build summary statements called assertions, for a variant in a specific cancer context. Assertions admit an AMP Tier and Level when supported by clinical evidence, and admit a five-tier pathogenicity assessment for predisposing germline evidence. ACMG codes supporting predisposing assertions are captured in germline EIDs. Multiple groups including CIViC are developing models for assessment of somatic variant oncogenicity/pathogenicity, and the outcome of these discussions will be captured in the CIViC data model using the assertion architecture.

Collaborative Refinement of Predicted Gene Function with Apollo

Nathan Dunn 1
Deepak Unni 1
Colin Diesh 2
Dan Keith 3
Luke Sargent 4
Helena Rasche 5
Robert Buels 2
Nomi Haris 1
Ian Holmes 2

1 Lawrence Berkley National Lab, Berkeley, CA, USA
2 University of California, Berkeley, Berkeley, CA, USA
3 Oregon State University, Covallis, OR, USA
4 Oregon Health Sciences University, Portland, OR, USA
5 University of Freiburg, Freiburg, Germany

Improving the quality of genome annotations allows a more accurate prediction of gene function and thus a more accurate imputation of its role in disease etiology. To this end, the Apollo genome annotation editing tool (https://github.com/GMOD/Apollo / http://genomearchitect.org/ / https://github.com/GMOD/docker-apollo) provides a way for teams of researchers to collaboratively refine automated gene predictions using experimental evidence in a manner similar to Google Docs. Our current work focuses on annotating the effects of individual variants on gene structure in order to identify candidate causal variants for changes in gene expression and phenotype.

Apollo servers are highly customizable and can scale to support the annotations of multiple genomes. Its suite of web services enables Apollo to be readily integrated into bioinformatics workflow environments. These web services can be accessed directly via a Python library http://bit.ly/python-apollo-client-reference, and are also used by the Monarch Initiative (e.g. https://monarchinitiative.org/gene/MGI:104537) and the Alliance of Genome Resources (e.g. https://www.alliancegenome.org/gene/MGI:104537) to display gene structure on disease pages using a cross-browser widget (https://www.npmjs.com/package/genomefeaturecomponent). Additionally, because it is built on top of the JBrowse genome browser (https://github.com/GMOD/jbrowse) and its rich plugin ecosystem (https://gmod.github.io/jbrowse-registry/), Apollo has been integrated into external web services such as http://myvariant.info/ via a simple plugin (http://bit.ly/jbrowse-myvariantinfo-plugin / demo [http://bit.ly/jbrowse-myvariantinfo-plugin-demo]) to provide clinical genome annotation data with minimal configuration.

Apollo provides many time-saving features for editing genomic elements such as maintaining a revertible history of structural edits and the ability to change the type of a genome feature. Apollo can read and display multiple tracks of variants as well as predictions of those variant effects.

Apollo is used in over one hundred genome annotation projects around the world, ranging from annotation of a single species to lineage-specific efforts supporting the annotation of dozens of species at a time.

Title: Harmonizing clinical interpretation of intragenic sequence and copy number variants in monogenic disease

Authors:
Ali Entezam, Yuya Kobayashi, Rebecca Truty, Rachel Lewis, Keith Nykamp, Invitae Clinical Genomics Group, Swaroop Aradhya

Affiliation: Invitae Corporation, San Francisco, CA

character limit for body: 2,500 characters (including spaces). Current, 2482

Establishing rigorous rubrics to interpret the significance of vast amounts of genetic information from clinical testing for germline disorders in a consistent manner across large numbers of individuals represents a major challenge. The recent revision of American College of Medical Genetics and Genomics (ACMG) guidelines for interpreting sequence variants (SNVs) was a significant step to address this challenge. Existing ACMG guidelines for interpreting copy number variants (CNVs) describe criteria that address chromosomal structural abnormalities and their clinical implications. However, neither guidelines have attempted to harmonize interpretation rules for SNVs and CNVs and have not included explicit rules for interpreting single gene or sub-genic CNVs.

We previously described an interpretation schema, Sherloc, which built on ACMG guidelines. One key feature of Sherloc was the organization of evidence criteria into categories based on five clinical molecular genetics concepts: “variant type”, “population data”, “clinical observations”, “experimental studies”, and “computational/predictive data” focused heavily on SNVs. Based on experience gained from testing a large clinical cohort, we iteratively refined Sherloc for interpretation of CNVs. Importantly, most of the improvements occurred within the “variant type” section, allowing the application of other evidence categories to be consistent across SNVs and CNVs.

To establish CNV “variant type” criteria, deletion and duplication CNVs are categorized based on (1) the likelihood of resulting in a null product through the nonsense-mediated decay (NMD) pathway, and (2) whether loss-of-function (LOF) variants in this gene are known to cause disease. These CNV criteria are variably weighted in a manner consistent with similar SNV categories. For example, whole gene deletions are scored the same as nonsense variants that are expected to result in NMD, and would reach a pathogenic classification in genes where LOF is the established disease mechanism. In total, we identified 13 possible CNV categories for a total of 26 Sherloc criteria to capture both LOF and non-LOF situations.

In summary, we have constructed a series of explicit interpretation criteria for intragenic and cytogenetic CNVs taking molecular genetic principles into account. These criteria were built within the existing Sherloc interpretation schema, enabling us to create a unified interpretation of framework for both SNVs and CNVs.

Using RefSeq Functional Element annotations for interpreting non-coding variation

Catherine M. Farrell, Terence D. Murphy, and the RefSeq Curation and Development Team

National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

The human genome contains many non-genic elements that are known to function in gene regulation, chromosome organization, recombination, repair or DNA replication. Sequence variation in such elements can result in human disease, with many genome-wide association studies showing a predominance of disease-associated variation in non-genic regions. While several large-scale epigenomic mapping projects have predicted the locations of gene regulatory elements, those data are not generally visible in traditional genome annotation, and interpretation of such data can require specialized research knowledge and customized displays that are not easily amenable to the bioclinical community. Therefore, NCBI has introduced a new RefSeq Functional Elements (www.ncbi.nlm.nih.gov/refseq/functionalelements/) dataset that is annotated on the human genome alongside conventional gene annotations. This curated dataset, which is defined based on published experimental results, includes richly annotated RefSeq records and accompanying descriptive records in the Gene database (www.ncbi.nlm.nih.gov/gene/). The dataset includes known enhancers, silencers, recombination regions, and other non-genic regions with experimentally-validated function. Functional element features annotated on the genome include descriptive details, experimental evidence and publication links. As of NCBI's Updated Annotation Release 109.20190125 on the GRCh38.p12 human genome assembly, the dataset includes over 3.7K GeneIDs and 8.8K feature annotations, with further growth expected for future NCBI annotation releases. The dataset is publicly available for FTP download (ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens). Feature annotation can be visualized in the 'Biological regions' track available in NCBI browsers, including the Variation Viewer (www.ncbi.nlm.nih.gov/variation/view/) and Genome Data Viewer (www.ncbi.nlm.nih.gov/genome/gdv/). This data track is particularly useful when viewed alongside variation data tracks from dbSNP, ClinVar and dbVar, with the additional option of viewing it alongside study-specific variation data from custom tracks or track hubs. This presentation will provide further information on how to use and access the dataset, and specific non-coding variant examples will be used to demonstrate how function can be interpreted from functional element feature annotations.

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

Title: "Unique Challenges in Fetal Genomic Curation"
Authors: The Fetal Sequencing Consortium

Determining the pathogenicity of any given variant in the genome requires accurate phenotyping as well as bioinformatics resources and expertise. Since 2015, our consortium has found 13-32% of fetuses with structural anomalies have a causative genotype based on ACMG variant classification criteria. We have also appreciated that curating fetal genomic variants poses unique challenges. Here, we outline these challenges with illustrative cases and offer possible solutions. All cases have a non-causative microarray and trio exome sequencing. **CHALLENGES (1)** There is incomplete phenotypic fetal data due to the limitations of ultrasound and the temporal limits inherent to fetal development. We submit a case of prenatally diagnosed hydrocephalus and aqueductal stenosis. Testing revealed a *de novo* p.Thr1092Ser variant in *DYNC1H1,* associated with dominant intellectual disability with neuronal migration defects. Post-natal imaging did not confirm hydrocephalus, however did confirm dysgenesis of the corpus callosum linking this variant to the proband's phenotype. **(2)** There is a lack of data on prenatal phenotypes in existing databases, possibly because the most severe phenotypes are embryonic lethal or cause fetal demise and are therefore never tested or documented in databases. We present a fetus with bilateral renal agenesis, anhydramnios and a complex congenital heart defect. Postnatally, the neonate had skin sloughing on the scalp consistent with aplasia cutis and nail dysplasia. A *de novo* p.Asn29Ser variant in *KCTD1*, associated with autosomal dominant scalp-ear-nipple (SEN) syndrome was found. This is the first published prenatal presentation of SEN to our knowledge. These prenatal findings further expand the phenotype of this condition**.** **(3)** There may be differences in inheritance models between previously described childhood cases and fetal cases. *RYR1* is associated with malignant hyperthermia when dominant and central core disease when recessive. We describe a case of pleural effusions, scalp edema, ascites, IUGR and demise found to have p.P816L and p.P2567S in *trans*. Both parents did not have a history of symptoms of malignant hyperthermia. **SOLUTIONS** Standardized data collection and collaborative dialogue to further generate prenatal genotype-phenotype correlations is required to expand fetal genomics into meaningful clinical use for patients and providers.

Assessing the Strength of Evidence for Genes Implicated in Aminoacidopathies Using the ClinGen Clinical Validity Framework

Jennifer Goldstein[1], Jennifer McGlaughon[1], Justyne Ross[1], Courtney Thaxton[1], Amanda Thomas[2], Kathleen Wallace[1], Diane Zastrow[3], Meredith Weaver[4], Heather Baudet[1], Annette Feigenbaum[5], Uta Lichter-Konecki[6], Marzia Pasquali[7], William Craigen[8]

1. Department of Genetics, University of North Carolina at Chapel Hill, NC; 2. Clinical Molecular and Cytogenetics, Columbia University, NY; 3. Stanford University, Palo Alto, CA; 4. American College of Medical Genetics and Genomics, Bethesda, MD; 5. Department of Pediatrics, University of California San Diego; Rady Children's Hospital, San Diego, CA; 6. Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA; 7. Department of Pathology, University of Utah and ARUP Laboratories, Salt Lake City, UT; 8. Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The aminoacidopathies are a diverse group of disorders, with widely varying clinical presentations, which are typically caused by defects in enzymes and transport proteins important in amino acid metabolism. Newborn screening can identify some, but not all, of these conditions. In order to examine the strength of evidence between specific aminoacidopathies and the genes implicated in these disorders, the NIH-funded Clinical Genome Resource (ClinGen) has established an Aminoacidopathy Gene Curation Expert Panel. This group, currently composed of 4 curators and 6 experts in the field, is using the ClinGen gene-disease clinical validity framework to evaluate 75 gene-disease pairs, prioritized in 3 tiers based on predicted strength of evidence and clinical utility. The curator first collects and assesses available case level data, case-control studies, and experimental data implicating the gene in the disease, and assigns a provisional clinical validity classification (definitive, strong, moderate, limited, no reported evidence or refuted/disputed). This is followed by expert review, and further discussion with the curator and the group if necessary, before a final classification is made and posted on the ClinGen website (www.clinicalgenome.org). Thus far, we have completed the curation of more than 30 gene-disease pairs. We will present the results of our curation efforts to date, and discuss the challenges of assessing clinical validity in this disease area.

Opinions of ClinGen Members on the Use of ClinGen Clinical Validity Classifications in Genetic Testing

Jennifer L Goldstein[1], Julianne O'Daniel[1], Marina DiStefano[2], Adam Buchanan[3], Kelly Ormond[4], Laura Milko[1], Heidi Rehm[2], Jonathan S Berg[1]

[1] Department of Genetics, School of Medicine, UNC-Chapel Hill, Chapel Hill, NC
[2] The Broad Institute of MIT and Harvard, Boston, MA
[3] Genomic Medicine Institute, Geisinger Health System, Danville, PA
[4] Department of Genetics, Stanford University, Stanford, CA

In recent years, advances in genetic testing technologies have led to the continuous identification of new gene-disease relationships; however, the evidence base for each new claim varies, creating challenges in deciding which genes to include in testing panels or to report to patients. The Clinical Genome Resource (ClinGen) has developed a standardized framework to classify the clinical validity of gene-disease pairs. However, no recommendations have been made on how clinical validity classifications should be used in genomic testing. To understand the perspectives of ClinGen members on this topic, we conducted an online survey. There were 98 responses (12% response rate). Three different genetic testing scenarios were presented to assess perspectives regarding which genes (based on clinical validity classification) to report from a clinical sequencing panel for 1) preconception carrier screening, 2) genetic diagnosis of a phenotype for which multiple genes have been implicated, and 3) testing of family members for an adult-onset disorder with invasive management. For each scenario, participants were asked to indicate their level of agreement on reporting genes with different clinical validity classifications. There was 100% agreement for the inclusion of gene-disease pairs with a "Definitive" classification and >95% agreement for "Strong" classifications in all three scenarios. Inclusion of gene-disease pairs with a "Moderate" classification was more accepted for diagnostic panel testing (83%) and family testing (72%), than for preconception carrier testing (47%). For "Limited" evidence genes, 44-76% felt that these genes should not be reported, depending on the scenario. Most respondents indicated that genes with "No evidence" or "Conflicting evidence" should not be reported in any of the scenarios (>77%). Regarding who should participate in making official recommendations on which genes to include on clinical sequencing test reports, the top categories were "experts in disease area" (100%), "ClinGen" (91%), and "testing laboratories" (83%). We will describe our results in detail, including limitations and future directions. The results of this survey may be used to guide ClinGen policies, and for informing the design of future research studies involving the wider genetics community on developing guidance for the return of results. These may also be useful for laboratories as they design genomic tests.

Further progress in the assessment of genes' associations with the RASopathies

Andrew R. Grant[1], Hélène Cavé[2], Mitchell W. Dillon[3], Bruce D. Gelb[4], Karen W. Gripp[5], Jennifer A. Lee[6], Heather Mason-Suares[7], Katherine A. Rauen[8], Marco Tartaglia[9], Lisa M. Vincent[10], Martin Zenker[11]

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts
[2]Département de Génétique, Hôpital Robert Debré and Institut Universitaire d'Hématologie, Université Paris Diderot, Paris-Sorbonne-Cité, Paris, France
[3]Molecular Genetic Testing Laboratory, Icahn School of Medicine at Mount Sinai, New York City, New York
[4]Departments of Pediatrics and Genetic and Genomic Sciences, Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York City, New York
[5]Department of Pediatrics, Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware
[6]Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, South Carolina
[7]Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, Massachusetts
[8]Department of Pediatrics, UC Davis Children's Hospital, Sacramento, California
[9]Ospedale Pediatrico Bambino Gesù, Rome, Italy
[10]GeneDx, Gaithersburg Maryland
[11]Institute of Human Genetics, University Hospital Magdeburg, Magdeburg, Germany

Disputing a gene's association with a disease domain is important for the processes of developing genetic testing panels and classifying a variant's contribution to a patient's disease. Following a large effort to formally assess the association of 19 genes with the RASopathies using the ClinGen gene curation framework (Grant et al. 2018 30311384, Strande et al. 2017 28552198), the ClinGen RASopathy Gene Curation Expert Panel (RAS GCEP) disputed two non-historical RASopathy genes recently linked to the disease domain (*KAT6B* and *NSUN2*) in addition to the previously disputed *A2ML1* and *RASA1* genes. In disputing these four genes' associations with the RASopathies, we established a clearer understanding of the genes that should, and should not, be linked to the RASopathies based on clinical presentations and the impact on the RAS-MAPK pathway. This distinction is important for diagnostic testing and diagnoses of patients with features that may be consistent with a RASopathy.

The RAS GCEP's reasons for disputing gene-disease links are variable. In the case of *A2ML1*, few variants have been reported in patients with disease, and all the variants that have been described in patients with Noonan syndrome have been frequent in the general population and/or lack evidence for pathogenicity and were not scored by the RAS GCEP (gnomad.broadinstitute.org, Vissers et al. 2015, van Trier et al. 2015). Alternatively, patients with *RASA1* alteration who possessed arteriovenous features suggestive of RASopathies possessed genetic causes secondary to *RASA1* variation (Illari, Agosta, & Bacino, 2016; Macmurdo et al. 2016).

The *KAT6B* and *NSUN2* genes were altered in patients described with a "Noonan-like" disease and functional/animal models indicating a possible link to RASopathies (Kraft et al. 2011 21804188, Martinez et al. 2012 22577224, Fahiminiya et al. 2014 24102521, Clayton-Smith et al. 2011 22077973, Blanco et al. 2011 22144916). However, the evidence indicated inconsistent phenotypic overlap with RASopathies and the experimental models lacked specificity to RASopathies. Therefore, *NSUN2* was curated for complex neurodevelopmental disorder (MONDO:0100038) by the ClinGen ID/Autism GCEP, and *KAT6B* will be curated for Say-Barber-Biesecker-Young-Simpson syndrome at a later date.

Curating the Clinical Genome Abstract: RASopathy GCEP Update

This work sets precedents for the different ways gene curation groups can dispute gene-disease assertions in the future and provides useful insight for non-historical genes in the RASopathy disease domain.

Large whole genome sequencing datasets facilitates curation of reported variants and detection of unreported changes: Gabriella Miller Kids First Data Resource Center as an example

Yiran Guo[1], Yuankun Zhu[1], Allison P. Heath PhD[1], Bo Zhang[1], Miguel Brown[1], Krutika Gaonkar[1], Pichai Raman PhD[1], Jena Lilly MS[1], Deanne M. Taylor PhD[1,7], Phillip B. Storm MD[1,9], Angela J. Waanders MD[1,7], Vincent Ferretti PhD[2,10], Michele Mattioni PhD[3], Brandi Davis-Dusenbery PhD[3], Zachary L. Flamig PhD[4], Robert L. Grossman PhD[4], Samuel L. Volchenboum MD PhD[4,8], Sabine Mueller MD PhD[5], Javad Nazarian PhD[6], Nicole Vasilevsky PhD[11], Melissa Haendel PhD[11,12], Adam Resnick PhD[1,9]

[1]Children's Hospital of Philadelphia, Philadelphia, PA, USA; [2]Ontario Institute of Cancer Research, Toronto, Ontario, Canada; [3]Seven Bridges Genomics, Cambridge, MA, USA; [4]Center for Research Informatics, The University of Chicago, Chicago, Illinois, USA; [5]University of San Francisco, San Francisco, CA, USA; [6]Children's National Medical Center, Washington, DC, USA; [7]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [8]Department of Pediatrics, Biological Sciences Division, The University of Chicago, Chicago, Illinois, USA; [9]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [10]Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada; [11]Oregon Health & Science University, Portland, Oregon, USA; [12]Oregon State University, Corvallis, Oregon, USA.

Gene panels, as well as whole exome sequencing (WES), have been widely used in the genetic diagnosis in rare disorders and cancer. Previous publications of pathogenic/likely pathogenic genetic changes, e.g. those cataloged in the Human Gene Mutation Database® (HGMD), are an essential reference for determining if variants identified in clinical practice are related or even causal to disease presentations. While the majority of the published changes are exonic (87.1%, or 137,815 of the total 158,223 variants labeled as DM[?]/[likely] pathological in the 2018Q4 professional version of HGMD) plus splicing (10.9% or 17,230), the non-coding variants, especially intronic ones published in the literature (1.1% or 1,501) are significantly less likely to be charted by clinical labs such as ClinVar submitters or reviewed by curators like the ClinGen Expert Groups. In the HGMD, 63.7% (87,845 out of 137,815) of exonic variants are not submitted/reviewed (according to ClinVar version 20180603), and the numbers are significantly higher as 70.0% (12,056 of 17,230; $p<1e-15$) and 69.2% (1,039 of 1,501; $p<1e-15$) for splicing and intronic changes, respectively. Previous pilot efforts with small sample sizes showed that whole genome sequencing (WGS) can provide more insight about non-coding regions of the human genome, increase genetic diagnostic rate in clinical settings, facilitate curation of reported variants, and discover novel genomic changes responsible for the clinical manifestations that could be missed by panel sequencing or WES.

Implementing WGS, the NIH Common Fund Gabriella Miller Kids First Pediatric Research Program (Kids First) represents a national collaboration focused on large-scale genomic and clinical data sharing for childhood cancers and structural birth defects, which present a huge burden to families and the healthcare system. As part of this program, the Gabriella Miller Kids First Data Resource Center (DRC) empowers collaborative discovery across Kids First datasets. Through newly developed platforms and cloud-based resources, researchers are able to access standardized genomic and clinical data in a timely fashion. Datasets harmonized by the Kids First DRC also enable curation of reported variants and discovery of new genetic changes, both coding and non-coding. In total there will be data for >27,600 whole genomes

and currently, ~30,000 genomic/phenotypic files are available on the Kids First Data Resource Portal, making it one of the largest pediatric data resources across a wide range of diseases.

New Avenues of Inquiry: Visualizing OMIM Phenotype-Gene Relationships at OMIM.org

Ada Hamosh[1], Carol A. Bocchini[1], Cassandra K. Arnold[1], François Schiettecatte[2], Marla M. J. O'Neill, Anne Stumpf[1], Joanna Amberger[1]

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine
[2]FS Consulting, LLC

The relationship between diseases and genes is a rapidly growing and complex body of information. As of February 2019, OMIM has over 24,900 entries describing genes and phenotypes and has compiled the molecular relationship between over 4,000 genes and 6,300 phenotypes. In addition to curating these relationships, OMIM has been centrally involved in the classification and naming of genetic phenotypes. In the same way that grouping genes of similar function or structure by name opens a window to biology, grouping similar phenotypes using clinical naming provides a clinical view of the human genome. Clinical naming as an organizing principle of OMIM permits collection and comparison of the genes underlying clinically related phenotypes (Phenotypic Series). OMIM has more than 440 Phenotypic Series comprising more than 3,600 disorders. To facilitate molecular genetic inquiry, OMIM has leveraged this classification of genetic disease to create visualizations of phenotype-gene relationships. These PheneGene graphics are available in linear and radial formats. Nodes represent phenotypes, genes, and Phenotypic Series; branching occurs when a gene causes more than one phenotype or when a phenotype is genetically heterogeneous. These views can be altered to place a different node at the center, thus illuminating further connections or placing emphasis on a particular gene or phenotype. The graphics are created on-the-fly from current OMIM data. No hierarchy is implied with either graphic format. These dynamic views of the connections between phenotypes and genes may suggest novel biological or clinical relationships. Clinical naming of genetic disorders enhances analysis of the relationship between phenotypes and their molecular basis, and provides new avenues of investigation through increased understanding of the molecular basis of disease.

A de novo deletion of 3p21.31 disrupting *SETD2*

Domagoj Hodko[1], Loretta W. Mahon[1], Shailesh M. Asaikar[2], and Sibel Kantarci[1*]

[1] Quest Diagnostics Nichols Institute, San Juan Capistrano, CA, USA

[2] Sutter Medical Center, Sacramento, CA, USA

[*]Email: Sibel.X.Kantarci@questdiagnostics.com

The SET domain-containing protein 2 (SETD2) [MIM: 612778] is involved in genome-wide transcriptional activation and functions in methylating histone H3K36, a hallmark of actively transcribed chromatin. SETD2-dependent H3K36 trimethylation facilitates homology-directed DNA repair and maintains genome stability. Heterozygous pathogenic variants in histone methyltransferases are associated with overgrowth and developmental disorders; for example, histone methyltranferase NSD1 deficiency is associated with Sotos syndrome. Similarly, several heterozygous sequence variants in *SETD2* have been reported in patients with Luscan-Lumish syndrome (LLS), which is characterized by overgrowth, macrocephaly, autism spectrum disorder, intellectual disability, developmental delay, and seizures [MIM# 616831].

Several heterozygous variants, including 4 de novo loss-of-function (LoF) variants, have been reported in patients with *SETD2*-related phenotypes. The Clinical Genome Resource (ClinGen) compiled some evidence for dosage pathogenicity (haploinsufficiency score of 2) for *SETD2*. In addition, the Exome Aggregation Consortium (ExAC) database reported a Probability of LoF Intolerance (pLI) score of 1.00, which predicts extremely high LoF-intolerance for SETD2.

Here, we report a de novo 114-kb deletion of 3p21.31 using SNP-based chromosomal microarray analysis in an infant male with macrocephaly and entropion of the right eye. This deletion interval includes a noncoding gene and several exons of the 5' portion of *SETD2*. This deletion interval does not overlap with any copy number losses in the Database of Genomic Variants (DGV) or in our internal database.

To our knowledge, this is the first report of a de novo partial deletion of SETD2 that is expected to disrupt gene expression. Larger deletions of 3p21.31 encompassing multiple genes, including *SETD2*, have been rarely reported in the literature; however, smaller deletions encompassing only SETD2 have not been reported yet. Thus, our finding further supports the association of SETD2 haploinsufficiency with macrocephaly.

Data sharing, supported decision making and clinical duties in genomics

Amalia M. Issa, PhD, MPH, FCP
University of the Sciences in Philadelphia

We currently live in the era of data sharing. Certainly patients and consumers in general are calling for more data sharing of genomic and other clinical, family and health-related data, as well as through apps, mobile and digital health. Indeed, it seems that we are rapidly entering a disruptive shift in the culture of biomedicine toward greater data sharing with patients to the degree that the patient views him or herself as the owner of all the data and the decisions associated with the data. Such new technological and social trends are illustrative of the increasing complexity related to a the debate about the duty of clinicians to re-contact patients as the meaning of data is updated and changes over time. From a clinical perspective, any duty to re-contact patients needs to be considered within the broader question of when the clinician-patient relationship begins and ends. The new and emerging technological trends delineated above call into question whether there is or ever can be a definitive end to the clinician-patient relationship. To date, certain circumstances, such as when a patient moves to another locale or when a pediatric patient reaches the age of majority and needs to find another physician for care, dictates the end of the clinician-patient relationship. However, in the context of the new precision medicine ecosystem where new interpretations of genomic data and previously collected and stored DNA sequences are possible, it may be that a physician's moral and legal responsibility (and therefore a duty to re-contact and update knowledge) to a particular patient endures for the remainder of the patient's natural life. If there is a duty to re-contact, should it be governed differently under different circumstances (e.g. pediatric cases until the age of majority vs adult patients)? Should limits be placed on the duty to re-contact depending on the type of information, with the clinician acting as final arbiter (i.e. for new interpretations from stored genomic data that provides a possible new diagnosis but with no known interventions or treatments)? The objective of this presentation is to examine how novel and emerging social trends and technologies for clinical decision support might influence clinician duties and obligations, and the potential effect on genomics, precision medicine and data sharing, and what might be some of the regulatory governance options.

Assessment of VCEP Functional Study Specifications and Development of Structured Framework for Evidence Curation

Meera K. Jairath[1], Dona M. Kanavy[1], Shannon M. McNulty[1], Sarah E. Brnich[1,2], Chris Bizon[3], Bradford Powell[1], Jonathan S. Berg[1]
[1]Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[2]Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[3]Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

The Clinical Genome Resource (ClinGen), since its inception in 2011, has established sixteen Variant Curation Expert Panels (VCEPs). The goal of these VCEPs is to make gene-level modifications to the 2015 ACMG/AMP guidelines to improve the evaluation of evidence for variant interpretation through completion of a four-step approval process. As of 2019, six VCEPs have been fully approved and their recommendations have been published. We are particularly interested in the disease-specific modifications VCEPs have made to the functional study guidelines (PS3/BS3) specified by the ACMG/AMP. These guidelines broadly approve any “well-established” functional assay, but this terminology is not well defined. VCEPs have interpreted the meaning of “well-established” for their disease area by identifying specific assays and validation parameters that are sufficient to provide evidence at a strong, moderate, or supporting level for classifying a given variant. To determine the attributes VCEPs identified as important for functional assays used to support PS3/BS3 assertions, we assessed whether each VCEP specified requirements for the following parameters: replication, controls, thresholds, and validation. In our assessment of their recommendations, we observed a high level of variation in the frequency and methods by which these parameters were specified, with most VCEPs only detailing a need for one to two of these parameters to be fulfilled by an individual instance of a functional assay. Given this variation, we developed a standardized framework for assessing VCEP recommendations and curating evidence provided by functional assays for clinical variant interpretation. This includes identifying the disease mechanism and general class of assay used and cataloguing details of the specific instance of assay being performed using a structured ontology. We used this framework to analyze the primary literature cited by each individual VCEP. We anticipate that application of a standardized approach for evaluating the clinical validity of functional assays and variant-level functional evidence will lead to greater consistency between VCEPs. The results from this work will also be used to inform a SEPIO-compliant data model to structure evidence regarding functional data in the Variant Curation Interface.

RefSeqGene (RSG) and Locus Reference Genomic (LRG) resources for reporting of clinically relevant sequence variants

K. McGarvey[1], V. Joardar[1], J. Morales[2], A. Astashyn[1], R. Bennett[2], C. Davidson[2], A. Frankish[2], L. Gil[2], M. Kay[2], J. Loveland[2], A. McMahon[2], G. Threadgold[2], F. Cunningham[2], Murphy[1]

[1]National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, U.S.A. 20894
[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD

RefSeqGene (RSG) is a subset of NCBI's Reference Sequence (RefSeq) project that defines manually curated genomic sequences to be used as reference standards for well-characterized genes (https://www.ncbi.nlm.nih.gov/refseq/rsg/). Each RSG sequence provides gene-specific genomic sequence, including intronic regions and upstream and downstream flanking regions, and is annotated with one or more representative transcripts as reference standards. These sequences serve as a stable foundation for reporting mutations, establish conventions for exon- and intron-numbering, and for defining the coordinates of other sequence variations. RSG sequences are versioned and may be updated, but the sequences are independent from upgrades to the reference genome assembly and may represent better supported alleles, making them useful for consistent variant reporting over time. A subset of RSGs form the foundation of the Locus Reference Genomic (LRG) project, a manually curated resource designed specifically for the reporting of clinically relevant variants (http://www.lrg-sequence.org/). LRG records, created jointly by the NCBI (https://www.ncbi.nlm.nih.gov) and EMBL-EBI (http://www.ebi.ac.uk), are similar to RSGs but are not versioned and have fixed sequence, and include additional annotation. Transcripts for which there is currently good biological understanding are selected for inclusion, with input from members of the clinical research community. The number of records available for community use currently stands at 6,847 for RSG, and >1,260 for LRG. The RSG and LRG resources are coordinating with the collaborative Matched Annotation from NCBI and EMBL-EBI (MANE) project to facilitate convergence on key high value annotations to ensure identity between RefSeq and Ensembl transcripts, including those used as reference standards for RSGs and LRGs. However, MANE transcripts must match the GRCh38 genome sequence, whereas RSGs and LRGs can be used to represent alternate alleles where the GRCh38 sequence is defective or otherwise not suitable for clinical reporting. Thus, RSGs and LRGs serve as valuable clinical reporting standards especially for individual genes, complementing the genome-wide MANE project.
This work was supported in part by the intramural research program of the National Library of Medicine, National Institutes of Health.

Toward complete interpretation of known *RYR1* variants

J.J. Johnston[1], D. Ng[1], S.G. Gonsalves[1], L.G. Biesecker[1,2]
[1]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
[2]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

RYR1 is one of the 59 genes recommended by the American College of Medical Genetics and Genomics (ACMG) for opportunistic screening and it is recommended that only known pathogenic (KP) variants are to be returned. As is true for several cardiomyopathy genes, nearly all variants that cause this trait are amino acid substitutions and many are rare. Interpreting such variants is challenging, as evidenced by the lower overall concordance of interpretation for cardiomyopathy variants. To address this challenge, we set out to work toward a complete interpretation of known *RYR1* variants, which could be made available to the genetics community. We set out to do this through review by a ClinGen expert panel including medical geneticist, clinical molecular geneticists, physiologists, and anesthesiologists. The ACMG-AMP criteria for variant interpretation were modified for this analysis based on the biology and genetics of *RYR1* and Malignant Hyperthermia Susceptibility (MHS). Interpretation of functional data for PS3/BS3 necessitated creation of a decision tree placing higher value on functional studies that isolate identified *RYR1* variants from other alterations potentially present in patient cells. Minor allele frequency (maf) values for BA1/BS1 were set at 0.004 and 0.0004, respectively, based on disease prevalence, contribution of *RYR1* to MHS and minor allele frequency in gnomAD for the common pathogenic variant p.Arg614Cys. Criteria omitted from the analyses included: PVS1, PM4 and BP1, as the majority of variants associated with MHS are missense alterations; PM3, as MHS is inherited as a dominant trait; PP2 as the gnomAD missense constraint metric (z-score=2.34) fell below our cutoff of 3.0; PP4, as alterations in *CACNA1S* have been identified as causative for MHS; BS2, as MHS is not fully penetrant; and BP7, as our analysis was restricted to coding variants. The Bayesian calculator of Tavtigian et al will be used to give a posterior probability for pathogenicity and the variants will be classified using the five tier system from pathogenic to benign. Variant interpretations will be reviewed by a ClinGen expert panel and finalized assessments will be made available to the genetics community through ClinVar. Access to variants curated by this process will allow more efficient review of secondary variants and support improved interpretation of *RYR1*/MHS genomic testing results and improve the ability to implement opportunistic screening for *RYR1*.

Pathogenicity Predictions Using Multiple Sequence Alignment-Based Algorithms For *APC* Missense Variants Reported in ClinVar

Alexander Karabachev[1], Dylan Martini[1,2], David Hermel[1,3], Dana Solcz[1], Indra Neil Sarkar[4], Marc Greenblatt[1], University of Vermont, Burlington VT[1]; Emory University, Atlanta, GA[2]; Keck School of Medicine of USC[3], Los Angeles, CA; Brown University, Providence, RI[4]* Marc.Greenblatt@uvmhealth.org

Aim:

Variants in the *APC* gene cause Familial Adenomatous Polyposis. Most are non-missense, protein-truncating changes. However, 40% of the 4,891 total variants in the ClinVar database are reported as missense. The total number of missense variants increased from 47 in 2013 to 1,988 in 2018. We used Protein Multiple Sequence Alignment (PMSA)- based methods to assess the proportion of missense *APC* variants that are Pathogenic/Likely Pathogenic (P/LP), Benign/Likely Benign (B/LB), or Variants of Uncertain Significance (VUS); whether altered APC splicing may be a mechanism of P/LP; and how PMSA-based *in silico* tools predict the effects of missense variants.

Methods:

We created PMSAs for APC of different evolutionary depths. We curated the APC missense variants that have assertions of P/LP and B/LB in ClinVar. We used 5 PMSA-based methods (Align-GVGD, SIFT, PolyPhen2, MAPP, REVEL) to generate pathogenicity predictions for *APC* missense variants.

Results:

An *APC* PMSA using 10 species (evolutionary depth through sea urchin), which required significant curation to optimize alignment, achieved our threshold for statistical significance for the conclusion of “absolute amino acid conservation”. Among the reported missense variants in ClinVar in July 2018, 91.5% are reported as VUS. All but 2 missense variants listed as P/LP occur at canonical splice sites or Exonic Splice Enhancer sites. *In silico* tools differed widely in predictions of P/LP (range 17.5–75%) and B/LB (range 25–82.5%) for all *APC* missense variants found in ClinVar. When applied only to the 22 variants that have been classified in ClinVar as Benign, the 5 methods ranged in accuracy from 78-100%. Some methods favored predictions of B/LB and some favored P/LP.

Conclusions:

PMSA-based *in silico* methods can be excellent classifiers for variants of some hereditary cancer genes. However, missense classifiers are not appropriate for variants in *APC,* and likely other hereditary cancer genes, in which missense is not a known mechanism of pathogenicity. Some variants in ClinVar that are classified as missense actually affect splicing. *In silico* missense classifiers vary widely when applied to the large set of *APC* VUS. There may be characteristics of the *APC* gene and protein that confound the results of *in silico* algorithms. A systematic study of these features could greatly improve the use of predictive algorithms in hereditary cancer genes and improve assessment of the health risks of inherited genetic variation.

From ACMG/AMP framework specifications to scaling expert-panel approved ClinVar submissions: pilot and perspectives of ClinGen's Cardiomyopathy Variant Curation Expert Panel

Melissa A Kelly[1], Christina Austin-Tse[2,3], Hana Zouk[3,4], C. Lisa Kurtz[5], Linnea Baudhuin[6], Charlotte Burns[7], Patricia Harper[8], Jodie Ingles[7], Michelle Kluge[6], Katrina Kotzer[6], Dani Macaya[9], Heather McLaughlin[9], Mitzi Murray[9], Emily Qian[10], Audrey Schaffer[8], Matthew Sdano[11], Matteo Vatta[12], Jessica Woodley[13], Heather Zimmermann[14], Birgit Funke[3,10], ***on behalf of the ClinGen Cardiomyopathy Variant Curation Expert Panel***

1. Genomic Medicine Institute, Geisinger – Danville, PA USA
2. Center for Genomic Medicine, Massachusetts General Hospital – Boston, MA USA
3. Harvard Medical School – Boston, MA USA
4. Partners HealthCare Personalized Medicine – Cambridge, MA USA
5. UNC Chapel Hill – Chapel Hill, NC USA
6. Department of Laboratory Medicine and Pathology, Mayo Clinic – Rochester, MN USA
7. Molecular Cardiology Program, Centenary Institute – Sydney, NSW AUSTRALIA
8. Genetics Diagnostic Laboratory, CHEO – Ottawa, Ontario CANADA
9. GeneDx – Gaithersburg, MD USA
10. Veritas Genetics – Danvers, MA USA
11. ARUP Laboratories – Salt Lake City, UT USA
12. Invitae – San Francisco, CA USA
13. West Midlands Regional Genetics Laboratories, Birmingham Women's and Children's NHS Foundation Trust – Birmingham UNITED KINGDOM
14. Ambry Genetics – Aliso Viejo, CA USA

Variant curation is challenging to standardize due to differences in data sources, rule interpretation, experience, and clinical judgement. Adapting the ACMG/AMP variant interpretation framework to increase its specificity for diseases and genes is expected to increase the reproducibility of variant classifications. The Cardiomyopathy Variant Curation Expert Panel (CMP-VCEP) has implemented a derivative classification framework for *MYH7* and is now establishing a sustainable variant curation effort to scale submission of 3-star status variants to ClinVar. Assembling a streamlined variant curation process requires standardization and well-trained curators at minimum. The CMP-VCEP relies on a network of 10 cardiomyopathy testing laboratories that each use their established infrastructure to curate *MYH7* variants using the adapted framework. However, such a networked approach is vulnerable to introducing variability. To assess the robustness of the *MYH7*-specific rules and determine implications for scaling variant curation using this network, we undertook dual curation of a set of priority *MYH7* variants with a focus on those with conflicting interpretations in ClinVar.

In >75% of variants, the final classification and/or the rules applied differed between curators and even when curators agreed on the final classification, the majority applied rules differently. Of the rules applied, >90% differed in usage in at least 1 variant, while >70% differed in usage for ≥50% of variants. Concordance increased with the frequency of rule usage. Reasons for discordance included deviation from the calculated classification based on clinical judgment, different sources of evidence, availability of internal data, incorrect rule application, and transcription errors. These data suggest that increased disease/gene specificity of the rules do not reduce variability and focus attention on process management to achieve scale. Towards this goal, the following areas need to be addressed to maximize throughput while maintaining quality and minimizing the amount of time spent on variant review by senior leadership: (1) assembling case data (literature and internal laboratory data) prior to curation, (2) using a standard curation process per gene or disease area that includes the ClinGen Variant Curation Interface (VCI), (3) continual re-education on complex clinical scenarios and difficult rules and (4) the addition of a senior level review to ensure compliance with rules and processes.

Character limit: 2,500 (including spaces)
Character count: 2,500 (including spaces)

Identification of Novel Mutations of Myosin 7A from families with different Mendelian Inheritance of Hereditary Hearing Loss

Hye-Min Kim[1,2], Nari Ryu[1,2], Byunghyeon Lee[1,2], Ye-Ri Kim[1,2], Min-A Kim[1,2], Deok-Gyun Choi[1,2], Hong-Joon Park[3], Un-Kyung Kim[1,2]

[1]Department of Biology, College of Natural Sciences, Kyungpook National University, Daegu, Korea

[2] School of Life Sciences, KNU Creative BioResearch Group (BK21 Plus Project), Kyungpook National University, Daegu, Republic of Korea

[3] Soree Ear Clinic, Seoul, Korea

Abstract

Hereditary hearing loss is one of the most common genetically heterogeneous defects in human. About 70% of hereditary hearing loss is defined as non-syndromic hearing loss showing loss of hearing ability without any other symptoms. Up to date, the identified genes associated with non-syndromic hearing loss are 115, including 45 genes for DFNA and 73 genes for DFNB. Because of high levels of heterogeneity, it is difficult to identify the causative factors for hearing loss using Sanger sequencing. Therefore, we used next generation sequencing technique to investigate whole-exome sequences of two Korean families with non-syndromic hereditary hearing loss. One of the families showed autosomal dominant inheritance pattern of hearing loss, and the other showed autosomal recessive inheritance pattern. Our aim was to detect causative factors and investigate pathogenic mutations, which co-segregates within the candidate families. As a result, we identified a novel missense variation, c.1978G>A in *MYO7A* gene, in the family with the autosomal dominant inheritance pattern and a novel compound heterozygote mutation, c.4501G>A and c.6070C>T in *MYO7A* gene, in the family

with the autosomal recessive inheritance pattern. c.1978G>A produced p.Gly660Arg in the motor head domain of Myosin VIIA disrupt the ATP- and actin-binding motif function. c.4501G>A and c.6070C>T produced p.Val1501Met and p.Arg2024Stop, respectively in the first and second MyTH4 domains. This study is the first to report pathogenic mutations within *MYO7A* gene in Korean families and our data would facilitate diagnosing the primary cause of hereditary hearing loss in Korean.

Title: Evaluation of the Accuracy and Clinical Utility of Variant Interpretations Generated by Artificial Intelligence

Authors: Afia Hasnain, Gregory Downs, Natalie Stickle, Rong Zeng, Tracy L. Stockley, Ian King

Affiliations (all): Advanced Molecular Diagnostic Laboratory, Princess Margaret Cancer Center, University Health Network, Toronto, Canada

The time, effort and expertise needed to perform variant Interpretation and curation is a limiting factor in clinical genomics for both inherited disease and cancer. In recent years tools using artificial intelligence (AI) to perform some or all aspects of variant interpretation for clinical reporting have come to market, and these tools have been adopted by some clinical laboratories as a solution to the scaling of variant interpretation, especially for cancer testing. However, the accuracy and clinical utility of AI driven variant interpretation has not been systematically evaluated.

We have compared somatic variant interpretations generated by the IBM Watson for Genomics platform to clinically reported interpretations for 800 cases of non-small-cell lung cancer and 1000 cases of hematological malignancy. Both the clinically reported and AI-generated interpretations were organized into the Minimum Variant Level Data format to enable direct comparison between the datasets. Comparisons were made for approximately 5000 variants (~1500 from the lung cancer cases, ~3500 from the haematological cases). Allele descriptive and allele interpretive interpretations were compared for all variants in the dataset. Some somatic interpretive data were also compared – variants predictive of therapy response were assessed for lung cancer cases and prognostic factors were assessed for haematological cases.

Agreement between AI-generated and clinically reported interpretations was high for descriptive categories. However, disagreement was common in allele interpretive categories. 7 genes, accounting for 17% of all variants, were assigned to different RefSeq transcripts in the haematological case set. As a result, DNA substitution interpretations were discordant in 10% of variants, and protein substitutions were discordant in 8%, leading to clinically significant differences in interpretation of variant pathogenicity. Interpretive disagreements were also observed in gene-disease association, and in interpreting cases with multiple variants of prognostic significance in acute leukemia.

AI based interpretation of somatic variants was largely accurate and extremely rapid (~2 minutes per case), demonstrating a basic clinical utility as an aid to variant analysis. However, significant discrepancies between AI-generated and clinically reported interpretations suggest that they are not yet appropriate as a stand-alone interpretation tools.

Development of the ClinGen Variant Curation Training Process

C. Lisa Kurtz[1], Sarah Brnich[1], Marina DiStefano[2], Steven Harrison[3], Jessica Mester[4], Julianne O'Daniel[1] Natasha Strande[1], Karen E. Wain[5], Danielle Azzariti[3] and Erin Rooney Riggs[5]

1. University of North Carolina at Chapel Hill, Department of Genetics, Chapel Hill, NC
2. Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA
3. The Broad Institute of Harvard and MIT, Cambridge, MA
4. GeneDx Laboratories, Gaithersburg, MD
5. Geisinger, Autism & Developmental Medicine Institute, Lewisburg, PA

The Clinical Genome Resource (ClinGen)'s Expert Panel pathogenicity assertions in NCBI's ClinVar database is the first public dataset to be recognized by the FDA as a source of clinically valid variant assertions supported by scientific evidence. As part of this recognition, the FDA approved a protocol for variant curation expert panels (VCEPs) to attain 3 star review status, including a step-wise application, variant evaluation, and reanalysis/discrepancy resolution processes. Integral to the accuracy and fidelity of the process are the biocurators tasked with assembling and analyzing available evidence and providing a preliminary pathogenicity assertion for review by a designated VCEP. To maintain standardization and proficiency in variant classification across all VCEPs, the ClinGen Education, Coordination and Training Working Group (ECT WG) developed a tiered approach to variant curation training. Level 1 training provides guidance on variant classification using ClinGen-approved protocols and tools for all biocurators. Level 2 training builds on this by introducing demonstrable proficiencies for VCEP-specific biocurators.

The ECT WG assessed ClinGen resources for variant curation and identified areas for training development. Curation tools such as the Variant Curation Interface (VCI) and ClinGen Allele Registry were in place, but lacked an organized general sequence variant curation standard operating procedure (SOP) for use. A committee consisting of clinical molecular geneticists, genetic counselors and senior biocurators developed this SOP, which was reviewed and approved by the ClinGen Genomic Variant WG. Level 1 training consists of a 90 minute web conference providing a tutorial on use of the VCI and review of the new variant curation SOP, the ACMG/AMP variant classification guidelines, and videos on literature searching and use of the Allele Registry. Level 2 training is VCEP-specific, focusing on disease/gene specifications to the ACMG/AMP guidelines. Proficiency will be assessed for three core competencies: evidence identification, curation, and assessment, and will be overseen by a senior biocurator. Here we detail our comprehensive training process and resources, and identify metrics to support the utility of this method. This tiered training process ensures the standardization and oversight required for continued FDA recommendation and aids in the effort to scale variant curation by providing a larger pool of qualified biocurators.

ClinVar and Assertion Criteria: Providing Transparency for Rules Used for Variant Classification

Melissa J Landrum, Shanmuga Chitipiralla, Garth Brown, Baoshan Gu, Jennifer Hart, Vitaly Lyoshin, Rama Maiti, George Riley, George Zhou, Donna Maglott, J Bradley Holmes, Brandi L Kattman

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

The ClinVar database at the National Institutes of Health currently holds more than 782,000 submitted interpretations for almost 500,000 unique variants. Data in ClinVar are submitted by clinical testing laboratories, research laboratories, curation groups, expert panels, and others. Many organizations that submit to ClinVar base their classification of variants on the 2015 guidelines from the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (PubMed ID 25741868). Some organizations have customized these rules, as suggested in the original recommendation; other organizations use a different set of rules, such as rules for copy number, pharmacogenetic, or somatic variants. ClinVar encourages submitters to provide their classification rules, known as “assertion criteria” or “assertion method”, as part of their submission. As of January 1, 2019, 44% (488/1119) of the organizations that submit to ClinVar had provided this documentation, namely 55% (285/563) of laboratories submitting data from clinical testing and 44% (239/543) of laboratories submitting data from research. Assertion criteria are required for submissions from expert panels and practice guidelines. The transparency that assertion criteria provides is valuable to ClinVar users, and so the availability of assertion criteria contributes to a record’s rating, or review status, which is depicted graphically on the ClinVar web site with gold stars. Variants with one or more stars have at least one submission with assertion criteria (https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/#revstat_web), along with supporting evidence. Web users can filter their search results for a specific review status, or by selecting “one or more stars”. Users who download one of ClinVar’s data files can also filter variants of interest using the text values of review status. Submitters who wish to include assertion criteria in their ClinVar submission can cite a publication or upload their documentation in Word or PDF format(https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/#ac). The ClinVar team is working on improvements to make submission of assertion criteria easier, including the ability to cite more than one document for assertion criteria in a single submission spreadsheet.

Title:

Establishing a Framework for Classification and Reporting of Risk Alleles and Low-Penetrance Variants: Progress from the ClinGen Low-Penetrance/Risk Allele Working Group

Authors and Affiliations:

Matthew S. Lebo, PhD[1], Ozlem Senol-Cosar, PhD[2], Erin Rooney Riggs, MS, CGC[3], Erica F. Andersen, PhD[4], Pinar Bayrak-Toydemir, MD, PhD[5], Mythily Ganapathi, PhD[6], Saurav Guha, PhD[7], Vaidehi Jobanputra, PhD[8], Deqiong Ma, MD, PhD[9], Emily Qian, MS, CGC[10], Heidi L. Rehm, PhD[11], Ryan J. Schmidt, MD, PhD[12] (on behalf of the ClinGen Low-Penetrance/Risk Allele Working Group)

[1] Laboratory for Molecular Medicine, Partners HealthCare, Boston, MA; Department of Pathology, Brigham and Women's Hospital; Broad Institute of MIT and Harvard, Cambridge, MA

[2] Laboratory for Molecular Medicine, Partners HealthCare, Boston, MA, Department of Pathology, Brigham and Women's Hospital, Boston, MA

[3] Autism & Developmental Medicine Institute, Geisinger, Lewisburg, PA, USA

[4] University of Utah, ARUP Laboratories, Salt Lake City, Utah, USA

[5] University of Utah, ARUP Laboratories, Salt Lake City, Utah, USA

[6] Department of Pathology & Cell Biology, Columbia University Medical Center, New York, NY, USA

[7] New York Genome Center, New York, NY; past affiliation: Myriad Women's Health (formerly Counsyl Inc), South San Francisco, CA

[8] New York Genome Center and Department of Pathology & Cell Biology, Columbia University Medical Center

[9] DNA Diagnostic Lab, Department of Genetics, Yale University, New Haven, CT

[10] Veritas Genetics, Danvers, MA, USA

[11] Broad Institute of MIT and Harvard, Cambridge, MA; Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

[12] Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Keck School of Medicine of USC, Los Angeles, California, USA

There are well-recognized challenges in interpreting variants in genes associated with Mendelian disease that have significantly reduced penetrance or small effect sizes, such as p.Ile200Thr in *CHEK2*, p.Ile1307Lys in *APC*, and recurrent 16p13.11 deletions. Risk alleles (those with very low penetrance or small effect sizes) and low-penetrance variants (those with reduced but appreciable penetrance) represent a significant source of the interpretation discrepancies in ClinVar. These types of variants often have population frequencies that are incompatible with highly penetrant Mendelian disease despite genetic evidence suggesting an association with disease. Further, as there is no established terminology to describe the role these variants may play in disease, laboratories are describing them differently, leading to confusion between laboratories, clinicians, and patients.

The current ACMG/AMP germline sequence variant interpretation guidelines do not address whether low-penetrance Mendelian variants should be a separate category that warrants a distinct classification or terminology framework but acknowledge that additional guidance covering interpretation of “risk alleles” would be an important area of future standards development. ClinGen’s Low-penetrance/Risk Allele Working Group aims to address this problem by 1) developing consensus on the terminology needed to categorize both risk alleles and low-penetrance Mendelian variants and 2) developing a standardized classification framework to evaluate these types of variants.

As a first step toward defining the terminology, thresholds, and data that will be used to create a standardized classification framework, we have developed a Delphi survey that will be deployed to the ClinGen community (currently over 800 individuals) to understand community preferences surrounding risk alleles and low-penetrance variants. We will present the current progress of our working group, including the results of this survey, which will guide our work over the coming year.

Future steps for the working group will involve setting the terminology and framework based on the survey results and applying the classification methods to a set of SNVs, indels, and CNVs with known or suspected low effect size and/or low penetrance.

A rigorous study of the need for Sanger confirmation in clinical genetic testing

Stephen Lincoln[1], Matthew Lebo[2,3], Justin Zook[4], Rebecca Truty[1], Chiaqo-Feng Lin[2], Marc Salit[3,5], Heidi Rehm[4], Robert Nussbaum[1,6]

[1] Invitae, San Francisco, CA, United States
[2] Harvard Partners Laboratory for Molecular Medicine, Cambridge, MA, United States
[3] Brigham and Womens Hospital, Boston, MA, United States
[4] National Institute of Standards and Technology, Gaithersburg, MD, United States
[5] Joint Institute for Metrology and Biology, Stanford University, Palo Alto, CA, United States [6] University of California, San Francisco, CA, United States

Background: The confirmation of variants identified by next-generation sequencing (NGS) by using orthogonal assays is standard practice, particularly in situations where false positives can have substantial consequences (e.g. clinical testing). Published studies have examined this issue, concluding that confirmation of the highest-quality NGS calls may not always be necessary. However, these studies are generally underpowered and explore limited aspects of the underlying data. Rigorously choosing criteria that separate high accuracy calls from those which benefit from confirmation remains a critical open issue.

Methods: We examined NGS data from two clinical laboratories. Over 80,000 patient specimens and five well-characterized reference samples (Genome in a Bottle) were analyzed. In total, 200,000 variant calls with orthogonal data were examined in detail including 1,600 primary variant caller false positives.

Results: An optimization algorithm used these data to identify criteria that confidently flag 100% of false positive calls (CI lower bound: 99.74% overall) while minimizing the number of flagged true positives. Rather than relying on one or two quality metrics, as most published methods do, a battery of criteria proved superior, consistent with current clinical NGS guidelines [Roy et al. JMD 2018]. Indeed, our expanded criteria identify false positives that the currently published criteria miss. We also find that historical performance (observing a variant as a true positive some number of times) is an ineffective criterion.

Discussion: Although we see limitations with the currently published criteria, our large multi-laboratory study reaffirms prior findings that high accuracy variant calls can be separated from those which merit confirmation. Our rigorous methodology for determining test and laboratory-specific criteria can be generalized into a practical approach which can be applied by many laboratories. This approach may help reduce the cost and time burden of confirmation without impacting clinical accuracy.

Complex genetic variants are prevalent causes of disease: Implications for clinical genetics and test validation

Stephen Lincoln[1], Andrew Fellowes[2], Shazia Mahamdallie[3], Shimul Chowdhury[4], Eric Klee[5], Justin Zook[6], Rebecca Truty[1], Russell Garlick[7], Marc Salit[8], Nazneen Rahman[3], Stephen Kingsmore[4], Robert Nussbaum[1], Matthew Ferber[5], Brian Shirts[9]

[1]Invitae, San Francisco, CA, USA
[2]Peter MacCallum Cancer Centre, Melbourne, Australia
[3]Institute of Cancer Research, London, UK
[4]Rady Children's Hospital, San Diego, CA, USA
[5]Mayo Clinic, Rochester, MN, USA
[6]National Institute of Standards and Technology, Gaithersburg, MD, USA
[7]Seracare, Gaithersburg, MD, USA
[8]Stanford University, Palo Alto, CA, USA
[9]University of Washington. Seattle, WA, USA

Next-generation sequencing is commonly used by clinical genetic tests to detect SNVs, small indels, and, sometimes, CNVs in relatively accessible regions of a patient's genome. However, conventional NGS methods have limitations. An analysis of over 80,000 patients, tested for constitutional alterations using sensitive methods, showed that variants of other, technically challenging types comprise between 9 and 19% of the reportable pathogenic findings depending on clinical indication. Approximately 50% of these variants are of challenging types (large indels, single exon CNVs, etc.), 20% are in challenging genomic regions (homopolymers, non-unique sequences, etc.), and 15% are in regions poorly covered by standard exome capture kits. A further 15% presented multiple challenges.

The high prevalence of these complex variants can be an under-recognized challenge and the sensitivity of particular genetic tests for such variants is not always described. One reason is that positive control specimens harboring such variants can be difficult to obtain. To investigate one possible solution to this problem we developed a synthetic specimen containing 22 challenging variants of diverse types in 7 commonly tested genes. Raw NGS data for these synthetic variants was found to mimic that of the endogenous variants and presented similar technical challenges for NGS.

This specimen was sequenced using 10 different NGS tests by an international group of collaborating laboratories. These tests employed different sequencing platforms, library methods, and bioinformatics pipelines. All 10 tests detected all of the relatively "easy" SNVs and small indels present (with one exception). However, only 10 of the 22 challenging variants were detected by all tests, and just 3 tests detected all 22. Limitations with large indels, homopolymer associated variants, variants in non-unique regions, and other challenges were observed. Many, but not all of these limitations appeared to be bioinformatic in nature.

We believe that both our prevalence data and control specimens such as ours may be a valuable asset to the ClinGen community for optimizing clinical tests. The particular specimen described herein is now available and additional specimens with over 100 variants are under construction. Additionally, we have expanded the prevalence data to include over 200,000 patients and additional genes, and this resource will be available as a published data set.

De Novo Partial Trisomy of Distal 17q: A Case Series

Jaime Lopes[1], Nicole Hoppman[1], Hutton Kearny[1], Umut Aypar[1]
1. Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN, USA

Duplication of 17q is rare chromosomal disorder described in patients with dysmorphic features, microcephaly, intellectual disability, brachyrhizomelia, and cardiac abnormalities among other features. There are few reported cases of 17q duplication specific to the distal region and the phenotypic consequence of these variants is not well understood. Here we describe a series of four cases with partial trisomy of distal 17q detected by chromosomal microarray analysis and confirmed by fluorescence in situ hybridization (FISH).

Two cases were mosaic in nature and a product of an unbalanced translocation. The first featured a 5.7 Mb duplication at 17q25.3 involving approximately 139 genes in a 43 year old female with severe intellectual disabilities, dysmorphic facial features, and hypothyroidism. FISH confirmed this variant to be in 47% of metaphases. The second was a 14.2 Mb terminal duplication at 17q24.2 including 275 genes in a 1 day old male with reported craniosynostosis and facial anomalies. This variant was present in 80% of metaphases.

Another two cases were constitutional in nature with a 2.5 Mb duplication from 17q25.1 to 17q25.3 involving 69 genes. This variant was observed in a 3 year old female with reported global developmental delay, autism spectrum disorder, insomnia, speech delay, and aggressive behavior. The second constitutional variant uncovered was a 10 Mb terminal duplication at 17q25.1 involving approximately 246 genes. This variant was observed in an 8 year old male with failure to thrive, short stature, gross motor delay, learning disability, seizures, structural brain anomaly, ventricular septal defect, dysmorphic facial features, microcephaly, hydro nephrosis, and undescended testes.

While the genes involved in these duplications are not known to correlate with specific phenotypes with increased dosage, no other copy number variant of clinical relevance was uncovered in any of the described patients. The clinical findings of intellectual disability; motor delay, craniofacial dysmorphisms, and seizures in these cases are consistent with those previously reported with duplication of 17q23 → qter. Additionally, all four cases were found to be de novo events. Therefore, these variants are considered likely pathogenic.

Myeloid Malignancy Variant Curation Expert Panel: an ASH-sponsored ClinGen Expert Panel to optimize and validate ACMG/AMP variant interpretation guidelines for genes associated with inherited myeloid neoplasms

X. Luo[1,2], S. Mohan[5], S. Feurstein[3], S. Jackson[6], J. Ross[5], C. Porter[11], C. Kesserwan[10], A. Agarwal[9], P. Baliakas[21], A. Bertuch[1,2], A. Brown[7], M. Chicka[16], C. DiNardo[12], M. Fleming[15], S. Keel[4], Z. Li[3], M. Luo[18], L. Malcovati[14], S. McWeeney[9], K.E. Nichols[10], A. Raimbault[8], L. Rawlings[7], M. Routbort[12], G. Ryan[13], J. Soulier[8], N. Speck[19], T. Vulliamy[20], M. Walsh[17], Y. Wang[6], L. Zhang[17], S.E. Plon[1,2], D. Wu[4], L. Godley[3]*.*

1) Baylor College of Medicine, Houston, TX;
2)Texas Children's Hospital, Houston, TX;
3) University of Chicago, Chicago, IL;
4) University of Washington, Seattle, WA;
5) University of North Carolina at Chapel Hill, Chapel Hill, NC;
6) GeneDx, Gaithersburg, MD;
7) SA Pathology, Adelaide, Australia;
8) University of Paris Diderot, Paris, France;
9) Oregon Health & Science University, Portland, OR;
10) St. Jude Children's Hospital, Memphis, TN;
11) Emory University, Atlanta, GA;
12) University of Texas MD Anderson Cancer Center, Houston, TX;
13) American Society of Hematology, Washington, DC;
14) University of Pavia, Pavia PV, Italy;
15) Boston Children's Hospital, Harvard Medical School Boston, MA;
16) Prevention Genetics, Marshfield, WI;
17) Memorial Sloan Kettering Cancer Center, New York, NY;
18) Children's Hospital of Philadelphia, Philadelphia, PA;
19) University of Pennsylvania, Philadelphia, PA;
20) Queen Mary College London, London, UK;
21) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden

Clinical Genome Resource (ClinGen) is a National Institutes of Health (NIH)-funded effort dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. ClinGen has developed both gene and variant expert panels to adapt the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines for consistent and accurate variant classification of specific genes and diseases. Here, we describe a new effort initiated in 2018 and supported by the American Society of Hematology (ASH) in collaboration with ClinGen to develop expert panels. This effort was motivated by the increasing use of genomics in clinical hematology and the lack of resources containing expert interpretation of germline variation. This panel, named the ClinGen Myeloid Malignancy Variant Curation Expert Panel, is focused on the curation and annotation of variants in genes associated with familial/inherited risk for myeloid malignancies. Our team consists of expert clinicians, clinical laboratory diagnosticians, and researchers interested in developing and implementing standardized protocols for sequence variant specific annotations of genes in inherited myeloid malignancies. The optimization of the ACMG/AMP guidelines encompasses disease-/gene-informed specifications or strength adjustments of existing rules, including defining gene-specific population frequency cutoffs, and specifying recommendations for the use of computational/predictive data, as supported by published functional and clinical data in addition to guidance on ACMG/AMP variant interpretation provided by the ClinGen effort. Our initial focus has been to organize sub-groups of teams to develop approaches for evaluating ACMG/AMP codes to interpret germline variants of the *RUNX1* gene. Once the curation of *RUNX1* variants is underway, we will extend our focus to include *CEBPA, DDX41, ETV6*, and *GATA2*. These efforts will be bolstered by encouraging submission of existing

variant interpretations to ClinVar or other public variant databases by the Hematology community. In summary, the ClinGen Myeloid Malignancy Variant Curation Expert Panel aims to develop recommendations to optimize ACMG/AMP criteria for standardization of variant interpretation in myeloid leukemia genes and make expert-reviewed and interpreted variants available to the hematology community through ClinVar and the ClinGen website (www.clinicalgenome.org) to support patient care and research.

Polygenic risk information for coronary artery disease - P5.fi FinHealth

A. Minttu Marttila[1], B. Teemu Paajanen[1], C. Heidi Marjonen[1], D. Niko Kallio[1], E. Ari Haukkala[2], F. Helena Kääriäinen[1], G. Kati Kristiansson[1], H. Markus Perola[1,3]

[1]National Institute for Health and Welfare, Helsinki, Finland
[2]Faculty of Social Sciences, University of Helsinki, Finland
[3]Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland

Introduction: We have tested how polygenic risk scores (PRS) for coronary artery disease (CAD), type 2 diabetes and venous thromboembolism affect the risk of disease within the following ten years in participants from a population-based study FINRISK. We will also utilize NMR metabolomic risk information in disease prevention.

Materials and Methods: Based on follow-up data from national health care registries we modelled PRS in whole genome genotyped population based FINRISK cohorts with multiple registry follow-up for incident cases (N=21726) using Cox regression models. We modelled the impact of genetic and traditional risk factors on a risk of disease within the next 10 years.

Results: We compared the CAD classification of Cox regression model with traditional risk factors and polygenic additive model with 6,6M variants. From basic model risk class 10-20% PRS reclassified 205 participants to highest >20% risk class. In FINRISK participants risk for CAD was higher for men and smoking adds to risk of disease in all PRS classes. Risk for disease increases with BMI being highest in BMI class 30-35.

Conclusions: The validation showed that adding PRS to the traditional risk factors significantly changed the risk enabling reclassification of as many as 17% of the participants. PRS and metabolic risk information are returned to volunteering participants through an internet portal. Changes in life style are followed up using questionnaires through the portal and the morbidity will be collected through health registers. P5 study is a pilot for P6 which will recruit 200 000 participants and study 10-20 diseases.

Grants:
The Finnish Innovation Fund Sitra
The Finnish Foundation for Cardiovascular Research
Yrjö Jahnsson Foundation

Authors: Elizabeth M. McCormick[1], Kierstin Keller[2], Lishuang Shen[3], Danuta Krotoski[4], Xiaowu Gai[3,5], Marni J. Falk[1,6], Zarazuela Zolkipli-Cunningham[1,6], and Shamima Rahman[7]

Institution: [1]Mitochondrial Medicine Frontier Program, Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [2]Center for Mitochondrial and Epigenomic Medicine, Department of Pathology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [3]Center for Personalized Medicine, Department of Pathology & Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA; [4]IDDB/NICHD, National Institutes of Health, Bethesda, MD, USA; [5]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [6]University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; [7]Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom.

Title. Standardization of gene-disease curation for Leigh syndrome spectrum within the ClinGen Gene Clinical Validity Curation Framework.

Background. Primary mitochondrial disease (PMD) encompasses a broad group of disorders caused by a deficiency in energy metabolism. Manifestations of PMD range from adult-onset isolated organ system involvement to neonatal lethal multi-organ system disease. Leigh syndrome (LS) is classically defined as a typically progressive neurodegenerative disorder characterized by neuropathology findings of bilateral necrotic brainstem and basal ganglia lesions. This clinical entity is the most common manifestation of PMD in children but may also present in adults. Since first being described in 1951, the definition of LS has evolved to now also include evidence of impaired oxidative phosphorylation. LS is commonly caused by pathogenic variants in mitochondrial and nuclear genes encoding mitochondrial proteins but has also been associated with other potentially treatable genetic etiologies, necessitating accurate gene-disease association.

Methods and results. Within the ClinGen Gene Clinical Validity Curation Framework, we sought to evaluate the relationship between 90 genes (across both nuclear and mitochondrial genomes) and LS. Prior to starting this process we identified inconsistencies in how PMD experts define LS. A clearly defined phenotype is essential to the ClinGen gene curation framework so we proposed to the ClinGen Mitochondrial Disease Gene Curation Expert Panel a draft consensus on the definition of this phenotype. This phenotype was refined after incorporating feedback from the global experts and is now referred to as Leigh syndrome spectrum (LSS). We systematically reviewed the current ClinGen Gene Clinical Validity Curation Framework and established criteria that must be met to confirm a LSS phenotype in the case level line of genetic evidence. We identified commonly utilized experimental methods that, for some assays, are unique to PMD and LSS research, and established scoring guidance for each assay or model based on our LSS definition.

Conclusion. We identified areas within the current ClinGen Gene Clinical Validity Curation Framework that could be interpreted and scored differently as it relates to LSS and PMD, assessed where commonly used PMD terms and assays might be applied, and established a scoring rubric for LSS. This improved intra- and inter-biocurator curation and expert panel scoring. The additional standardization provided here will be generalizable to additional PMD phenotypes as they are curated over time.

ClinGen Somatic Cancer Working Group: The Pancreatic Cancer Taskforce

Matthew McCoy[1], Shruti Rao[1], Kaylee Barber[2], Wan-Hsin Lin[3], Andrea Vaags[4], Gagandeep Brar[5]

1. Innovation Center for Biomedical Informatics, Georgetown University Medical Center
2. Bio-Reference Laboratories, Inc
3. Mayo Clinic, Department of Cancer Biology
4. Trillium Health Partners, Department of Laboratory Medicine and Genetics, Mississauga, Ontario, Canada.
5. National Cancer Institute, Thoracic and GI Malignancies Branch

Pancreatic cancer is among the deadliest cancers, largely due to its late stage diagnosis. The poor prognosis and survival underpin the motivation for the Pancreatic Cancer Taskforce to curate variants that are commonly seen in the patient population. The taskforce is part of the Clinical Genome (ClinGen) Resource Somatic Cancer Working Group, and is composed of genetic counselors, clinical professionals, and curation experts with a range of expertise in pancreatic cancers. With the goal to centralize and share freely accessible genomic data, the team has prioritized 25 genes that are commonly mutated in pancreatic cancer patients for curation. We are currently compiling data elements from literature required to curate variants within the updated MVLD (Ritter DI, et al., 2016) framework, and depositing these findings in the Clinical Interpretation of Variants in Cancer (CIViC) database. Among the many challenges facing the curation is the limited availability of experimental evidence for many variants, especially in the context of pancreatic cancers. The taskforce is developing consensus-based strategies to explore how the specific functional context and assertions made in other cancer types can inform the curation of these variants of unknown significance.

Updates to the Clinical Genome (ClinGen) Resource Somatic Cancer Working Group's Minimum Variant Level Data (MVLD): v2.0

Shruti Rao 1, Deborah Ritter* 2, Arpad Danos 3, Killannin Krysiak 3, Gordana Raca 4, Angshumoy Roy 2, Dmitriy Sonkin 5, Matthew McCoy 2, Malachi Griffith 3, Nikoletta Sidiropoulos 7, Obi L. Griffith 3, Shashikant Kulkarni 2,8, Subha Madhavan 1, on behalf of The Clinical Genome Resource Somatic Cancer Working Group.*co-first authors*

1. Innovation Center for Biomedical informatics, Georgetown University
2. Baylor College of Medicine and Department of Pediatrics, Texas Children's Hospital
3. McDonnell Genome Institute, Washington University School of Medicine
4. Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, University of Southern California
5. Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute
6. Larner College of Medicine, University of Vermont Medical Center
7. Baylor Genetics

The Clinical Genome (ClinGen) Resource Somatic Cancer Working Group (WG) had published the first version of a consensus-based Minimum Variant Level Data (MVLD) framework (Ritter DI, et.al., 2016) to curate somatic cancer variants for clinical utility, and enable the rapid communication and transfer of this data between multiple end users including clinicians, molecular pathologists, research and data scientists, as well as the public. The pediatric and pancreatic cancer task forces of the Somatic Cancer WG have since been using the MVLD framework to inform their curation efforts. This process highlighted several areas of improvement to satisfy the need and harmonize ongoing efforts in the somatic cancer community.

Here, we describe updates to the MVLD framework, i.e. MVLD v2.0. Firstly, the current levels and sub-levels of evidence in MVLD have been replaced by the somatic interpretation guidelines published by the Association of Molecular Pathologists (AMP), the College of American Pathologists (CAP) and the American Society of Clinical Oncology (ASCO). Secondly, we are expanding MVLD to capture other relevant somatic events, such as RNA fusions, gene amplifications, and chromosomal rearrangements, in addition to the currently acceptable single nucleotide variants, multi-nucleotide variants, insertions, and deletions. Thirdly, we are implementing MVLD 2.0 as the input format for case reports in cancer journals such as Cancer Genetics. These case reports in the MVLD 2.0 format will be automatically entered into Clinical Interpretations of Variants in Cancer (CIViC), which will, in turn, submit variants to ClinVar on a semi-annual basis. This workflow will help streamline and mediate the process of ClinVar submissions for case study publication.

Using MVLD, the Somatic Cancer WG is engaged in curation efforts in three disease areas of focus: pediatric cancers, pancreatic cancers and non-small cell lung cancers. Significant efforts

Undergraduate Annotation Pipeline: Using a Tiered Approach to ClinGen Gene Curation Training

Jennifer McGlaughon, Courtney Thaxton, Jennifer Goldstein, Kathleen Wallace, Daniela DeCristo, Megan Mayers , May Flowers, Misbah Ahmad, Erin Blalock, Layna Perini, Autumn Self, Courtney Davis, Jonathan Berg

Department of Genetics, University of North Carolina at Chapel Hill, NC, USA

The ClinGen gene curation process involves the identification, annotation, and interpretation of relevant data from the literature by a biocurator, and deposition of data into the ClinGen gene curation interface (GCI). Curation of applicable evidence can be a time-intensive task, so the UNC Biocuration Core implemented an undergraduate pipeline to assist in the annotation of peer-reviewed articles containing curatable evidence using Hypothes.is, a web-based annotation tool. This pipeline not only assists the biocurators in gathering data, but also allows undergraduates to develop skills that are relevant to any field, such as critical evaluation of peer-reviewed articles. Gene curation is a complicated process that requires advanced knowledge of genetics and genomics, therefore we implemented a 3-tiered approach to introduce undergraduates to gene curation in a step-wise manner with increasing complexity. Training materials have been developed for each level. Using Hypothes.is, the undergraduates identify, annotate, and tag information in a peer-reviewed article that is relevant to the gene curation. The first level of the pipeline involves the annotation of demographic and phenotypic information about individuals with variants in the gene of interest. Level 2 builds on the annotations from level 1, adding information about variants found in individuals (e.g. ClinVar ID, gnomAD frequency), genotyping methods, and previous testing. Finally, level 3 includes any additional information not captured in the first two levels, such as family segregation data, variant-level evidence, and experimental evidence. Once an article has been annotated at each level, a senior biocurator reviews the information and enters it into the GCI. Since its launch in August, 2018, the UNC undergraduate annotation pipeline has trained 7 undergraduates that have annotated more than 50 articles for over 30 genes of interest for curation. Here, we will present an overview of the pipeline, training process and metrics obtained. The undergraduate annotation pipeline can also be adapted to assist in other curatorial efforts, such as variant curation, and can serve as a proxy for crowd-sourced curation on a larger scale.

Establishing the definition of "well-established" functional assays: a case study of PS3/BS3 application by ClinGen Variant Curation Expert Panels

Shannon M. McNulty[1], Dona M. Kanavy[1], Meera K. Jairath[1], Sarah E. Brnich[1,2], Chris Bizon[3], Bradford Powell[1], Jonathan S. Berg[1]

[1]Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[2]Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[3]Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Under the 2015 ACMG/AMP variant interpretation guidelines, "well-established *in vitro* or *in vivo* functional studies" can be applied as strong evidence for or against pathogenicity (PS3/BS3). While functional assays can provide compelling information about the effect of a variant on gene product function, selecting appropriate assays and studies to satisfy PS3/BS3 often requires expert-level knowledge of the gene and disease. To enhance variant interpretation in the context of specific diseases, six ClinGen Variant Curation Expert Panels (VCEPs) published gene-specific recommendations for evidence application, including providing guidelines for PS3/BS3 use and identifying "well-established" assays. These adapted guidelines were then applied to a set of pilot variants. We utilized the Hearing Loss (HL) and *PTEN* VCEP pilot variant classification as a case study to better define the characteristics of functional assays that satisfy PS3/BS3 criteria by conducting a comprehensive biocuration of instances of assays cited by each VCEP. Our analysis of the primary literature indicated that the parameters VCEPs specified as preferred for functional studies (e.g. assay-specific controls and validation measures) were satisfied to varying degrees. For example, although the HL VCEP approved electrical coupling assays to assess *GJB2* variants and specified water-injected controls should be incorporated, this control was only included in 4/7 studies cited by the panel. Our curation efforts also highlighted the relatively low frequency of PS3/BS3 utilization in pilot variant interpretation by the HL VCEP (10/51 variants) and *PTEN* VCEP (6/42 variants), partially due to a paucity of functional data for variants in the studied genes. Importantly, the *PTEN* VCEP reviewed a high-throughput assay of protein activity that tested over 8000 *PTEN* variants. Assays such as these are uniquely powered to rapidly generate functional data about many variants, including those that have not yet been observed in a clinical setting, and could be integral in reclassifying variants of uncertain significance. Given the potential for functional evidence to improve variant interpretation, this work highlights the need for VCEP recommendations to be conveyed to research labs for incorporation into study designs to generate data better suited for use in clinical variant interpretation, as well as continued development of standardized, disease-specific guidelines for functional evidence application.

Resolving Classification of PTEN Variants Using Gene-Specific ACMG/AMP Guidelines: Results from the First Year of the ClinGen PTEN Variant Curation Expert Panel

Jessica L. Mester[1], Charis Eng[2-4], and Madhuri Hedge[5,6], on behalf of the ClinGen PTEN Variant Curation Expert Panel (Helio Costa, Felicia Hernandez, Kathleen Hruska, Robert Huether, Rachid Karam, Katherine Lachlan, Joanne Ngeow, Melody Perpich, Tina Pesaran, Kaitlin Sesock, Liying Zhang)

1. GeneDx, Inc, Gaithersburg, MD. 2. Case Comprehensive Cancer Center, Cleveland, OH. 3. Case Western Reserve University School of Medicine, Cleveland, OH. 4. Cleveland Clinic Genomic Medicine Institute, Cleveland, OH. 5. Emory University, Atlanta, GA. 6. PerkinElmer Genetics, Pittsburgh, PA.

The ClinGen PTEN Variant Curation Expert Panel (VCEP) provides assessment regarding the pathogenicity of germline *PTEN* variants with respect to a PTEN Hamartoma Tumor syndrome (MIM+601728) phenotype using PTEN-specific variant classification criteria based on the 2015 ACMG/AMP framework. Of highest priority are variants with conflicting ClinVar assertions (CONF) or those classified as variants of uncertain significance (VUS) by multiple submitters, with pathogenic (PATH), likely pathogenic (LPATH), benign (BEN), likely benign (LBEN), and other VUS also undergoing review. Since receiving full EP status, the PTEN VCEP has finalized classifications for 98 variants utilizing the Variant Curation Interface and uploaded assertions and detailed evidence summaries to ClinVar. We present a summary of the variants curated to date, highlighting PTEN VCEP resolution of variants with CONF or VUS assertions.

Classification other than VUS was achieved for 10/18 (55.6%) CONF variants, with 6 classified by the VCEP as BEN or LBEN and 4 as PATH or LPATH. Among 16 variants with VUS assertions by multiple submitters, the VCEP was able to reclassify 3 (18.8%), 1 as LBEN and 2 as LPATH. Shared internal laboratory data provided evidence critical for the resolution of 8 CONF variants and reclassification of 2 VUS. Most of the variants with ClinVar assertions of PATH/LPATH (45/48, 93.8%) or BEN/LBEN (14/16, 87.5%) retained that classification following VCEP review. The 26 variants currently classified as VUS by the VCEP include missense variants (n=13), variants in the promoter region or 3' UTR (n=6), intronic or synonymous variants for which *in silico* tools are uninformative or conflicting (n=5), and in-frame indels (n=2). Notably, 12 of the 26 VUS (46.2%) require only one additional benign supporting criterion to meet the classification of LBEN.

All VUS and LPATH will undergo re-review, two years from the last review, with the intent of seeking new evidence that may aid in variant classification. Likewise, the PTEN VCEP plans to make updates to the current PTEN-specific variant classification criteria that may permit future resolution of variants currently classified as VUS. The presence of the PTEN VCEP's assertions and evidence in ClinVar will enable laboratories and others performing *PTEN* variant analysis to consider aligning their own criteria and classifications with those of the VCEP, ultimately decreasing inter-laboratory classification discrepancies.

The Open Custom-Ranked Analysis of Variants Toolkit (OpenCRAVAT): A customizable annotation and prioritization pipeline for genes and variants

Kymberleigh Pagel[1], Lily Zheng[1,2], Rick Kim[3], Kyle Moad[3], Michael Ryan[3], Rachel Karchin[1,4,5,6]

[1]The Institute for Computational Medicine, The Johns Hopkins University, Baltimore, MD
[2]Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD
[3]In Silico Solutions, Falls Church, VA
[4]Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD
[5]Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD
[6]Department of Computer Science, The Johns Hopkins University, Baltimore, MD

Abstract

More than 3,000 known Mendelian diseases have unknown genetic causes, and discovery and interpretation of genetic variants underlying Mendelian and rare disease remains a complex and daunting task. The modern genetics researcher is confronted with literally hundreds of published methods to annotate and represent variants. There are databases of genes and variants, phenotype-genotype relationships, algorithms that score and rank genes, and *in silic*o variant effect prediction tools. Because variant prioritization is a multi-factorial problem, a welcome development in the field has been the emergence of decision support frameworks, which make it easier for users to integrate many resources in an interactive environment. Current decision support frameworks are limited by closed proprietary architectures, access to a restricted set of tools, lack of customizability, web dependencies that expose protected data, and limited scalability.

OpenCRAVAT is an open source, scalable decision support system with an extensive catalog of resources to support variant and gene prioritization in Mendelian and rare diseases. It is highly customizable, does not expose protected data over the web and scales to very large datasets. As the term decision support implies, associating variants to disease ultimately relies on manual human expert review and interpretation. Automated systems can make these tasks tractable, by reducing the number of variants and genes to be considered to the most promising few. Interactive environments for visual exploration of results are critical to manual review. To our knowledge, OpenCRAVAT is the first open tool that provides both fast, configurable variant and gene prioritization and an integrated graphical interface that facilitates expert manual review of prioritized results.

41

ClinGen Allele Registry and Evidence Repository provide access to FDA-recognized, expert-curated human variant classifications and supporting data

Ronak Y Patel, Sameer Paithankar, Andrew R Jackson, Sai Lakshmi Subramanian, Kevin Riehle, Tam P Sneddon, Julianne O'Daniel, Erin Riggs, Larry Babb, Tristan Nelson, Chris Bizon, Matt W Wright, Hannah Wand, Steven Harrison, Jonathan Berg, Heidi L Rehm, Carlos D Bustamante, Sharon E Plon, Aleksandar Milosavljevic, on behalf of **Clinical Genome Resource**

Department of Molecular And Human Genetics, Baylor College Of Medicine, Houston, TX 77030
Renaissance Computing Institute, University of North Carolina, Chapel Hill, North Carolina.
Geisinger's Autism and Developmental Medicine Institute, Lewisburg, Pennsylvania.
Department of Genetics, University of North Carolina, Chapel Hill, North Carolina.
Department of Biomedical Data Sciences, Stanford University School of Medicine, Palo Alto, California.
National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland.
Department of Pediatrics, Baylor College of Medicine, Houston, Texas.
Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA.
Broad Institute of MIT and Harvard, Cambridge, MA, USA.

The Clinical Genome resource (ClinGen) project is dedicated to building an authoritative resource that defines the clinical relevance of genes and variants. The Food and Drug Administration (FDA) has recently recognized ClinGen's variant dataset as valid scientific evidence. This recognition includes the variant pathogenicity classification, as well as the evidence and processes underlying the classification. Each ClinGen variant curation expert panel (VCEP) clearly defines and maintains guidelines for variant classification in focused clinical domains. The guidelines broadly are a specification of the ACMG/AMP - 2015 guidelines tailored for a specific domain of interest. Once guidelines are approved by ClinGen, the VCEP analyzes and documents evidence for or against variant pathogenicity and provides pathogenicity assessments. The FDA human variant database program emphasizes data validity, transparency, and access including (1) unambiguous naming of variants; and (2) public access to evidence that supports pathogenicity assessments. Key to our approach is the (1) ClinGen Allele Registry, CAR (reg.clinicalgenome.org); and (2) ClinGen Evidence Repository, ERepo (erepo.clinicalgenome.org). CAR provides readily available, globally unique variant identifiers that enable aggregation of information about variants from different sources. More than 910 million distinct variants are currently registered, including those from key resources such as gnomAD, dbSNP, ClinVar and a smaller number registered by users. Several major resources including CIViC, Ensembl-EBI, and ClinVar have adopted the CAR identifiers, thus enabling the linking of variant information across these resources. The ERepo is browsable and searchable using various types of queries through a dedicated web-page. ERepo

stores evidence using the ClinGen-SEPIO machine-readable format and is accessible via both a web User Interface (web-UI) and programmatically via Application Programming Interfaces (web-APIs). Currently, the ERepo hosts 538 variant classifications and supporting evidence generated by six ClinGen VCEPs and is linked to the VCEP ClinVar submission. We present data access use cases and demonstrate key aspects of these resources.

Automated classification of ~7,000 variants shows near perfect concordance with expert panel assessments

Jennifer L. Poitras[1], Daniel Richards[1], Hua Su[1], Tara Love[1], Rupert Yip[1]

[1]QIAGEN, Redwood City, CA, United States

Gathering the most current and accurate information is critical to variant interpretation. The QIAGEN knowledgebase includes a manually curated database of variant specific publications, and data from public and proprietary databases. This resource is the cornerstone of QIAGEN Clinical Insight (QCI), which facilitates automated variant classification using all 28 ACMG rules (including many of the newly amended *MYH7* context specific rules), while transparently providing the underlying evidence. Here, we compare the concordance of QCI's automated variant classification with expert panel assessments across hereditary breast cancer and cardiomyopathy disease indications. Expert panel (ENIGMA) reviewed *BRCA1* and *BRCA2* variants (n=6154), and variants reviewed by the ClinGen Inherited Cardiomyopathy Expert Panel (n=102) were exported from ClinVar. Variants reviewed by the RASopathy expert panel, and those related to drug sensitivity alone were excluded. Resulting VCFs were uploaded into QCI, and concordance of automated classifications was compared with expert panel assessments in ClinVar. With respect to clinical actionability, automated variant classifications were extremely concordant, reaching 99.6% concordance with ENIGMA assessments of *BRCA* variants, and 96.1% concordance with cardiomyopathy variants assessed by the ClinGen expert panel. The small number of differences seen in the ENIGMA dataset could be attributed to functional studies used by the automated algorithm not considered by the expert panel. The 3.9% of discrepant classifications in the cardiomyopathy set likely result from the small number of MYH7 rules we are still in the process of implementing. Through updated content, and continued alignment with professional guidelines, automated variant classification in QCI demonstrates extremely high accuracy across multiple disease contexts. This level of accuracy speaks to the quality of the clinical, functional, and population level data curation, as well as the robustness of the underlying ACMG classification algorithm.

Curating the Clinical Genome (CCG) Annual Conference 2019 Platform Abstract

Title: Perspectives and Use of Race, Ethnicity, and Ancestry in Clinical Curation

Authors: Popejoy, A.B.[1,2], Bonham, V.L.[3], Carpten, J.[4], Crooks, K.[5], Fullerton, S.M.[6], Gignoux, C.R.[5], Hindorff, L.A.[7], Kenny, E.E.[8], Koenig, B.A.[9], Lee, S.S.[10], Oh, S.[11], Ormond, K.E.[10,12], Plon, S.E.[13], Ramos, E.M.[7], Ritter, D.I.[13], Rivas, M.[1], Thornton, T.A.[14], Wright, M.W.[1], and Bustamante, C.D.[1,12]

Affiliations: 1. Department of Biomedical Data Science, Stanford University School of Medicine; 2. Center for Computational, Evolutionary, and Human Genomics (CEHG), Stanford University; 3. Social and Behavioral Research Branch, National Human Genome Research Institute (NHGRI); 4. Institute of Translational Genomics, Keck School of Medicine of the University of Southern California; 5. Colorado Center for Personalized Medicine, Anschutz Medical Campus; 6. Department of Bioethics & Humanities, University of Washington School of Medicine; 7. Division of Genomic Medicine, National Human Genome Research Institute (NHGRI), National Institutes of Health; 8. Center for Genomic Health, Icahn School of Medicine at Mount Sinai; 9. UCSF Bioethics, University of California at San Francisco; 10. Stanford Center for Biomedical Ethics (SCBE), Stanford University School of Medicine; 11. Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco; 12. Department of Genetics, Stanford University School of Medicine; 13. Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine; 14. Department of Biostatistics, University of Washington School of Public Health.

Abstract: The Ancestry and Diversity Working Group (ADWG) of The Clinical Genome Resource (ClinGen, clinicalgenome.org) and members of Clinical Sequencing Evidence Generating Research (CSER) are working to improve the utility and application of clinical genomics for all demographics, by investigating perspectives and practices around the use of race, ethnicity, and ancestry (REA) in genetic testing. In this study, ADWG and CSER developed a survey to reveal trends in current perspectives and practices on the use of REA in the context of genome curation and patient care. The survey was validated through 11 cognitive interviews and disseminated through email lists of ClinGen and CSER, as well as the American College of Medical Genetics and Genomics (ACMG), the American Board of Genetic Counselors (ABGC), and the American Society of Human Genetics (ASHG). Survey respondents (N=193) include clinical geneticists, laboratory directors, genetic counselors, and human genetics researchers from eight different countries, with ~90% of respondents in the U.S., spread out across 32 states.

Our results indicate that the majority of respondents (N=188, 97%) use information such as population allele frequencies, self-reported race or ethnicity, and/or ancestral origin(s) of a patient or family members in some aspect of their research (N=45, 23%), clinical practice (N=88, 46%), or both (N=55, 29%). Respondents who order genetic tests or return results to patients (N=113) were asked an additional subset of questions related to the importance of diversity measures in their work. Among these respondents, more than half report that race (N=60, 53%), ethnicity (N=75, 66%), and ancestry (N=81, 72%) are “somewhat important”, “important”, or “very important” when ordering a genetic test. Similarly, the majority of those who conduct clinical variant interpretation (N=114), report that race (N=66, 58%), ethnicity (N=65, 57%), and ancestry (N=79, 69%) are at least somewhat important for their clinical curation work. However, nearly one-third of all survey respondents reported feeling “not at all confident” in their command of these measures (N=56, 29%) and the majority (N=124, 65%) felt that new guidelines would be helpful for directing the collection, use, and communication of REA data. We anticipate that this work, and the future development of guidelines, will be beneficial for researchers and clinical professionals, with strong potential to improve the quality of clinical curation and care.

MANE Select: a set of matched representative transcripts from NCBI-RefSeq and EMBL-EBI GENCODE gene sets for every human protein-coding gene.

S Pujar[2], J. Loveland[1], J. Morales[1], A. Astashyn[2], R. Bennett[1], C. Davidson[1], O. Ermolaeva[2], C. Farrell[2], L. Gil[1], J.M. Gonzalez[1], V. Joardar[2], M. Kay[1], V. Kodali[2], K. McGarvey[2], A. McMahon[1], S. Rangwala[2], G. Threadgold[1], F. Cunningham[1], A. Frankish[1], T. Murphy[2]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK CB10 1SD.
[2] National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, U.S.A. 20894

NCBI's RefSeq and the EBI's GENCODE annotations have provided independent, high-quality reference datasets describing the human gene complement. While curated RefSeq transcripts have served as clinical reference standards to report clinically significant mutations, several genomic tools and browsers based on clinical variation data use Ensembl transcripts as default reference sequences. This may lead to inconsistency in clinical reporting, especially when the corresponding RefSeq and Ensembl transcripts have differences. To address the need for convergence on key high value annotations, NCBI and EBI have initiated a new collaborative project called Matched Annotation from NCBI and EBI (MANE), which provides a matching set of transcripts for human protein-coding genes. As a first step in this project, we have recently released a beta version of the MANE Select set (MANE Select v0.5), which defines a matched representative transcript per gene.

To build the MANE Select set, initially, independent pipelines at NCBI (RefSeq Select pipeline) and EBI picked their own select transcripts, which were then matched. The criteria used to pick the select transcripts include expression levels, conservation data, prior curation, transcript and protein length and concordance with UniProt canonical isoform. Matched transcripts from both annotation sets were then updated to match the start and end coordinates based on high-throughput data. MANE Select transcripts are stable, perfectly match the human GRCh38 reference assembly and represent the biology of the gene. Curators from both the annotation groups work together to improve the automated pipelines, make select transcript decisions in cases where the automated choice is not ideal and to define additional datasets (such as MANE Plus) in this project.

One of our goals is to align the MANE Select set with the workflow of the clinical variation community, and to make this set available as a stable transcript set annotated identically by NCBI and EBI and backed by robust curation support from both the groups. Towards this end, we are working with the LRG (Locus Reference Genomic) group at EBI to carefully review and pick MANE Select transcripts to serve as LRG reference transcripts for new LRG records. This work was supported in part by the intramural research program of the National Library of Medicine (NIH), and grants from the Wellcome Trust, and the National Human Genome Research Institute.

Computational Methods to Predict the Actionability of Variants of Unknown Significance in the Protein Targets of Small Molecule Inhibitors

Matthew D. McCoy[1,2], Shruti Rao[1,2], Vikram Shivakumar[2], Subha Madhavan[1,2]

1 - Georgetown-Lombardi Comprehensive Cancer Center, Department of Oncology, Georgetown University Medical Center, Washington, D.C., USA

2 - Innovation Center for Biomedical Informatics, Georgetown University, Washington D.C., USA

A major challenge facing genomic curation is how to shift variants of unknown significance as benign or pathogenic when no experimental evidence exists to support any type of assertion. Computational methods offer the only means of gaining any insight into the impact of a mutation on protein function, but face several limitations. No current method performs perfectly, and in most cases the methods only provide a statistical measure of the likelihood a variant will disrupt function. However, even for the most confidant predictions, the utility can be limited because it doesn't help make a clinical decision. Curation guidelines appropriately don't allow for the exclusive use of computational prediction algorithms as the sole source of evidence for variant curation, however they may be reliable in limited contexts. For example, computational simulations of protein structures can be used to assess the impact of mutations on the binding of targeted inhibitor molecules. The results of molecular dynamics simulations can be used to build variant specific scaffolding for small molecule ligand docking, and have the potential to identify when somatic variation will lead to disrupted binding of a targeted inhibitor and inform the optimal selection of a second line therapy. Currently, we are utilizing a computational simulation workflow, SNP2SIM, to study drug resistant variants in a number of targeted protein kinases, and extending the capability to include more complex interactions between other proteins or nucleic acids. While all computational methods are not suitable evidence to make strong claims about the impact of a VUS on pathogenicity, there may be limited contexts where innovative tools may be useful. Future work should not only try to develop methods to address these actionable curation contexts, but also to develop standards to determine when their predictions can be used with confidence.

are also ongoing to adapt and extend the use of MVLD 2.0 to curation activities in publically accessible clinical cancer databases such as CIViC and ClinVar. To date, we have 217 evidence items submitted in CIViC, three assertions, and 133 revisions, comments, and source suggestions. We have created a curation volunteer outreach process, and trained 10 new volunteers to date. We are actively seeking to develop additional cancer curation task teams and seeking input of the cancer community on areas of interest.

DNA vs Enzyme Testing for Tay-Sachs Carrier Screening: Lessons Learned

Kristina Robinson, Shan Yang, Xu Li, Nicole Faulkner, Robert Nussbaum
Invitae - San Francisco, CA

Carrier screening for Tay-Sachs disease has traditionally been recommended for members of high-risk ethnic groups, but advances in genetic sequencing and the difficulty of defining ancestry in a multiethnic society have led to more pan-ethnic offerings. Tay-Sachs carrier screening can be performed by DNA or enzyme testing, but the best method to identify carriers among lower-risk individuals remains undetermined.

We compared carrier screening results for 25,926 individuals in a pan-ethnic population tested by both DNA and enzyme analysis. NGS-based HEXA sequencing, including targeted detection of the del7.6Kb variant, was performed. Individuals heterozygous for pathogenic or likely pathogenic variants were classified as DNA-POS. All others were classified as DNA-NEG, including those with benign, likely benign, and/or variants of uncertain significance (VUSes). Hexosaminidase A activity was measured in leukocytes by two reference laboratories (~50% of samples tested at each) and categorized as carrier (ENZ-POS), indeterminate (ENZ-IND), or negative (ENZ-NEG) using lab-specific reference ranges.

We detected 6 B1 variant carriers (0.023%) and 112 individuals (0.43%) with a benign pseudodeficiency (ψD) allele. Among non-B1, non-ψD results, 94.7% (24,447) had concordant DNA and enzymes and 0.9% (226) were discordant (all DNA-NEG:ENZ-POS). An additional 1135 (4.4%) were ENZ-IND (1126 DNA-NEG:ENZ-IND and 9 DNA-POS:ENZ-IND). Use of different cutoffs by the reference labs yielded very different ENZ-IND rates (7.3% vs. 1.8%). Comparing the frequency distribution of DNA-NEG:ENZ-IND individuals across enzyme values to that of the NEG:NEG and POS:POS groups demonstrated that ENZ-IND results are an obvious extension of the NEG:NEG curve, suggesting that most DNA-NEG:ENZ-IND individuals are truly non-carriers. Rare missense VUSes (<10 alleles in gnomAD) were significantly enriched (14.2%) in DNA-NEG:ENZ-POS individuals, but nearly absent (0.2%) from the ENZ-NEG group and infrequent (1.2%) in ENZ-IND individuals.

Enzyme testing has high sensitivity, but the burden of indeterminate and potentially false positive results is significant when used for first-line screening in pan-ethnic populations. Our data suggest that identifying rare VUSes by DNA analysis, with follow-up enzyme testing in VUS carriers, may provide a better balance between sensitivity and specificity. Furthermore, accumulating enzyme results of VUS carriers may aid in future interpretation of those variants.

Specifications of the ACMG/AMP Variant Curation Guidelines for the Analysis of *ITGA2B/ITGB3* Sequence Variants: Recommendations by ClinGen's Platelet Disorders Expert Panel

Justyne Ross[1], Kristy Lee[1], Brian Branchford[2], Paul Bray[3], Stefanie Dugan[4], Kathleen Freson[5], Paula Heller[6], Andrew Johnson[7], Walter Kahr[8], Michele Lambert[9], Claire Lentaigne[10], Minjie Luo[9], Shruthi Mohan[1], Andrew Mumford[11], Juliana Perez Botero[4], Matthew Rondina[3], Gabriella Ryan[12], Leighton Telling[1], Sarah Westbury[11], Bing Zhang[13], Wolfgang Bergmeier[14], Jorge DiPaola[2]

1. Department of Genetics, University of North Carolina, Chapel Hill, NC, US.
2. Hemophilia and Thrombosis Center, University of Colorado, Aurora, CO, US.
3. Program in Molecular Medicine, University of Utah, Salt Lake City, UT, US.
4. BloodCenter of Wisconsin, Medical College of Wisconsin, Milwaukee, WI, US.
5. Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, University of Leuven, Leuven, BE.
6. Institute of Medical Research Alfredo Lanari. CONICET. University of Buenos Aires, Buenos Aires, AR.
7. Population Sciences Branch, National Heart, Lung and Blood Institute's The Framingham Heart Study, Framingham, MA, US.
8. Department of Paediatrics, University of Toronto, Toronto, ON, CA.
9. The Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, US.
10. Centre for Haematology, Imperial College London, UK.
11. School of Clinical Science and School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK.
12. American Society of Hematology, Washington, DC, US.
13. Department of Pathology, Stanford University School of Medicine, Stanford, CA, US.
14. Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC, US.

The Platelet Disorders Expert Panel (EP), organized by the ClinGen Hemostasis/Thrombosis Clinical Domain Working Group, with support from the American Society of Hematology, is composed of clinicians, researchers, and molecular diagnosticians with expertise in platelet disorders tasked with the development of specifications to the 2015 ACMG/AMP Sequence Variant Interpretation Guidelines. Accordingly, the Platelet Disorders EP has generated gene-specific specifications of ACMG/AMP guidelines for the consistent and accurate classification of variants in the *ITGA2B* and *ITGB3* genes. Pathogenic variants in the *ITGA2B* and *ITGB3* genes cause Glanzmann thrombasthenia (GT), an inherited platelet disorder in which platelets fail to aggregate to physiological agonists due to quantitative or qualitative defects in integrins αIIb or β3, but have a normal response to ristocetin. Disease-/gene-informed specifications of the ACMG/AMP guidelines included application of gene-specific population frequency cutoffs, and recommendations for the use of computational/predictive data, functional data, and clinical evidence. *ITGA2B* and *ITGB3* based specifications were made to 13 ACMG/AMP criteria, including 5 with adjustable strength based on the quantity or quality of evidence. Another 8 criteria were deemed not applicable. The *ITGA2B/ITGB3* GT-related specifications were tested with 60 variants, representing 30 from each of *ITGA2B* and *ITGB3*. Comparing previously reported *ITGA2B/ITGB3* variant classifications to Platelet Disorders EP recommendations, with the integration of internal data from participating diagnostic laboratories, will result in reduced numbers of variants of uncertain significance by reclassification to the appropriate pathogenicity classification. These modified rules provide increased specificity for use in Glanzmann thrombasthenia and provide a basis for specification of ACMG/AMP guidelines to additional platelet disorders commonly tested in hematology practices.

Neurodevelopmental defects due to haploinsufficiency of *KMT2E*

Samin A. Sajan[1], Deepali N. Shinde[1], Sha Tang[2], Rachel Saunders[3], Vikas Bhambhani[3], Sarah L. Dugan[4], Kelly Radtke[1]

[1] Department of Clinical Genomics, Ambry Genetics, Aliso Viejo, CA

[2] WuXi NextCODE, Cambridge, MA

[3] Department of Medical Genetics, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN

[4] Department of Pediatrics, Division of Medical Genetics, University of Utah School of Medicine, Salt Lake City, UT

The lysine-specific methyltransferase 2E (*KMT2E* aka *MLL5*) gene is known to play a role in cell cycle regulation and hematopoiesis. It has recently also been shown to result in neurodevelopmental defects due to three patients having been reported with *de novo* alterations predicted to result in haploinsufficiency (for example DECIPHER patient #263273). An additional line of support for the involvement of *KMT2E* in neurological function is that proper control of histone methylation is critical for neural development and pathogenic alterations in other histone lysine methyltransferase genes such as *KMT2A*, *KMT2B*, and *KMT2D* are well known to cause neurodevelopmental disorders. Moreover, the Developmental Disorders Genotype-Phenotype Database (DDG2P) database already regards *KMT2E* to be a "confirmed" gene for neurodevelopmental disorders. Here we present three additional patients with alterations in *KMT2E* predicted to cause its haploinsufficiency. The first patient had the *de novo* nonsense alteration NM_182931 p.Y779* (c.2337C>A) and had global developmental delay, craniosynostosis, high head circumference, abnormal gait, hypertonic/hypotonic mixed muscle tone, static encephalopathy, congenital torticollis, and mild dysmorphic facial features. The second patient had the maternally inherited nonsense alteration p.C723* (c.2169T>A) and had developmental delay, intellectual disability, mild scoliosis, behavior problems, some dysmorphic facial features, and tetralogy of Fallot. The mother was reported to have required learning assistance in grade school and high school. The third patient had the frameshift p.W1067QFS*2 (c.3198_3234DEL37) and presented with developmental delay, intellectual disability, nystagmus, overgrowth including macrocephaly, and dysmorphic features. This patient's healthy mother and healthy paternal uncle did not harbor this alteration. Our report bolsters the notion that gene disrupting alterations in *KMT2E* cause a neurodevelopmental disorder in humans and expands its molecular (both *de novo* and inherited) and phenotypic spectrum.

ClinGen's Patient Data Sharing Program: Leveraging Data Sharing Experience from GenomeConnect to Broaden Patient Data Sharing Efforts

Authors and Affiliations:

Juliann M. Savatt[1], Danielle R. Azzariti[2], W. Andrew Faucett[1], David H. Ledbetter[1], Vanessa Rangel Miller[3], Emily Palen[1], Heidi L. Rehm[2,4,5], Jud Rhode[3], Les Rogers[6], Sandra Talbird[7], Laura Trutoiu[8], Jo Anne Vidal[3], Erin Rooney Riggs[1], Christa Lese Martin[1]

[1]Geisinger, Danville, Pennsylvania; [2]The Broad Institute of Harvard and MIT, Cambridge, Massachusetts; [3]Invitae, San Francisco, California; [4]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts; [5]Harvard Medical School, Boston, Massachusetts, [6]CFC International, St. Petersburg, Florida, [7]CureCADASIL, Plainsboro, New Jersey, [8]Association for Creatine Deficiencies, Carlsbad, California

Abstract:
The Clinical Genome Resource (ClinGen) collaborates with laboratories, researchers, clinicians, and patients to promote broad sharing of genomic and phenotypic data to inform variant classification and gene–disease validity. ClinGen has facilitated patient data sharing through its patient registry, GenomeConnect, since 2014. The registry is open to anyone who has undergone genetic testing, and, to date, has shared 969 sequence variants with ClinVar. These efforts have resulted in 491 (51%) previously unsubmitted sequence variants being added to ClinVar, 18 variant classification updates being identified and shared with interested participants, and numerous requests for additional information from clinicians and researchers. Given GenomeConnect's success in engaging patients in the genomic data sharing process, we wanted to make those same opportunities available to patients participating in other registries who are not currently sharing data.

These external registries, many of which are disease-specific, can have ClinGen genetic counselors review participant genetic test reports to ensure standardized data collection and submit de-identified data from participants that opt to share with databases, such as ClinVar. Registry participants control their participation in data sharing efforts via an online consent within their registry. Those that opt to share are asked to upload a copy of their genetic testing report to their account, may be invited to complete additional surveys, and are provided with the option to receive updates regarding the reporting laboratory's interpretation of their variant. Since July 2018, three registries have begun recruiting participants into data sharing efforts – Association for Creatine Deficiencies, Cardio-Facio-Cutaneous Syndrome (CFC) International, and CureCADASIL. From these registries, 67 participants have opted to share data and 44.8% (n=30/67) have uploaded their genetic testing report. From these 30 participants, a total of 49 variants were shared, of which 55.1% (n=27/49) were novel to ClinVar. Of variants that were previously submitted by the reporting laboratory, 14% (n=2/14) were out of date compared to the current submission. Both participants have opted to receive updates regarding their results and will be informed of the update via email. Patient data sharing contributes valuable information to the public knowledge base that may not be otherwise available, benefiting both patients and the genetics community.

Real-Time Implementation of Precision Oncology Decision Support Demonstrates Improvements in Clinical Trial Matching and Overall Survival

Kenna R. Mills Shaw, Jia Zeng, Amber M. Johnson, Md Abu Shufean, Vijaykumar Holla, Dong Yang, Michael Kahle, Russell Broaddus, Scott Kopetz and Funda Meric-Bernstam. Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, MD Anderson Cancer Center, Houston, TX

Precision medicine can be defined as the use a biomarker to identify therapies for which an individual patient might receive therapeutic benefit. In oncology, there are just over fifty FDA approved biomarker-therapeutic indication combinations that help guide patient treatment decisions. There are additional treatment options that target specific biological pathways but do not require the presence of a biomarker as part of their indication. We maintain a dynamic decision support system called the Precision Oncology Decision Support (PODS) system that allows us to provide clinicians with real-time actionability classification of biomarkers in their patients regardless of the source of the biomarker (internal or commercial, NGS or protein, etc). This system couples expert manual curation with natural language processing-assisted curation to yield functional annotations on all alterations found in patient tumor samples. In coordination with proactive clinical trial alerts, and engaged follow-up with clinicians, we have seen a doubling of matching patients to clinical trials based on biomarker status as well as an improvement in overall outcome. Since being established in 2015, we have annotated 18,492 variants and issued 6,942 reports. It is important to note that only 25% of variants annotated to date can be considered therapeutically potentially actionable or actionable. Many patients are unable to be matched not because of a lack of associated trial, but because of progression of disease, suggesting that earlier biomarker testing could improve clinical trial enrollment rates. Moreover, less than half of our patients have a biomarker that could be characterized as actionable or potentially actionable for any of the 2,784 drugs annotated in PODS which could signify the need for both innovative drug development strategies as well as broader biomarker screening methods beyond traditional methods like NGS and immunohistochemistry.

A new syndromic curation method useful for gene curation prioritization

Rebecca K. Siegert[1], Sarah E. Hemphill[2], Madeline Y Hughes[1], Andrea M. Oza[1,3], Ahmad N. Abou Tayoun[4], Mustafa Tekin[5], Lisa A. Schimmenti[6], Margaret Kenna[3,7], and Heidi L. Rehm[1,2], on behalf of the ClinGen Hearing Loss Clinical Domain Working Group

[1]The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[2]Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA
[3]Dept. of Otolaryngology and Communication Enhancement, Boston Children's Hospital, Boston, MA, USA
[4]Al Jalila Children's Specialty Hospital, AL Jaddaf, Dubai, United Arab Emirates
[5]John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA
[6]Department of Otorhinolaryngology, Clinical Genomics and Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, USA
[7]Harvard Medical School, Boston, MA

The ClinGen Hearing Loss Gene Curation Expert Panel (HL GCEP) is focused on defining the validity of gene-disease pairs involving both nonsyndromic and syndromic HL. While HL is the only feature in nonsyndromic cases, it can vary in age of onset and penetrance compared to the other phenotypes in a given syndrome. Additionally, many genes have been linked to multiple syndromes. Before curation, the HL GCEP developed a precuration method, modified from the process developed by the ClinGen Lumping and Splitting Working Group, in order to parse through gene-disease pair claims and prioritize a list for curation. The method, referred to as 'syndromic curation,' was performed to prioritize the curation of 105 genes associated with nonsyndromic HL and 104 genes associated with syndromic diseases in which HL is a presenting feature or in which additional features of the syndrome could be overlooked during clinical evaluation (DiStefano et al. 2019, https://doi.org/10.1101/534040).

The syndromic curation method involves gathering phenotypic information for syndromic patients reported to have pathogenic variants in the gene of interest, including penetrance, age of onset, diagnostic or presenting features. We have developed a table display format for the recording of this information. The curated information was then reviewed by experts from the HL GCEP, though it did not undergo the rigorous approval process applied to full curations. The syndromic curation of 104 genes led to the inclusion of 46 gene-syndromic disease pairs in the HL GCEPs first round of full curation, along with 105 gene-nonsyndromic HL pairs. The 46 gene-syndromic disease pairs were selected because HL has been reported as a presenting feature in at least one individual.

The syndromic curation method can be utilized for gene list prioritization by other Clinical Domain Working Groups whose disease domain involves a phenotype that presents as nonsyndromic and syndromic. Use is also intended to inform inclusion of certain genes on sequencing panels that cover primarily nonsyndromic presentations but wish to include syndromes where other features might be missed. Additionally, this information is useful to inform clinical care of patients where the characteristics and penetrance of the isolated feature is not well understood within the syndrome. The syndromic curation has currently been performed for 104 HL gene-syndromic disease pairs, with planned discussion for adoption by other ClinGen EPs.

ClinGen Actionability curation tools and knowledge repository

Sai Lakshmi Subramanian[1], Andrew R. Jackson[1], Neethu Shah[1], Sameer Paithankar[1], Ronak Patel[1], Jessica Ezzell Hunter[2], Elizabeth M. Webber[2], Kathleen Mittendorf[2], Kristy Lee[3], Bradford Powell[3], Katrina Goddard[2], Aleksandar Milosavljevic[1], on behalf of the **Clinical Genome Resource**

1. Molecular and Human Genetics Department, Baylor College of Medicine, Houston, TX
2. Center for Health Research, Kaiser Permanente Northwest, Portland, OR
3. Department of Genetics, University of North Carolina, Chapel Hill, NC

Abstract:

The ClinGen Adult and Pediatric Actionability Working Groups (AWG) aim to assess the clinical implications of human genes and use a standardized protocol to produce evidence-based summary reports with scores for the actionability of gene-disease pairs. These reports of expert-curated gene-disease pairs will provide initial guidance to researchers and clinicians regarding findings based on clinical actionability and facilitate clinical decision making.The actionability workflow includes the Actionability Curation Interface (ACI), which is a web-based curation portal and REST-APIs to programmatically access structured content. The ACI and associated services are developed using Genboree Stack, following the microservices architecture and implemented as Redmine Ruby-on-Rails plugins. The application includes services for authentication, role based access control, allows tracking version histories and has a user friendly web interface for browsing, filtering, searching and export of curation data. The microservice model allows us to update components of the application to meet the requirements of the curators.

The ACI has been developed to streamline the curation process in both adult and pediatric contexts, which includes searching the literature for evidence, documenting the search process and references, performing an early rule out survey to determine if the topic is actionable, and abstracting data into a structured format. ACI also allows application of semi-quantitative scoring metric by experts, generation of summary reports, and dissemination of these reports in both human and machine readable formats. The application stores the latest guidelines and provides a transparent and systematic way to access supporting evidence and previous summary report versions, thus facilitating critical evaluation of current as well as prior recommendations in light of evolving evidence.

The AWGs actively use the ACI to curate and score topics of interest. As of February 2019, 86 genetic disorders have been evaluated and released in the adult context and 10 in the pediatric context. The reports are publicly accessible through actionability.clinicalgenome.org and are integrated with other curation efforts of ClinGen through clinicalgenome.org. REST-APIs enable retrieval of reports and summaries of individual and consensus scores. Together, these efforts

provide a centralized resource for the AWG curation and scoring activities and dissemination of knowledge to the public.

Annotation assisted biocuration expedites curation and enhances data capture.

Courtney Thaxton, Jen McGlaughon, Jenny Goldstein, Justyne Ross, Megan Mayers, May Flowers, C.Lisa Kurtz, Jonathan Berg

Curation of biologically relevant evidence (aka biocuration) is a fundamental aspect of the expert evaluation of genes and variants performed by ClinGen. The process of biocuration includes four major steps:

(1) searching, acquiring, and analyzing relevant literature (including clinical, genetic, and experimental evidence);
(2) assessing the evidence supporting the pathogenicity of the genetic variant(s) described;
(3) annotating the phenotypic features of each clinical case of interest; and
(4) applying the evidence based framework for gene or variant assessment.

Currently, ClinGen biocuration is largely a manual process, which can be laborious and time consuming. There are several areas within the biocuration process that may benefit from software development in order to expedite the curation process and reduce the amount of manual curation. The UNC Biocuration Core evaluated Hypothes.is, a freely available open source web annotation software platform, as a potential tool to reduce curation time, for both gene and variant level curation. For gene curation, two skill-matched biocurators were tasked to compare the time expenditure in curating 6 gene-disease relationships, one using Hypothesis annotation and the other using standard manual processes. In order to assess whether Hypothesis assisted annotation of literature could enhance variant curation, we tasked undergraduate volunteers to preannotate evidence on two genes of interest, *GAA* and *MYH7*. Blinded, skill-matched expert biocurators then compared the time and evidence captured over specific variants using either standard search platforms versus Hypothes.is pre-annotated information. Here, we will present the results of the gene and variant time trials which indicate that annotation not only expedites curation, but also enhances the curation experience through standardized and transparent data capture. Furthermore, we will outline our development of standardized tagging, future plans for automating Hypothes.is annotations and integration of the data within our software platforms, as well as present future aims to expand the crowd-sourcing to the greater community using Hypothes.is based annotation in order to “curate the clinical genome.”

Barriers to Genetic Data Sharing as Experienced by the Phelan-McDermid Syndrome International Registry.

Emily VandenBoom MS, CGC; Brittany McClarney MS, CGC; Megan O'Boyle; Phelan-McDermid Syndrome Foundation

Established in May of 2011, the Phelan-McDermid Syndrome International Registry (PMSIR) serves as a central repository for clinical and genetic data related to deletions of the 22q13 region and variants within the *SHANK3* gene. Caregivers provide data regarding overall health and development of affected individuals through questionnaires. Ideally the caregiver will upload a copy of genetic test results to the registry. Genetic reports are first reviewed by two independent genetic counselors who ensure the uploaded data are complete. Genetic data is then extracted from the report, deidentified, and shared with PhD level geneticist(s) for confirmation of a PMS diagnosis. Once confirmed, this genetic data is available for use in approved research studies. However, several barriers exist that prevent sharing of genetic data with external researchers, including: (a) Incomplete or illegible genetic test reports missing critical information needed to determine the diagnosis; (b) Lack of updated consent to include new data sharing preferences; and (c) Lab report is provided in native language and not easily translated to English. To resolve these issues, the caregiver is contacted which takes time and may be unsuccessful due to lack of response or inability to obtain needed information from the genetics clinic or laboratory. Further, in 2016, the PMSIR entered into a collaboration with Harvard Medical School, and all participants of the PMSIR were required to update their consent preferences to determine if they wanted to participate in the PMS_DN. PMSIR is an international registry and many uploaded laboratory reports require language translation services to minimize interpretation errors. The laboratory may also need to be contacted directly for clarification of reported results, as errors in identifying interstitial versus terminal deletions on chromosomal microarray, for example, have occurred. Curation of the genetic data as performed in a sequence of steps by the genetic counselors and PhD-level geneticist(s), may also be a rate-limiting step for data output, increasing the time to analysis and publication. In addition, whole exome sequencing is now routinely ordered and yields a new set of data interpretation challenges related to confirmation of a PMS diagnosis. As the field of genetics evolves, accurate genetic data is necessary for genotype-phenotype correlations, and barriers to obtaining this data can limit research progress in a rare disorder like PMS.

Curating the Clinical Genome 2019
Washington, D.C. | May 29-31

Link: https://www.clinicalgenome.org/events-news/events-conferences/curating-the-clinical-genome-2019/

Submitting Author Information
Submitting Author Name: Jameson Voss
Submitting Author Email Address: jameson.d.voss.mil@mail.mil

Abstract

Population Level Screening for Genetic Risk of Sudden Cardiac Death in the US Military

Jameson Voss, MD, MPH, FACPM,[1] Albert Bonnema, MD, MPH,[2] Erin McAuley, PhD,[3] Pavithra Shivakumar, PhD,[3] Zeke Maier, PhD[3]
[1] Air Force Medical Support Agency, Defense Health Headquarters, Falls Church, VA, 22042, USA
[2] Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, 45433, USA
[3] Booz Allen Hamilton Inc., McLean, VA, 22102, USA

Background

Approximately 30% of ClinGen classified medically actionable genes are linked to sudden cardiac death. The US Military has substantially higher rates of sudden cardiac death than age matched competitive athletes; moreover, the implications of sudden death can be particularly grave among a population who operate costly and lethal weapon systems and routinely engage in high risk missions critical for national security. To enable early detection and intervention, multiple international specialty societies recommend genetic screening among those with a family history of sudden cardiac death.

Methods

We used the Defense Medical Epidemiology Database to determine current rates of a diagnosis of family history of sudden cardiac death based on the International Classification of Diseases 10 code Z82.41. We also explored the literature to estimate the yield and benefit if the military health system were to implement a comprehensive screening system involving standardized identification of those with a family history, genetic testing of those with a 1° relative who had

sudden cardiac death before age 40 (SCDY), counseling those with positive results, and providing beta-blockers over 10 years.

Results

From 2016-2017, there were a total of 465 diagnoses of Z82.41 in any diagnostic position over approximately 2.5M person years of observation. Prior population based survey results in Michigan suggest active case finding could identify 2.7% or ~34K with a 1° relative who had SCDY; thus current diagnostic practice appears to represent only ~1.4%. If all 34K were identified, genetic testing costs an estimated ~$6.8M and if it yielded 25% with actionable genetic diagnoses, there would be ~8.6K diagnoses. Genetic counseling and beta blocker use for 10 years costs an estimated ~$2.6M among this 8.6K. Assuming 2.5% absolute mortality risk reduction over 10 years would lead to an estimated 200 lives saved (>20% of all sudden cardiac deaths in DoD over 10 years) at estimated cost of $44K/life.

Conclusions

Overall yield is sensitive to each assumption. A pilot study is needed to determine real world cost and yield of a screening protocol. Available literature does not adequately characterize harms of screening, which may include unique harms in the military population. A digital repository of genetic data of military members would enable population specific estimation for program planning and retrospective screening of newly curated actionable variants.

Acknowledgements

The authors thank DHA/JPC-1 for funding #B0900-0521/0050. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Joint Program Committee 1, the Defense Health Agency, the Department of Defense, or the U.S. Government.

Curating the Clinical Genome (CCG) 2019 Abstract

Title:
ClinGen's Variant Prioritization Tool (VPT): A Tool for Prioritizing Clinical Variants at Scale

Authors:
Wand, H., Preston, C.G., Wright, M.W., Lopez Pineda, A., Wulf, B., Ritter, D.I., Patel, R. Y., Milosavljevic, A., Costa, H., Plon, S.E., Bustamante, C.D., Clinical Genome Resource (ClinGen)

Abstract:

ClinGen's Variant Curation Interface (VCI) is a community curation tool that was designed to support variant curation based on the ACMG/AMP Guidelines and serves as a platform for the standardized curation of clinical variants by ClinGen's Variant Curation Expert Panels (VCEPs). Prior to accessing the VCI, curators need to assess which variants should be prioritized for curation. Selecting the variants of interest is based upon gene- or domain-specific requirements set by each VCEP, and relies on researching clinically relevant data about each variant from multiple data sources. This task is now facilitated by the Variant Prioritization Tool (VPT), a tool external to the VCI which enables user-defined filtering and sorting of variants to produce a set of prioritized variants for import and curation at scale within the VCI.

Users of the VPT start with a list of variants within a single gene. This list can be a user-given list of variants, or all the variants within a given gene. Variants are identified using the ClinGen Allele Registry identifiers (CAids). The VPT then programmatically retrieves various types of genomic data about these variants (e.g. population frequency, *in silico* predictor scores, ClinVar status, molecular consequence, etc.) from ClinGen's Linked Data Hub (LDH), a publicly accessible, contributor-managed, web-API accessible repository of evidence. Users can simultaneously view aggregated information about the variants, and prioritize their list of variants for curation according to user-specified filtering and/or sorting requirements of these genomic data fields. This tool is open to any user as an anonymous search, but registered VCI users will have the ability to take their prioritized list of variants and batch import them directly into the VCI to start the curation process. Registered users can also store their search inputs and filter settings.

It is intended that the VPT will be expanded to include further functionalities necessary for curation at scale. For instance, it paves the way for future implementation for automated criteria evaluation whereby variants meeting specified criteria thresholds could be automatically scored with the associated ACMG evidence code. We also welcome your feedback on any features you would like see included in the VPT.

The VPT was initially developed via a collaboration between the Stanford University & Baylor College of Medicine ClinGen teams and their cloud provider, Amazon Web Services (AWS).

Alistair Ward[1,2], Tonya Di Sera[1], Matt Velinder[1], Aditya Ekawade[1], Gabor Marth[1,2]
[1]University of Utah, Department of Human Genetics, Salt Lake City, UT
[2]Frameshift Genomics, Boston, MA

Web tools for physician-driven genomic analysis

DNA sequencing is already supporting the diagnosis of inherited diseases, cancers and other genetic disorders. It is becoming increasingly important for subspecialty clinicians to carry out analysis in their own patients' genomic data: however, existing genomic analysis software tools require substantial involvement from bioinformaticians. We are developing a tool to guide physicians through all necessary steps of clinical diagnostic genomic analysis, from data quality review, candidate gene generation, variant interrogation and prioritization, to reporting of all findings, in a visual, compelling fashion. This tool, ***clin.iobio***, is based on our popular IOBIO web apps, and performs analyses in real time, without requiring the help of bioinformaticians.

Clin.iobio provides a directed workflow through the analysis process, by combining multiple IOBIO apps into a single platform. Following the lead of other medical procedures, a checklist of tasks promotes analysis consistency and repeatability. A user begins by building a candidate gene list using ***genepanel.iobio*** (http://genepanel.iobio.io). This app combines NCBI's genetic testing registry, the Phenolyzer tool, and the user's knowledge of likely genes to generate this gene list. These genes are then propagated to ***gene.iobio*** (http://gene.iobio.io) for real-time variant annotation and analysis. Variants are automatically prioritized by ***gene.iobio***, and users are encouraged to assign a significance level and attach notes to variants to promote collaboration. Finally, variants drawn from the whole exome or genome, independent of the candidate gene list, and conforming to predetermined filters are presented for review. This ensures all potentially disease-causing variants in all genes are examined. The analysis findings are summarized in a final, interactive report. An important feature of ***clin.iobio***, is that it facilitates dynamic analysis and re-analysis. If no candidate variants are discovered, the user can quickly create a more expansive gene list, bringing in additional variants for review. This dynamic and collaborative analysis platform will help move genomic analysis into the clinical practice of subspecialty clinicians.

Structured Narrative of Functional Assays to Support the Determination of Damaging Effect on Protein Function

Matt W. Wright, PhD[1], Michael Iacocca, MS[2], Arturo Lopez Pineda, PhD[1], Hannah Wand, MS[1], Bryan Wulf, BS[1], Ronak Y. Patel, PhD[4], Aleksandar Milosavljevic, PhD[4], Deborah Ritter, PhD[4], Snehit Prabhu, PhD[1], Helio A. Costa, PhD[1], Joana R. Chora, MS[3], Mafalda Bourbon, PhD[3], Sharon E. Plon, MD, PhD[4], Carlos D. Bustamante, PhD[1], ClinGen's Cardiovascular Familial Hypercholesterolemia Variant Curation Expert Panel, and Clinical Genome Resource (ClinGen)

1. Stanford University, Stanford, California, USA
2. Western University, London, Ontario, Canada
3. Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisbon, Portugal
4. Baylor College of Medicine, Houston, Texas, USA

Abstract:

There is an important need to understand the functional consequence of a specific genomic change to a gene or protein when interpreting a variant in the context of pathogenicity determination. The joint guidelines of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) requires 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.

The Clinical Genome Resource (ClinGen)'s Familial Hypercholesterolemia (FH) Variant Curation Expert Panel manually curated a comprehensive list of 128 peer-reviewed publications describing functional studies of variants identified in FH patients. This body of literature identified 305 variants in three genes: APOB (19 variants), LDLR (263 variants), and PCSK9 (23 variants). We reviewed the diversity of biochemical experiments, and heterogeneity of results with the goal to find a pattern that could promote consistency across the aggregated results.

We propose the use of structured narratives with the structure of 1) method, 2) material, and 3) effect (with or without a quantifier), using standardized terminology from BioAssay Ontology (BAO), Methods and Materials Ontology (MMO), Cell Line Ontology (CLO), and Gene Ontology (GO). For example, a given publication could refer to the use of the luciferase reporter gene assay (method assigned to BAO:0002661) with Hep G2 cells (material assigned to CLO:0003704) finding 12% gene expression (effect assigned to GO:0010467, with a quantifier of 12%). For qualitative findings (i.e. abnormal transcript lenght), the quantifier would not be used.

We found that only 13 terms were used to describe methods, 17 for materials, and 22 for effects. For a panel of experts, like the ClinGen's FH-VCEP, deliberating about the correct assignment of a reduced number of terms is far less challenging than being familiar with all available literature.

Our structured narrative have the potential to be used by other expert panels and working groups in the ClinGen ecosystem. With the proper validation and expansion of existing ontologies, we believe that our framework can accelerate the curation of functional data in accordance with the BS3/PS3 criteria of the ACMG/AMP guidelines. In the future, with defined rules for numerical thresholds, this framework could pave the way for automated scoring of BS3/PS3.

The Effect of Removing Reputable Source Criteria from ACMG/AMP 2015 Guidelines: Using Variants from the Personalized Diabetes Medicine Program (PDMP) as an Example

Haichen Zhang[1], Kristin Maloney[1], Toni I. Pollin[1], Linda J.B. Jeng[1]

1. Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD

Introduction

The 2015 American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) variant classification guidelines have been broadly implemented and proven to increase consistency of variant classification among laboratories. Still, many groups have recommended changes to these guidelines based on their experience with particular genes and diseases. Recently, the ClinGen Sequence Variant Interpretation Working Group (SVI-WG) recommended that the reputable source criteria (PP5/BP6) be removed. By comparing the classification of variants from the Personalized Diabetes Medicine Program (PDMP), we evaluated whether and to what extent removing the PP5/BP6 criteria would result in changes in variant classification.

Methods

The analysis included 130 variants from nine monogenic diabetes-related genes in which at least one reportable variant was identified in the PDMP: *GCK* (23), *HNF1A* (13), *HNF4A* (8), *INS* (2), *KCNJ11* (7), *HNF1B* (9), *LMNA* (6), *WFS1* (53), and *MC4R* (9). The initial classification was performed according to the ClinGen SVI-WG modified ACMG/AMP guidelines (i.e., without PP5/BP6). A reputable source criterion was then added if the variant was recently (after 2000) classified as pathogenic in ClinVar or LOVD, but no case-level evidence was available. We then compared the classification outcomes of these two strategies.

Results

The PP5 criterion could be applied to 14 (10.8%) variants in five genes: *GCK* (7/23), *HNF1A* (3/13), *LMNA* (2/6), *WFS1* (1/53), and *MC4R* (1/9). The classification of one (0.7%) variant (*HNF1A*) changed from likely pathogenic to pathogenic after adding PP5, but this would not impact reportability in the PDMP. The BP6 criteria could be applied to 23 (17.7%) variants in seven genes: *GCK* (1/23), *HNF1A* (1/13), *HNF4A* (1/8), *KCNJ11* (1/7), *LMNA* (1/6), *WFS1* (17/53), and *MC4R* (1/9). The addition of BP6 to four (3.0%) variants (one in *LMNA*, three in *WFS1*) changed classification from VUS to likely benign, similarly not affecting reportability in the PDMP.

Discussion

The PP5/BP6 criteria could be applied to 28.5% of the analyzed variants. Adding PP5/BP6 changed the classification of five (3.8%) variants with no changes to reportability in the PDMP. In summary, we provide evidence that at least for monogenic diabetes, removing PP5/BP6 does not significantly impact variant classification. For clinical laboratories reporting VUS but not benign/likely benign variants, there may be a small impact on reportability.