Platform and Poster Presentations
Curating the Clinical Genome 2017

Platform Presentations

Thursday, June 29, 2017

Curating Gene-Disease Relationships

8:30am  Use of the ClinGen Clinical Validity Framework to Evaluate the Strength of Evidence for Genes Implicated in Hypertrophic Cardiomyopathy  
  Jennifer Goldstein, University of North Carolina at Chapel Hill

8:45am  The Impact of Community Curation of Gene-Disease Relationships for Clinical Genome Analysis  
  Ellen McDonagh, Genomics England

Perspectives and Regulations on Data Sharing

10:00am  The BRCA Exchange: Global Data Sharing and Knowledge Exchange to Enable Accurate Clinical Care  
  Lena Dolman, Global Alliance for Genomics and Health (GA4GH)

10:15am  Resolving Variant Interpretation Differences in ClinVar between 43 Clinical Laboratories  
  Steven Harrison, Harvard Medical School

Evolving Guidelines/Resources to Support Variant Assessment

2:15pm  Modeling the ACMG/AMP Variant Classification Guidelines as a Bayesian Classification Framework  
  Leslie Biesecker, National Institutes of Health

2:30pm  Merging Single Gene-Level CNV with Sequence Variant Interpretation Following the ACMGG/AMP Sequence Variant Guidelines  
  Tracy Brandt, GeneDx

Friday, June 30, 2017

Functional Genomics Aiding Clinical Interpretation

10:45am  Functional Annotation of Human Ion Channel Variants of Unknown Significance Using Automated Electrophysiology  
  Al George, Northwestern University Feinberg School of Medicine

11:00am  Multiplex, Prospective Identification of Unstable Pathogenic Variants of Clinically Important Genes  
  Kenneth Matreyek, University of Washington
Use of the ClinGen clinical validity framework to evaluate the strength of evidence for genes implicated in hypertrophic cardiomyopathy.

Jodie Ingles¹, Colleen Caleshu², Edward W. Corty³, Stephanie Crowley³, Kristen Dougherty⁴, Jennifer Goldstein⁵, Jennifer McLaughon⁵, Laura Milko⁵, Ana Morales⁵, Bryce Seifert⁵, Chris Semsarian¹, Natasha Strande³, Courtney Thaxton³, Kate Thomson³, Peter van Tintelen⁷, Kathleen Wallace⁸, Roddy Walsh⁸, James Ware⁸, Quinn Wells⁹, Nicola Whiffin⁸, Leora Wikowski¹⁰, Ray Hershberger⁵, Birgit Funke¹⁰, on behalf of the ClinGen Cardiovascular Clinical Domain WG

¹Sydney Medical School, University of Sydney, Australia; ²Stanford University, CA, USA; ³UNC Chapel Hill, NC, USA; ⁴Eastern Virginia Medical School, VA, USA; ⁵Ohio State University, OH, USA; ⁶Oxford Medical Genetics Laboratory, Oxford, UK; ⁷Academic Medical Center, Amsterdam, Netherlands; ⁸Imperial College London, London, UK; ⁹Vanderbilt University Medical Center, TN, USA; ¹⁰Harvard Medical School/MGH, MA, USA.

Hypertrophic cardiomyopathy (HCM) is defined as unexplained left ventricular hypertrophy (LVH) in the absence of other causes. The clinical presentation ranges from asymptomatic LVH to progressive heart failure, arrhythmia, and sudden cardiac death. In the majority of cases, HCM is caused by a pathogenic variant in one of 8 genes encoding components of the cardiac muscle sarcomere. Z-disc and calcium-handling genes have also been implicated. Commercial laboratory gene testing panels can include 50 or more genes with variable levels of evidence to support their association with HCM. Using the Clinical Genome Resource (ClinGen) gene-disease clinical validity framework, we curated publicly available data to classify the clinical validity for 55 genes and HCM. The gene list was based on (1) Presence on >30% of HCM gene panels in the NCBI Genetic Testing Registry, (2) Inclusion in the TruSight HCM gene panel (Pua et al, 2016, PMID: 26888179), and (3) Expert opinion. Curations were performed by biocurators, overseen by a cardiac expert, and discussed on regular conference calls to reach a final gene-disease validity classification. Initially, we targeted genes that appeared to be associated with isolated HCM. Of 37 genes curated to date, 8 had a definitive association, 2 strong, 4 moderate, 16 limited, and 7 had no reported evidence. As we began to curate genes associated with HCM as part of a wider syndrome (e.g. GLA, FHL1) we recognized the need to fully evaluate the range of disorders caused by all of the genes on our list, and whether it was appropriate to “split” the curation to include only individuals with isolated HCM or “lump” cases into a wider syndrome. Following the recommendations of the newly-formed ClinGen Lumping and Splitting working group, we binned all 55 genes on our list into one of three groups: genes causing isolated HCM, genes associated with a complex cardiac phenotype, and syndromic genes that include HCM as part of the phenotypic spectrum. We will present the rationale for the binning process and our curation results to date. We conclude that the ClinGen clinical validity classifications will be beneficial to the development and interpretation of genetic testing in patients with HCM.
The Genomics England 100,000 Genomes Project aims to sequence, analyse and interpret the genomes of around 50,000 rare disease patients and their relatives. The hope is to provide a diagnosis for the underlying cause of the disorder and identify treatment options. Simultaneously, the infrastructure for integrating genomic medicine into routine clinical practice within the National Health Service (NHS) in England is being established. Curation is a fundamental element in the interpretation of genomes within the project, as is engaging members of the clinical, research and industry community.

Analysis of rare disease genomes within the 100,000 Genomes Project includes the use of virtual gene panels, comprised of genes with a diagnostic-grade level of evidence for disease association. This helps filter the millions of variants in each genome to identify those that are potentially causative. It is well known that panel-based tests for the same disease can differ across diagnostic labs; the challenge for the Curation Team at Genomics England is to establish a consensus list of clinically-relevant genes with a high confidence level in order to allow genome analysis and reporting. The Genomics England “PanelApp” database (https://panelapp.extge.co.uk/crowdsourcing/PanelApp) was established to allow the curation of initial virtual gene panels, crowdsourcing of reviews by experts from the clinical and scientific community, and revision to establish a diagnostic-grade virtual gene panel. PanelApp has more than 520 registered reviewers from over 20 countries worldwide, contains >3600 genes and has 150 revised gene panels enabling genome analysis for more than 150 rare disease categories.

We present a preliminary analysis of diagnostic candidates using gene panels before and after review, clinical input and further curation, to show the importance of how a coordinated, combined curation effort can assist genome interpretation and help gain patient diagnoses. The existence of publically available curated resources such as OMIM, Orphanet and Gene2Phenotype is vital in supporting this process. Rules and procedures to evaluate evidence for a gene-disease association were established to gain concordance across different reviewers and curators, and iterative development of PanelApp tools to improve usability and aid curation. The challenges of crowdsourcing knowledge and the key lessons learned from the process will be presented.
The BRCA Exchange: global data sharing and knowledge exchange to enable accurate clinical care

Lena Dolman\textsuperscript{1,2}, Rachel Liao\textsuperscript{1,3}, Stephen Chanock\textsuperscript{4,5}, and Sir John Burn\textsuperscript{5,6} on behalf of the BRCA Challenge Steering Committee

Affiliations:
\textsuperscript{1}Global Alliance for Genomics and Health (GA4GH)
\textsuperscript{2}Ontario Institute for Cancer Research (OICR)
\textsuperscript{3}Broad Institute of MIT and Harvard
\textsuperscript{4}U.S. National Cancer Institute (NCI)
\textsuperscript{5}Co-Chair of BRCA Challenge Steering Committee
\textsuperscript{6}University of Newcastle upon Tyne

The \textit{BRCA1} and \textit{BRCA2} genes are frequently tested and associated with well-known hereditary cancer risks. However, variant interpretation efforts can be hampered by the lack of a comprehensive resource displaying all public \textit{BRCA1/2} variant information in one place, and the inability to access additional evidence not yet in the public domain. In response to this, the BRCA Challenge project was proposed in 2014 as a demonstration project of the Global Alliance for Genomics and Health (GA4GH) to tackle the challenge of bringing international communities and resources together around a key disease area. To advance understanding of the genetic basis of hereditary breast, ovarian, and other cancers, and to better enable accurate clinical care, this project proposed to: i) share \textit{BRCA1/2} variants publically; ii) create a curated list of \textit{BRCA1/2} variants, interpreted by expert consensus from the ENIGMA Consortium, and; iii) create an environment for collaborative variant curation with access to evidence. In 2016, the “BRCA Exchange” web portal (brcaexchange.com) was launched to support these goals. Today, the BRCA Exchange includes over 18,000 unique, de-duplicated variants, making it the largest public source for \textit{BRCA1/2} variant information. Using the GA4GH Genomics API, this portal aggregates variant data from public sources (including ClinVar, LOVD, 1000 Genomes, ExAC, BIC, ENIGMA, and ESP) with automated monthly updates. The portal consists of an “Expert-Reviewed” tier, which supports clinical care by displaying consensus expert classifications (and associated evidence) for almost 5,000 variants as assigned by ENIGMA, and an “All Public Data” research tier, which displays a range of evidence (including allele frequencies, predictive algorithms, and assertions of pathogenicity) for the entire dataset, derived from the original submitters. The full dataset is downloadable and versioned, enabling research and improved speed and efficiency for curators. Features currently under development include the production of a mobile app to allow patients and providers to “follow” variants of interest for classification changes, and the addition of a third, credentialled-access tier for aggregation of case-level evidence and collaborative community review to resolve VUSs. The BRCA Exchange ultimately enables cyclical knowledge exchange among researchers, care providers, and diagnostic labs, allowing for improved patient management in the hereditary cancer domain.
Resolving Variant Interpretation Differences in ClinVar between 43 Clinical Laboratories

Steven M. Harrison¹, Jill S. Dolinsky², and Heidi L. Rehm¹
¹Partners HealthCare Laboratory for Molecular Medicine, Cambridge, MA, USA; ²Ambry Genetics, Aliso Viejo, CA, USA

Sharing data in ClinVar provides open access to variant classifications from many clinical laboratories. While the majority of classifications agree, ClinVar has shed light on the important issue of interpretation differences between laboratories, providing a valuable opportunity to resolve differences and positively impact patient care. Recent work with four clinical laboratories found that 53% of interpretation differences were resolved by either updating ClinVar with current internal classifications or reassessment of an older interpretation with current classification criteria (PMID: 28301460). With this finding in mind, ClinGen’s Sequence Variant Inter-Laboratory Discrepancy Resolution team will encourage clinical laboratories with outlier interpretations to update ClinVar with current classifications and reassess remaining conflicts using current guidelines. To identify variants that could be resolved by this outlier strategy, interpretations from 43 clinical laboratories in ClinVar were compared, identifying 26,421 variants interpreted by ≥2 clinical laboratories. The majority of classifications were concordant (85.7%; 22,637 variants). Only 2.5% (667 variants) of all shared variants were medically significant differences (MSDs) with potential to impact medical management [pathogenic (P/LP) versus other (VUS/LB/B)]. These differences were investigated to determine if submitted interpretations could reach a majority consensus (agreement in classification of at least 2/3 of clinical laboratory submitters). Of the MSDs with ≥3 interpretations (249 variants), 87.6% (218 variants) reached a majority consensus, thus allowing for identification of outlier submissions most in need of reassessment. Outlier submitters on variants with majority consensus will be contacted with a custom report and be encouraged to update ClinVar, if the classification has already changed internally, and reassess remaining outlier interpretations. If the discrepancy remains, other clinical laboratories will be encouraged to share internal evidence to facilitate resolution. Based on our initial study results, it is anticipated that this process will resolve at least 79% of MSDs, reducing total MSDs to 0.5%. This process adds to the value of ClinVar and will help the community move toward more consistent variant interpretations which will improve the care of patients with, or at risk for, genetic disorders.
Modeling the ACMG/AMP Variant Classification Guidelines as a Bayesian Classification Framework

Sean V. Tavtigian,1 Marc S. Greenblatt,2 Steven M. Harrison,3 Robert L. Nussbaum,4 Snehit A. Prabhu,5 Kenneth M. Boucher,6 and Leslie G. Biesecker,7 on behalf of the ClinGen Sequence Variant Interpretation Task Team (ClinGen SVI)

1. Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City UT.
2. University of Vermont Robert Larner, MD, College of Medicine, Burlington, VT.
3. Partners HealthCare Laboratory for Molecular Medicine and Harvard Medical School, Boston, MA.
4. Invitae, San Francisco, CA.
5. Department of Genetics, Department of Biomedical Data Science, Stanford University, Palo Alto, CA.
6. Department of Internal Medicine, Division of Epidemiology and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City UT.
7. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published a guideline for variant interpretation meant to apply broadly to Mendelian disease gene variants (PMID: 25741868). The committee recognized that the guideline was a starting framework that would evolve over time. ClinGen’s Sequence Variant Interpretation (SVI) Working Group aims to standardize application of the ACMG/AMP guidelines by providing recommendations to adapting the guidelines and converting qualitative criteria to quantitative, where applicable. For this goal, the ACMG/AMP guidelines were evaluated to determine whether the ACMG/AMP rules are internally consistent and whether the systematic, qualitative, categorical ACMG/AMP “rules for combining criteria” to evaluate variants represent a Bayesian heuristic similar to that used for hereditary cancer variants (PMID: 18951446; PMID: 7932824). The ACMG/AMP criteria were translated into a naïve Bayesian classifier, assuming four levels of evidence and exponentially scaled odds of pathogenicity. Results of this analysis indicate that the majority of ACMG/AMP pathogenicity classifications were compatible. However, one ACMG/AMP likely pathogenic combination (1 Very Strong criterion and 1 Moderate criterion) was shown to be equivalent to pathogenic and one ACMG/AMP pathogenic combination (2 Strong criteria) was actually likely pathogenic. Combinations that include evidence for and against pathogenicity were also modeled, and while some resulted in a posterior probability in the VUS range, our approach demonstrated that weak evidence against pathogenicity, in combination with strong evidence for pathogenicity, can lead to posterior probabilities in the range of likely pathogenic or even pathogenic. By transforming the ACMG/AMP recommendations into a Bayesian framework, we provide a mathematical foundation for what was a qualitative heuristic. Only two of the 18 existing ACMG/AMP evidence combinations are mathematically inconsistent with the overall framework. This quantitative framework validates the approach adopted by the ACMG/AMP, provides opportunities to further refine evidence categories and combining rules, and supports efforts to automate components of variant pathogenicity assessments.
Merging single gene-level CNV with sequence variant interpretation following the ACMGG/AMP sequence variant guidelines

Tracy Brandt, Lisa Vincent, Dolores Arjona, Gabriele Richard, and Jeanne Meck

GeneDx, Gaithersburg, MD

In 2015 ACMGG, in conjunction with AMP, published standards and guidelines for the interpretation of sequence variants. These guidelines discuss the classification of sequence variants based on a quantitative criteria system, but also consider exon-level copy number variants (CNVs). Our clinical diagnostics laboratory rapidly incorporated these guidelines into our sequence variant interpretation workflow and, as recommended in the publication, improved upon them by developing gene and disease-specific modifications. In addition, our robust internal system makes this classification method computable, efficient and reproducible. Implementation of these same criteria for analysis of intragenic or whole-gene CNVs is the next logical step given the growing trend towards using Next Generation sequencing as a single platform to obtain both sequence and CNV data. There are interpretation challenges unique to CNVs. For example, compared to sequence-level changes, it can be more difficult to relate a CNV detected in a patient to CNVs reported in the literature or in databases due to different start/stop locations, sometimes secondary to alternative technologies. In addition, interpretation of duplications poses a greater challenge when location and orientation is not known. CNV interpretation utilizes specialized resources (e.g. Decipher, DGV) with different limitations compared to sequencing variant resources. However, many of the same ACMGG/AMP criteria are also directly applicable to CNVs with only minor customizations required. Examples of such criteria include those considering: parentage and segregation studies (PS2, PM6), prevalence in control populations (PM2, BA1, BS1, BS2), and gene-specific criteria like whether the variant impacts a functional domain (PM1). As overlapping technologies act as a driving force for the merger of clinical molecular genetics and clinical cytogenetics laboratories, the most efficient and consistent workflow would be to use the same variant interpretation system for assessing sequence and copy number alterations. Unification of sequencing and CNV classification guidelines would also reflect the integration of molecular and cytogenetics in the newly formed ABMGG training program known as Laboratory Genetics and Genomics. Furthermore, as we begin to adapt these guidelines for single-gene CNV interpretation, it may provide a natural progression towards inclusion of multi-genic events such as those found on chromosomal microarray.
Functional Annotation of Human Ion Channel Variants of Unknown Significance Using Automated Electrophysiology

Carlos G. Vanoye¹, Reshma R. Desai¹, Katarina L. Fabre¹, Franck Potet¹, Jean-Marc DeKayser¹, Jen Meiler², Charles R. Sanders³ and Alfred L. George, Jr¹.

¹Department of Pharmacology, Northwestern University Feinberg School of Medicine, Departments of Biochemistry and ³Chemistry and ⁴Center for Structural Biology, Vanderbilt University School of Medicine

Ion channels make up ~1% of all genes in the human genome and mutations in >50 ion channel genes cause several monogenic disorders collectively called channelopathies. The emergence of widespread clinical genetic testing and the use of next-generation sequencing in genetic research has resulted in explosive growth of known ion channel variants associated with disease traits and in populations. Differentiating pathogenic from benign variants and establishing genotype-phenotype relationships has become very challenging. Functional annotation experiments (e.g., patch-clamp recording) are the gold standard in assessing the likely pathogenicity of ion channel variants, but the extreme time- and labor-intensity of this approach is insufficient to tackle the thousands of known variants. To address this challenge, we have employed a novel technical approach for performing functional analyses of human ion channel variants at an unprecedented scale using an automated electrophysiology platform. In this study, we targeted 51 nonsynonymous variants in KCNQ1 encoding a cardiac potassium channel implicated in congenital long-QT syndrome (LQTS). Variants were collected from several sources including the Human Gene Mutation Database (HGMD), ClinVar, Exome Aggregation Consortium (ExAC) database and a series of engineered variants based on conservation of the gene among species. All variants introduced amino acid substitutions within the functionally critical voltage-sensor domain. We implemented a work flow that combined site-directed mutagenesis, high efficiency electroporation to achieve transient expression of KCNQ1 cDNA in cultured cells and automated planar patch clamp recording performed in 384-well format to generate high fidelity electrophysiological measurements of functional properties. Results from approximately 20% of the variants tested in this manner were validated by conventional patch-clamp recording. Results from high throughput electrophysiological analyses enabled classification of KCNQ1 variants into defined functional categories (normal, near normal, mild loss-of-function, severe loss-of-function, gain-of-function). These classifications can aid the genetic classification of variants, particularly those categorized as variants of unknown significance. Our study demonstrates the feasibility and robustness of automated electrophysiology for the functional annotation of human ion channel variants.
Multiplex, prospective identification of unstable pathogenic variants of clinically important genes

Kenneth A. Matreyek1,4, Lea M. Starita1,4, Jason J. Stephany2, Beth Martin1, Melissa Chiasson1, Vanessa E. Gray1, Martin Kircher1, Arineh Khechaduri1, Ronald J. Hause1, Jay Shendure1,3,*, Douglas M. Fowler1,2,*

1 Department of Genome Sciences, University of Washington, Seattle, Washington, USA.
2 Department of Bioengineering, University of Washington, Seattle, Washington, USA.
3 Howard Hughes Medical Institute, Seattle, Washington, USA.
4 These authors contributed equally to this work.
* Corresponding authors

Modern sequencing technologies have uncovered millions of novel missense variants. However, the use of these variants for the diagnosis and treatment of disease has been stifled by a lack of understanding of each variant’s effects. One-at-a-time variant phenotyping using protein-specific assays is prohibitively slow, so generalizable, high-throughput assays are needed to keep pace with the many thousands of missense variants that will be discovered for each of the thousands of clinically important genes. Since correct folding and expression is critical for the function of almost all proteins, we developed the Multiplex Assessment of the Abundance of ProteinS (MAAPS) assay. MAAPS combines a human cell-based, fluorescent reporter system with flow cytometry and deep sequencing to quantify the effect of missense variation on the intracellular abundance of thousands of variants of a protein simultaneously.

We applied MAAPS to PTEN and TPMT, two important, cancer-related proteins, obtaining abundance scores for approximately 3,700 PTEN, and 2,900 TPMT single amino acid variants, respectively. Roughly 80% of germline PTEN missense or nonsense variants classified as pathogenic in ClinVar were of low abundance in the MAAPS assay, demonstrating the predictive value of MAAPS. Accordingly, inclusion of our newly identified low abundance variants would nearly quadruple the list of pathogenic PTEN variants in ClinVar. We also observed differences in the patterns of PTEN variant abundance between different types of cancers like glioma and melanoma, possibly reflecting cancer type-specific driver pathways. In TPMT, we found that variants known to require lower thiopurine drug dosing were of low abundance. This suggests that other low abundance TPMT variants revealed by MAAPS will also require thiopurine dosing adjustment should they be found in patients.

Finally, we show that MAAPS can be applied to other proteins including VKORC1, CYP2C9 and CYP2C19. Thus, MAAPS is a broadly applicable method for quantifying variant abundance. We suggest that MAAPS will empower variant interpretation by enabling prospective and comprehensive identification of low abundance, pathogenic variants in clinically important genes.
1. Nicole J. Burns The TruGenome Predisposition Screen and the Donation of Over 106,000 Unique Variants Found in Ostensibly Healthy Individuals to ClinVar (Poster #4)
2. Edward C. Cooper An Informatics Infrastructure for KCNQ2 Associated Disorders Research Including Patient Registry, Database, Curation Platform, and Website (Poster #5)
4. Marina T. DiStefano Curating Clinically Relevant Transcripts for the Interpretation of Sequence Variants (Poster #9)
5. Ada Hamosh Criteria for OMIM for Establishing Gene-Phenotype Relationships in the Era of Whole Genome/Exome Sequencing (Poster #15)
6. Sarah E. Hemphill Progress in Evaluating the Clinical Validity of Gene-Disease Associations in Hearing Loss (Poster #16)
7. S. Mohsen Hosseini Gene Curation for Brugada Syndrome Questions the Clinical Validity of Some of the Previously Reported Gene-Disease Associations (Poster #18)
8. Melissa J. Landrum ClinVar Supports Multiple Approaches to Access to Variant Interpretations for the ACMG 59 Genes (Poster #22)
10. Peter McGarvey Combining Protein and Genome Annotations for Deciphering Functional Effects of Variation (Poster #28)
12. Deborah I. Ritter Cancer Variant Curation for Clinical and Public Use: Disseminating Minimum Variant Level Data (MVLD) through Collaboration and Curation (Poster #41)
13. Juliann M. Savatt Sharing Patient-Derived Data in ClinVar via GenomeConnect (Poster #44)
14. Scott Topper Challenges and Solutions in Implementing the ACMG/AMP Variant Classification Guidelines in a Large Reference Laboratory (Poster #51)
15. Kathleen Wallace Defining the Pediatric Actionability of Genetic Conditions for Utility in Newborn Screening (Poster #54)
16. Roddy Walsh A Quantitative and Disease- and Gene- Specific Approach to Variant Interpretation Improves the Yield of Genetic Testing (Poster #55)
17. Elizabeth M. Webber Beyond the ACMG 59: Identification of Clinically Actionable Secondary Findings by the ClinGen Actionability Working Group (Poster #57)
18. Zena Wolf Adapting the ACMG/AMP Variant Classification Framework for Inherited Cardiomyopathy: Recommendations by ClinGen's International Cardiomyopathy Expert Panel (Poster #58)
1. Development of Oncomine Knowledgebase Reporter, an information system to associate public evidence with cancer gene variants detected by next-generation sequencing tests

Santhoshi Bandla, David Galimberti, Ken Kopp, Sarah Anstead, Christopher Zurenko, Nikki Bonevich, Dinesh Cyanam, Jane Faershtein, Habib Hamidi, Jody McIntyre, Michael Hogan, Seth Sadis

Introduction: Next-generation sequencing (NGS) is displacing single-gene based molecular testing in many oncology applications. However, the increased scale of data generation and interpretation required for multi-variant NGS tests is a challenge for many laboratories. We asked whether a combination of programmatic and manual curation could create a scalable and sustainable information management system to associate relevant public evidence with gene variants detected by NGS tests.

Methods: To define recurrent somatic alterations in solid tumors, we created a compendium of variant calls from > 15,000 exomes, defined focal amplifications and deletions from > 30,000 arrays, and defined recurrent fusions from several thousand RNAseq profiles. We determined that drug labels, treatment guidelines, and clinical trial enrollment criteria were relevant public source information to help cancer researchers understand the relevance of gene variants detected by NGS assays. Candidate evidence was identified using automated text search pipelines followed by manual verification and validation before committing the content to a knowledgebase. To support generation of custom reports with relevant content, we developed a web application wherein the user can upload a variant call format file (VCF), select a cancer type, set appropriate content filters, review the custom report, and finally export in a PDF format.

Results: We developed a scalable information management system made up of two components—a content knowledgebase and a web based application to create custom reports. The knowledgebase contains gene-variant evidence for over 115 genes across 36 approved drug labels, 19 cancer type specific treatment guidelines and over 800 clinical trials with global recruiting locations (from over 50 trial registries). Reports can be customized to include select data sources, trial locations, and trial phases.

Conclusion: We demonstrate that a scalable information management system can be developed to return relevant evidence associated with cancer gene variants detected by NGS tests. The Oncomine Knowledgebase Reporter can be installed locally or accessed through the Thermo Fisher Cloud.

For Research Use Only. Not for use in diagnostic procedures.
2. Applying the ACMG variant classification guidelines within a UK NHS diagnostic exome sequencing service

Emma L Baple\textsuperscript{1,2}, K. L. Stals\textsuperscript{1}, J. Baptista\textsuperscript{1,2}, C. Tysoe\textsuperscript{1}, M. Wakeling\textsuperscript{2}, G. Jones\textsuperscript{1}, A. Parrish\textsuperscript{1}, R. Caswell\textsuperscript{2}, S. Ellard\textsuperscript{1,2}

\textsuperscript{1}Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom, \textsuperscript{2}University of Exeter Medical School, Exeter, United Kingdom.

We provide the only NHS diagnostic trio whole exome service, receiving referrals from across UK Regional Genetics Services. Children with a high prior probability of a rare monogenic disorder are selected for testing through multi-disciplinary assessment. Most patients have undergone extensive prior genetic testing. We started using the ACMG guidelines in 2016 to classify variants identified through a gene-agnostic, inheritance based whole exome strategy. Variants were reviewed by clinical scientists and final classification agreed in a multi-disciplinary team meeting with clinical input.

We achieved a 42% diagnostic yield through identification of pathogenic or likely pathogenic variants in 182 cases. Approximately 20\% of variants were in genes linked to the disease phenotype within the previous year and \textasciitilde50\% were novel, with limited information available for classification. Two cases had non-canonical splice site variants predicted to affect splicing by \textit{in silico} tools. Analysis of messenger RNA extracted from a PAXgene blood sample showed exon skipping or intronic retention resulting in a premature termination codon. This allowed use of the PS3 criterion to achieve a classification of likely pathogenic.

In one trio with a homozygous novel \textit{CLPB} variant, p.(Arg362Gln), we applied PP4 as the testing strategy used a gene agnostic approach and the patient had previously been tested for the two other genes associated with the phenotype. For couples with pregnancies affected by lethal disorders, where fetal DNA is limited we undertook a parental exome analysis to identify rare variants (MAF<0.0001) in the same gene. In this scenario we have used co-segregation data (PP1) from testing the fetal DNA sample(s). In several families additional information obtained from the international clinical genomics community allowed classification of variants as likely pathogenic. For example we identified a \textit{POMGNT1} variant reported on ClinVar as uncertain significance. Contacting the clinician confirmed both cases were compatible with the diagnosis, enabling a genetic diagnosis for both families.

The ACMG guidelines provide a useful framework for classifying very rare variants, through the provision of defined levels of evidence to establish pathogenicity. These cases illustrate...
how applying these guidelines within an NHS multidisciplinary service have resulted in changes in practice that have informed and improved the clinical management of the patient and their family.
3. Focused Exomes: A Hypothesis-Driven Genomic Test With a Safety Net

EG Bend, RJ Louie, RC Bend, S McGee, K Kubiak, JA Lee, MJ Friez, JR Jones

Greenwood Genetic Center, Greenwood, SC

Our diagnostic laboratory has begun offering a novel type of Next Generation Sequencing (NGS) assay using customized genomic testing catered to the patient’s specific clinical diagnosis. The test is called Focused Exome Sequencing (FES). FES samples are processed using our standard whole exome sequencing (WES) bench pipeline, but only 1 to 20 physician-requested genes are analyzed. This targeted approach offers key benefits over WES and fixed NGS panels for a subset of patients. First, customized gene lists can target a patient’s unique phenotype and accommodate our evolving understanding of the genetic causes of disease. This system empowers clinicians to practice hypothesis-based testing, and to their credit, the diagnostic yield of this test is high. We reported a pathogenic or likely pathogenic variant in over 50% of FES cases (n=36) since this test was launched. This is significantly higher than the yield for WES (36%, n=413) or that of our high-yield panels (e.g., Skeletal Dysplasia, 20%, n=131). Second, FES can potentially target genes for which clinical Sanger sequencing does not currently exist, increasing the diagnostic options for patients. Third, data analysis and reporting are more straightforward in comparison to both WES and panels. This has the potential of easing the burden on the laboratory, the clinicians, and the patients, since no variants of uncertain significance in unrelated genes will be reported. Finally, in the event of a negative result by FES, tests can easily be reflexed to WES analysis with no additional bench work. This provides an important diagnostic safety net for clinicians and patients. We will outline our FES methodology and results, and highlight interesting cases solved by FES. In an age when global hypothesis-free approaches to molecular diagnosis are popular (i.e., whole exome/genome sequencing), we think that genetics laboratories should not lose sight of the diagnostic strength of the clinician.
The Illumina Clinical Services Laboratory (ICSL) offers the TruGenome Predisposition Screen, a physician-ordered clinical whole genome sequencing screen intended for generally healthy adults. This screen provides a clinical interpretation of variants found in approximately 1,700 genes associated with more than 1,200 inherited disorders. Only genes with a well-established gene-disease association and variant callability of greater than 95% across the gene are included.

Variant interpretation is performed using ICSL’s variant classification system, which is based on the American College of Medical Genetics and Genomics (ACMG) guidelines and considers variant allele frequency, disease penetrance and prevalence, and inheritance mode, in addition to data reported in the literature. Variants too common to cause the associated disorder in a Mendelian manner are ruled out utilizing an autocategorization algorithm. ICSL also uses an additional classification category, variant of unknown significance-suspicious (VUS-S), for variants with insufficient information to classify as likely pathogenic in the context of a reportedly asymptomatic individual but with some evidence suggesting the variant might contribute to a disorder.

The ICSL database currently contains over 166,000 variants curated from nearly 1,600 genomes (2012 – present). A recuration program of all variants was initiated in June 2016 to ensure up-to-date classifications. In the interest of sharing these data with the wider community, ICSL has made two donations to ClinVar to date. The first donation contained more than 95,000 unique variants with interpretations and included those reviewed in the initial database update effort. The second donation marked the beginning of regular planned donations and included more than 11,000 unique variants with interpretations. Each donated variant was classified for all associated disorders and evidence summaries were provided for all variants with clinically significant and VUS-S classifications. The more recent submission also included evidence summaries for autocategorized variants. Subsequent donations will include evidence summaries for all variants and ICSL is actively working with ClinVar to resolve variant interpretations identified by ClinVar users as having conflicts or errors. These efforts highlight the power of whole genome sequencing in healthy individuals.
5. An informatics infrastructure for KCNQ2 associated disorders research including patient registry, database, curation platform, and website

Nishtha Joshi1, Maurizio Taglialatela2, Sarah Weckhuysen3, Gerry Nesbitt4, Edward C. Cooper1

1Department of Neurology and Departments of Neuroscience and Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
2Department of Medicine and Health Science, University of Molise, Campobasso, Italy
3Neurogenetics Group, Department of Molecular Genetics, VIB, Antwerp, Belgium
4Clirinx, Inc., Dublin, Ireland

KCNQ2/3 variants lead to a spectrum of early onset neurological disease, from self-limiting forms (BFNE, OMIM #121200) to KCNQ2/3 encephalopathy (EIEE7, #613720), which includes phenotypes with persistent seizures, profound developmental impairment, and autism. Factors affecting outcome are incompletely understood. Objectives of KCNQ2/3 variant curation include improving understanding of genotype-phenotype relationships, enabling clinical research developing better outcome measures, and identifying subgroups that may require different therapeutic approaches.

A custom informatics system has been established using Clirinx software incorporating a patient registry, curation platform and website. Data collected include clinical history, test results, and therapeutic responses, as available. Each entry is reviewed by a multi-institutional, multidisciplinary curation panel. Curation includes standardized abstraction of key clinical and laboratory data that ends in a “variant summary” used for classification. Variant scoring criteria for pathogenicity and severity assessment are based on ACMG guidelines (Richards, 2015), but are customized based on gene-specific knowledge. A novel feature, assessing variant severity and thus, actionability, has been implemented. A “point and click” online pathogenicity/severity calculator is used. Results, including parts of the evidence summary and the pathogenicity and severity classification, are reported on a locus specific website (www.rikee.org) and (in process) ClinVar.

Patients and pedigrees total 483 as of April 2017. 77 BFNE variants (88 pedigrees) and 50 EE variants (114 individuals) have been classified as likely pathogenic or pathogenic. Some previously published variants have been classified as benign, likely benign, or uncertain significance. Nearly all reported but as-yet uncurated patients have EE and de novo KCNQ2 variants. Correlations have been made between phenotype and variant characteristics, including type (e.g., missense vs. stop-gained), location in the quaternary structure, and in vitro functional effects. A growing percentage of reported cases are recurrences previously seen in an unrelated patient.

The KCNQ2/3 patient registry, database, curation platform, and public website provide an integrated informatics home for future research. Our work provides a model for addressing challenges in collaborative investigation of rare disease arising in genes with great phenotypic heterogeneity.
6. Evaluation of Qiagen Clinical Insight as a content resource for variant curation in a CLIA laboratory

Samuel Cox¹, Kambiz Karimi¹, Krista Moyer¹, Peter Kang¹, Eric Evans¹, Sara Tanenbaum², Dan Richards²

¹. Counsyl, 180 Kimball Way, South San Francisco, CA 94080
². QIAGEN, 1700 Seaport Blvd, Redwood City, CA 94063

Counsyl is a health technology company that offers an expanded carrier screen for >175 of the most relevant recessive diseases and a panel of up to 36 genes for hereditary cancer risk assessment. Our CLIA laboratory utilizes NGS to analyze all exons for panel genes, excluding some regions with high homology, and we have developed a comprehensive curation workflow to interpret the clinical impact of detected variants.

Automated proprietary software initially classifies variants with high population frequency or which have not been previously reported. The remaining variants undergo curation by PhD scientists and genetic counselors before approval by board-certified laboratory directors. Pathogenicity is assessed using ACMG guidelines and is based on published case and functional studies, variant databases, population frequency, conservation, and in silico predictors. Identification of articles mentioning a variant is crucial for accurate appraisal of pathogenicity, and Counsyl relies on variant databases, an in-house article library, and online searches. We investigated QIAGEN Clinical Insight (QCI™) as a solution for reducing manual searches, with quantitative and qualitative assessment of variant-specific coverage.

QIAGEN Clinical Insight (QCI) is a clinical decision support platform that provides manually curated clinical case evidence with computed ACMG classifications and a comprehensive bibliography of articles. Articles describing variants in the relevant genes are identified through natural language processing of abstracts and PubMed annotations. The full text of the articles is reviewed by scientists that have undergone a rigorous training program and is entered into a web-based curation tool. QIAGEN uses third party User Acceptance Testing to validate high-level coverage and accuracy in compliance with quality targets.

For Counsyl-curated variants, we first evaluated reference overlap with QIAGEN, and whether additional references identified by Qiagen would alter classifications; details will be presented. The analysis demonstrates the benefits of adopting QCI for reference selection, while validating the efficacy of our previous workflow. With initial integration, QCI has proven a valuable resource for increasing the efficiency of our in-house curation, as evidenced by ~75% time savings in a reference search process that can take up to 45 minutes. More time-intensive methods can now be focused on variants nearer evidence thresholds.
Establishing Clinical Validity Classifications for the RASopathies

Brandon Cushman¹ and Lisa Vincent² on behalf of the ClinGen RASopathies Expert Panel

¹Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, Massachusetts, USA
²GeneDx, Inc., Gaithersburg, Maryland, USA

Funding: NIH U41 HG006834

The RASopathies are a group of syndromic disorders caused by dysregulation in the RAS/mitogen-activated protein kinase (MAPK) pathway. The disorders include Noonan syndrome (NS), cardiofaciocutaneous syndrome (CFC), Costello syndrome (CS), Noonan syndrome with multiple lentigines (NSML), Noonan-like syndrome, neurofibromatosis type 1 (NF1), and Legius syndrome. RASopathies are estimated to affect 1:1000 newborns¹. While phenotypically heterogeneous, the disorders share some overlap in clinical presentation because of their common pathogenic mechanism resulting in dysfunction of the RAS/MAPK signaling pathway. Presentations include craniofacial dysmorphism, developmental delay, and generalized hypotonia, as well as musculoskeletal, ocular, and cutaneous abnormalities². Each disorder is differentiated based on both its clinical presentation and genetic etiology³.

Established in 2015, the ClinGen RASopathies Expert Panel (RAS EP) strives to establish the clinical validity of the gene-disease associations within the RASopathies and to refine the ACMG classification criteria⁴ for the interpretation of variants in those genes.

Using the semi-quantitative method for validating gene-disease associations established by the ClinGen Gene Curation Working Group⁵, the RAS EP preliminarily assessed 11 genes and their association with at least 1 of 5 RASopathies for a total of 25 gene-disease pairs. Of these, 9 genes were assessed for NS, 5 for CFC, 4 for CS, 5 for NSML and 1 for Noonan-like syndrome. As new gene-disease associations are published, the RAS EP aims to curate and assess their association with this group of disorders.

References
Multi-gene panel tests (MGPT) have grown in use with advances in sequencing technology; however, to responsibly design such panels, a comprehensive vetting process using a standardized clinical validity scoring system is necessary. Proper gene vetting aims to provide relevant information to identify the most clinically significant genes and thereby allow accurate and useful interpretation of results. To directly test the utility of clinical validity-based gene vetting, the results from hereditary cardiovascular MGPT from 3,524 cases including testing of a subset of 106 genes from a clinical laboratory were retrospectively reviewed. The detection rates of pathogenic and likely pathogenic variants (VLP) in clinically characterized genes (i.e. with moderate, strong, or definitive evidence for a role in disease) were compared to those with limited evidence of clinical validity. Variants were classified based on the clinical laboratory's classification scheme. Across all panels combined, 575 mutations or VLPs and 2,454 variants of uncertain significance (VUS) were identified. Using a newly published standardized scoring system, 42% of the genes on the panel were determined to be definitive while 17% were strong, 28% were moderate, and 13% had limited clinical validity. When these data were analyzed based on each clinical validity category, we identified approximately 12 mutations/VLPs per gene for genes in the definitive category, 1.7 mutations/VLPs per gene in the strong category, 0.5 mutations/VLPs per gene in the moderate category and 0.07 mutations/VLPs per gene in limited category. Among the 14 genes determined to have limited clinical validity, only one VLP was identified in the TGFB3 gene (amongst 2,122 total cases tested for the gene). These results show the importance of evidence-based, standardized gene vetting process to establish clinical validity when selecting gene content for MGPT. While there may be many reasons to consider gene inclusion on an MGPT, detection rates support prioritization of genes with greater clinical validity scores for maximum clinical actionability and to reduce VUS burden.
9. Curating clinically relevant transcripts for the interpretation of sequence variants

Marina T. DiStefano¹, Sarah E. Hemphill¹, Brandon J. Cushman¹, Andrea M. Oza¹, and Heidi L. Rehm¹²

¹Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA.
²Department of Pathology, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA USA

Proper analysis of genomic variants is critical for patient care. The ACMG has set forth guidelines for the interpretation of sequence variants (Richards et al. 2015, PMID 25741868); however, these guidelines depend on defining the biologically relevant transcripts necessary to accurately interpret the impact of variation on gene function. We have therefore developed a systematic strategy for designating primary and other clinically relevant transcripts and applied it to genetic hearing loss, a highly heterogeneous disease with over 100 genes causative for nonsyndromic hearing loss alone (Abou Tayoun et al 2016 PMID: 26562227).

Genes were divided into 3 categories using NCBI reference sequence (RefSeq) transcripts with Ensembl transcripts used for comparison. Category 1 (C1) contained genes that had a single transcript. While genes in category 2 (C2) had multiple transcripts, the longest transcript encompassed all exons, and these genes required minimal curation. Category 3 (C3) genes had multiple transcripts with unique exons and required thorough curation.114 hearing loss-associated genes, predominantly from the OtoGenome™ Test (GTR000509148.7), were categorized and 37 C1, 41 C2, and 35 C3 genes were identified. The 35 C3 genes were manually curated to identify a medically relevant transcript set for variant interpretation. All C3 transcripts were curared with respect to function, tissue specificity, and temporal expression. All pathogenic and likely pathogenic variants in each C3 gene were pulled from ClinVar, HGMD (DM variants), and internal clinical testing data and evaluated based on transcript location and predicted molecular consequence to the gene. Multiple genes contained variants that could be incorrectly classified without a curated transcript list. For example, a pathogenic variant in CLRN1 is a missense variant in the RefSeq transcript used by ClinVar, HGMD, and our lab (NM_174878.2:c.368C>A (p.Ala123Asp)), but is a nonsense variant in an alternative transcript (NM_001256819.1:c.540C>A (p.Cys180X)). In another example, the gene OTOF contains two alternate last exons, however only one of these is expressed in the human ear (Choi et al. 2009 PMID:19250381).Given the critical role transcript analysis plays in variant interpretation, we recommend each ClinGen Expert Panel or Gene Curation Team perform this analysis in their disease areas. Transcript curation such as this can greatly improve the quality of variant interpretation and patient care.
Molecular diagnostic laboratories find compliance with evolving guidelines a challenge in an environment of multiple regulatory and guideline agencies due to locale specific practices. In 2016, ClinGen Minimum Variant Level Data (MVLD) was released followed closely by AMP/CAP/ASCO guidelines for somatic variants reporting in early 2017. To maintain consistency, established labs are additionally burdened by custom philosophies established through legacy reports. The need for a resource to eliminate repetitive manual curation of the same evidence prompted development of Molecular Assertions (MA) which leverages decoupled tiering templates for generation of custom philosophies. The underlying data model is based on evidence attributes defined by MVLD and AMP/CAP/ASCO guidelines. Drawing upon our previous advances in publications, clinical trials and associative therapeutics search, MA establishes concrete linkage between a given condition, molecular alteration, clinical significance (prognostic, diagnostic, predictive), therapeutic if predictive, and citation or source. Source types include regulatory, clinical trial, in human case study, expert opinion, retrospective institutional study, preclinical or pathway inferred. Subtypes like prospective, retrospective, or meta-analysis further clarify source type clinical trials while cell line, PDx, biochemical assay, or mouse model pertain to preclinical sources. Source evidence without Pubmed IDs are cited with URL links indexed by unique ID. Distinctions between well powered versus small studies are based on phase of trials. Currently 1488 MA spanning lung, breast and colorectal cancers have been curated with the addition of another 2200 lung MA expected within 3 weeks time. Tiering templates applied to date include MVLD, AMP/CAP/ASCO, and customer specific 8 tier templates. Examples of challenges include vagueness in AMP/CAP guidelines on defining classification of variants with high prevalence but no functional difference from wild type, determining applicability of historically conditions based FDA approvals to individual variants or genes, and definition of off label scenarios. Applications of MA are broad, and adoption has been rapid for a variety of steps in a diagnostic laboratory’s workflow including assay panel design, reporting consistency, physician portals for clinical decision support, human readable assertions, and rationale for assigned tiers.

*Current affiliation GeneByGene

Dwight, S.\textsuperscript{1}, Wright, M.\textsuperscript{1}, Dalton, K.\textsuperscript{1}, Choi, M.\textsuperscript{1}, Zhen, J.\textsuperscript{1}, Thomas, C.\textsuperscript{1}, Azzariti, D.\textsuperscript{2}, Babb, L.\textsuperscript{2}, Bizon, C.\textsuperscript{3}, Buchanan, A.\textsuperscript{5}, DiStefano, M.\textsuperscript{2}, Ghosh, R.\textsuperscript{4}, Harrison, S.\textsuperscript{2}, Patel, R.\textsuperscript{4}, Pawliczek, P.\textsuperscript{4}, Rooney Riggs, E.\textsuperscript{5}, Seifert, B.A.\textsuperscript{6}, Strande, N.\textsuperscript{5}, Berg, J.\textsuperscript{6}, Bustamante, C.\textsuperscript{1}, Martin, C.\textsuperscript{5}, Milosavljevic, A.\textsuperscript{4}, Plon, S.\textsuperscript{4}, Rehm, H.\textsuperscript{2}, Cherry, J.M.\textsuperscript{1}

\textsuperscript{1}Stanford University
\textsuperscript{2}Laboratory for Molecular Medicine, Partners Health Care
\textsuperscript{3}Renaissance Computing Institute
\textsuperscript{4}Baylor College of Medicine
\textsuperscript{5}Geisinger Health System
\textsuperscript{6}University of North Carolina at Chapel Hill

Critical to the field of clinical genomics is an understanding of the strength of correlation between genotype and phenotype. The central mission of ClinGen (The Clinical Genome Resource, clinicalgenome.org) is to define the “clinical relevance of genes and variants for use in precision medicine and research.” Inherent to this mission is the ability to access and evaluate evidence for a gene or variant’s role in disease in a consistent and efficient manner. Towards this goal, ClinGen has built interconnected curation interfaces, one for gene curation and one for variant curation. The rich evidence captured by both interfaces is available in a structured format that allows it to be easily accessed via ClinGen’s public portal (clinicalgenome.org).

The ClinGen gene and variant curation interfaces have been developed according to the following important specifications: 1) variant curation follows the ACMG-AMP Standards and Guidelines, while gene curation follows ClinGen’s Clinical Validity Classification framework, 2) both interfaces centralize evidence from relevant resources to support efficient, consistent curation, 3) curated literature evidence and evidence retrieved from external resources are shared between the gene and variant curation tools, 4) the interfaces are designed to guide biocurators through the curation process, 5) use of controlled vocabularies and ontologies promote the capture of discrete evidence in a consistent manner to facilitate connections, 6) the ClinGen Allele Registry normalizes variants and provides an identifier for ClinVar submission, 7) all evidence is viewable by every user, while curated evidence can only be edited by its creator, 8) the interfaces support expert review of provisional classifications and interpretations, 9) contextual help and documentation are included to assist the biocurator, and 10) data is stored in standard JSON-LD format to define rich relationships and facilitate data exchange.

Both tools are now in production (curation.clinicalgenome.org) and currently accessible to ClinGen biocurators and approved groups from the broader community. A demo version of the interfaces (curation-test.clinicalgenome.org), which does not permanently save data, is available via both registered and generic (non-registered) access to allow exploration. We expect active use of the ClinGen curation interfaces will facilitate the implementation of the ACMG variant classification guidelines across diverse disease genes.

Alexander R. Henrie,¹ Brandon Cushman,² Sarah Hemphill,² Steven Harrison,² Marina DiStefano,² Danielle Azzariti,² Justin Aronson,³ Lisa M. Vincent,⁴ Scott R. Goehringer,⁵ Erin Rooney Riggs,⁵ Christa Lese Martin,⁵ Heidi L. Rehm,²,⁶ Karen Eilbeck¹

¹Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, Utah, USA
²Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA.
³Pierce Middle School, Brookline, MA, USA
⁴GeneDx, Gaithersburg, MD, USA
⁵Geisinger Health System, Danville, PA, USA
⁶Department of Pathology, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA USA

ClinVar is an international archive of variant-condition interpretations hosted by the National Center for Biotechnology Information (NCBI). Submitters include clinical labs, research institutions, database curators and consortia and data are collected from clinical testing, literature curation and smaller databases. Viewing and understanding the complex data within the database at a higher perspective, rather than variant perspective, facilitates both archive development and maintenance while also providing significant feedback to contributors on conflicts in interpretation with other submitters. Thus, the exploration of this data leads to revised and improved understanding of the data in ClinVar.

We have developed an interface for viewing data collected in ClinVar, called ClinVar Miner (http://malachite.genetics.utah.edu/clinvar-miner/), to explore the database at different levels of granularity and from different perspectives. Statistics for current data and in some cases for time series can be viewed relative to all submissions, submitters, conflicting submissions and genes. As an example, conflicting interpretations by submitters can be extracted and triaged easily by utilizing the ‘Conflicts by submitter’ tool, which enables a submitter to view all other submitters with whom they have conflicting interpretations. These submissions can further be filtered by review status and the method of submission. The differing interpretations between each submitter are separated into five levels of conflict status: synonymous conflict, confidence conflict, benign vs uncertain conflict, category conflict, and medically significant conflict. Filtering enables the user to prioritize conflict resolution efforts. The conflicting variants can be further examined by displaying the evidence in ClinVar for each interpretation.

In order to identify trends over time, submitters can view high-level summaries using the ‘Total submissions by method’, which displays the growth of ClinVar submissions over time, and across submission methods. Like the tool for individual submissions, this graph can be filtered by review status and method of submission. Another tool, ‘Total submissions by country’, displays a log-scale choropleth with high-submitting countries colored more darkly than low-submitting countries. These views can be used to inform policy decisions regarding submission and display of data in ClinVar. We will present the tool and demonstrate both the usage and findings.
13. Consent and Return of Medically Actionable Genomic Results in the Geisinger MyCode Community Health Initiative Biobank

W. Andrew Faucett1, Carroll Flansburg2, Adam H. Buchanan2, Marci L.B. Schwartz2, John Jurosky2, Michael Murray2, Amanda Lazzeri2, David H. Ledbetter1

1. OCSO, Geisinger Health System, Danville, PA, United States.
2. Geisinger Health System, Danville, PA, United States.

BACKGROUND: When the MyCode Community Health Initiative began in 2007 it did not include return of results to participants. Focus groups with diverse stakeholders were held and the consent was amended in 2013 to allow return of medically actionable results with placement in the electronic medical record. A communication plan for patients and providers was designed in conjunction with a clinical oversight committee, an ethics advisory council, and our IRB. We present a model for returning results in large research studies, and patients’ decisions for medical follow-up.

METHODS: We tracked the actions of patients after notification. When a result is identified, providers are notified via the electronic health record or phone. About 5 business days later, patients are notified by mail or patient portal that a result is available. This contact has four key aspects: (1) a thank you for participating in research, (2) a reminder of the research purpose, (3) identification of an actionable genomic result, and (4) follow-up recommendations. Three additional contact attempts are made. If necessary, a fifth contact is made in the form of a certified letter notifying the patient that they have a result. A final notification is sent to the physician of record stating that attempts to contact the patient were unsuccessful and we remain available for information.

RESULTS: 309 patients were identified from May 2015 to April 2017 with results. 141 patients were seen by genetics, 7 by a specialist, 36 by their primary care provider, 90 chose not to follow-up immediately, 4 are now deceased, 10 are pending additional notification attempts, and the remainder are pending further discussion. Early assessments have shown participants adapting well to results and using them to guide their care.

CONCLUSIONS: A patient contact method to return genomic findings identified through research involves a complex interplay between providing accurate and appropriate information in a manner that is supportive of the patient and their physician while not causing undue anxiety. Early results have shown patient and provider acceptance of the process and that multiple contact attempts may be required. We will continue to evaluate the model’s effectiveness for engagement, results delivery, and initiation of clinical management. Data on patient preferences and the efficacy of varied approaches may inform procedures for returning actionable results to participants in large cohorts.

KEYWORDS: genomics, results, biobank
Deciphering the germline genetic contributions to hematopoietic malignancies

Lucy A. Godley, The University of Chicago

Genetic predisposition to cancer is best described for the solid tumors, but there is an emerging appreciation for the contribution for inherited mutations to the development of hematopoietic cancers. In recognition of the important contribution that germline genetic factors play in leukemia etiology, the World Health Organization included germline predisposition to myeloid malignancies in its revised classification scheme for the first time in 2016, and clinical guidelines are beginning to include these genetic syndromes. At The University of Chicago, we have established a cancer risk clinic devoted to hematopoietic malignancies and a comprehensive testing panel comprised of next-generation sequencing (NGS)-based molecular testing to detect point mutations as well as microarrays to detect chromosomal rearrangements that allow us to perform clinical testing to identify an individual/family’s germline mutation. As part of our routine diagnostic work-up for leukemia patients, we also run a 1200 gene NGS panel that covers all of the known germline predisposition genes. To determine how often this prognostic panel detected germline variants, we systematically obtained germline material for every patient in whom a variant was detected in a gene known to confer germline susceptibility. Among 28 patients, 36 variants in four genes, \textit{RUNX1}, \textit{GATA2}, \textit{ETV6}, and \textit{TP53}, were detected in the prognostic panel. Germline DNA was obtained preferentially from cultured skin fibroblasts, but we also analyzed DNA from saliva and surgical biopsy specimens when patients were in clinical remission. Among those samples, four variants in three patients were found to be of germline origin, but all were considered variants of unknown significance. This work highlights the importance of considering the potential germline nature of variants detected by panels designed for prognostic purposes as well as the difficulty in assigning functional status to these variants.
15. Criteria for OMIM for establishing gene-phenotype relationships in the era of whole genome/exome sequencing

J. S. Amberger, C. A. Bocchini, N. Sobreira, A. F. Scott, A. Hamosh
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland

For decades, OMIM has been cataloging phenotype-gene relationships. These relationships have been defined by variants in genes and have been ascertained by a variety of technologies. The certainty of these relationships is highly variable. While whole exome/genome sequencing has enabled rapid growth in the publication of "causative" variants, it still does not confirm the certainty of the relationships. Since 2013, OMIM applies the following criteria to establish a gene-phenotype relationship: (1) the existence of multiple, unrelated individuals with pathogenic variants in the same gene; (2) the variants segregate with the phenotype in multiplex families; and/or (3) the variants occur *de novo* in a statistically significant number of individuals. Functional data and/or animal models support the causality but are not required. A qualified gene-phenotype relationship is established based on the following: (1) only one multiplex family is reported to have variants in a single gene and the variants segregate with the phenotype in the family, (2) there is supportive functional data such as in vitro enzyme activity, an animal model, or (3) a robust biological pathway connection; in this case the gene-phenotype relationship is qualified by noting that the variant has been identified in only “1 family”. In rare instances, a similar gene-phenotype relationship may be established on the basis of a single patient if there is robust supporting phenotype and functional data. The morbid map listing of a qualified phenotype is preceded by a “?” . These qualified relationships are targeted for frequent review and update to further substantiate the phenotype-gene relationship. A VUS (variant of unknown significance) may be created to mark the place of a variant that has undergone significant curation, but does not yet reach the qualified state. OMIM splits phenotypes by gene and Clinical Synopses reflect the unique qualities that may be associated with specific phenotype-gene combinations. To see the genetic heterogeneity of a phenotype or compare the clinical differences among phenotypes, OMIM provides views of Phenotypic Series and clinical synopsis comparisons, respectively. To help users follow OMIM entries, we implemented MIMmatch. MIMmatch members can designate entries or Phenotypic Series to follow and receive email alerts when these are updated and/or new gene-phenotype relationships are established; or find other researchers who share their interest in certain entries.
16. Progress in evaluating the clinical validity of gene-disease associations in hearing loss

Sarah E. Hemphill¹, Brandon J. Cushman¹, Andrea M. Oza¹, Heidi L. Rehm¹,², Sami S. Amr¹,², Ahmad N. Abou Tayoun³, and Marina T. DiStefano¹, on behalf of the ClinGen Hearing Loss Clinical Domain Working Group

¹Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA
²Department of Pathology, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA, USA
³Division of Genomic Diagnostics, The Children’s Hospital of Philadelphia, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

Using ClinGen’s clinical validity classification framework (Strande 2017, in press), the ClinGen Hearing Loss (HL) Clinical Domain Working Group (CDWG) is evaluating over 200 genes with reported associations with both nonsyndromic and syndromic HL. While it is becoming more financially and technically feasible to expand the number of genes on NGS panels, it does not necessarily increase detection rates (Alfares 2015 25616165), and may actually be detrimental to both laboratories and patients (Abou Tayoun 2016 26562227). Our goal in evaluating the clinical validity of HL genes is to establish expert consensus to inform test panel design and ultimately improve patient care.

For this genetically heterogeneous disease area, we have come up with several unique approaches to curation and review that have facilitated this process. First, before curation we performed a rapid literature review in order to triage genes by importance. Additionally, we bypassed one-on-one content review for definitive genes and presented primary curations directly to the CDWG for approval. Finally, we established an expedited curation procedure to classify well-established HL genes as definitive without curating case-level data, provided no conflicting evidence has been published. These methods have allowed our curators to focus their attention on the less-documented gene-disease pairs that are critical to test panel design.

As of April 2017, 89 total gene-disease pairs had undergone primary curation, 17 of which had been finalized by the CDWG. Of 89 completed curations, 16 were limited, 24 were moderate, 6 were strong, 40 were definitive, and 3 were disputed or refuted. 14 of the 16 limited genes, including HARS, CRYM, GJB3, and SLC26A5, appeared on one or more HL panels submitted to the GTR. Variants identified in such genes are difficult to interpret and can increase inconclusive results. Another gene that is on multiple tests, MYO1A, has had its association with HL refuted (Eisenberger 2014 24616153, Patton 2017 27759032). Meanwhile, 28 of the 46 strong or definitive genes, including CABP2, CEP78, FAM65B, PDZD7, PTPRQ, S1PR2, and SLITRK6, were absent from one or more HL panel tests. The clinical sensitivity of these tests is thus decreased, as causative variants in such genes will not be detected. Our results emphasize the importance of consistent and transparent gene curation to the accurate interpretation of genomic variants and to the efficacy of genetic testing in HL.
17. The Importance of Data Curation and Sharing—FNDC3B as an Example

Nicole Hoppman, Yang Cao, Erik Thorland

Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN

Chromosomal microarrays have enabled the discovery of novel genes and genomic regions that result in abnormal phenotypes due to copy number variation (CNVs). This “genotype-first” approach, in which genotypes are determined prior to defining phenotypic correlations, contrasts with the more traditional “phenotype-first” approach in which diseases are classified based on phenotype, and then the genotypic cause is sought. Now that many recurrent microdeletions have been identified and characterized, rarer genotype/phenotype correlations remain to be discovered. This discovery process is aided by extensive data curation and systems to continually re-evaluate cases that could potentially be re-interpreted in light of new information, both within as well as among clinical laboratories performing microarray testing.

Our laboratory previously identified two patients with abnormal craniofacial features and overlapping de novo 3q26.31 deletions (513 and 345 kilobases in size). These deletions met our laboratory reporting size criteria (>200 kilobases for deletions) and were therefore initially reported as CNVs of uncertain clinical significance since loss-of-function mutations in the genes within the interval had not been reported previously in the literature. The minimal region of overlap between these deletions contained only the FNDC3B gene. Since our laboratory maintains a database of all CNV calls encountered in approximately 10 years of chromosomal microarray testing, we were able to recognize that these two de novo deletions had occurred in patients with similar phenotypic features and we were able to publish these cases (Cao et al., Am J Med Genet A 170(12):3276-3281, 2016).

Subsequent to publication of this manuscript, we observed a third (84 kilobase) deletion within the FNDC3B gene in an infant with multiple findings, including abnormal craniofacial features. If it were not for publication of the first two individuals, this deletion would not have been reported as it was below the size cut-off in our reporting guidelines. Instead, the laboratory was able to issue a “likely pathogenic” report to the patient, explaining at least some of the infant’s abnormal phenotypic features. Furthermore, there is at least one additional patient per the ClinGen database with an uncertain deletion including FNDC3B whose variant could be re-classified in light of these data. These cases highlight the importance of data curation, mining, sharing, and publication.
18. Gene curation for Brugada syndrome questions the clinical validity of some of the previously reported gene-disease associations

S Mohsen Hosseini1, Raymond H Kim1,2,3, Rebekah Jobling1,3,4, Eriskay Liston1,3, Seema M Jamal3,4, Gregory Costain3, Marta Szybowska2, Chantal F Morel2, Sarah Bowdin1,3, Michael H Gollob5

1 Ted Rogers Centre for Heart Research, The Hospital for Sick Children, Toronto, ON, Canada
2 Fred A. Litwin Family Center in Genetic Medicine, University Health Network, Toronto, ON, Canada
3 Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada
4 Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada
5 Peter Munk Cardiac Centre, Toronto General Hospital, Toronto, ON, Canada

Introduction: Brugada syndrome (BrS) is an arrhythmia syndrome with a risk of sudden cardiac death, diagnosed based on a characteristic electrocardiogram pattern. When familial, it follows an autosomal dominant inheritance with high but incomplete penetrance. BrS is genetically heterogeneous; more than 20 genes have been reported to cause BrS and are available on commercial genetic panels. The validity of these disease-gene associations is critical to the accurate interpretation and application of genetic information in the care of patients and families. We sought to evaluate the clinical validity of the reported gene associations for BrS using an analytical framework proposed by the Gene Curation Working Group of Clinical Genome Resource (ClinGen).

Methods: Our group included 3 independent gene curation teams (3 curators per team) consisting of genetic counsellors, genome scientists, and led by a board-certified medical geneticist. We followed the proposed ClinGen Gene Curation framework which provides an evidence-based system for evaluating gene-disease associations. Based on a comprehensive review of the published literature, each gene was evaluated for genetic and experimental evidence supporting its causal role for BrS. Using a semi-quantitative scoring metric weighing the level of evidence, each curation team independently classified genes as demonstrating limited, moderate, strong or definitive evidence for disease association.

Results: Here, we report our initial results. Six genes have been curated independently by all 3 curation teams. Of these genes, only SCN5A reached the definitive evidence tier. For ABCC9, RANGRF and CACNA2D1 all 3 curation teams concluded limited evidence. For the GPD1L gene, curation teams were concordant for a moderate level evidence score. For the ANK2 gene, the three teams did not agree on evidence tier (two moderate and one limited). Overall, there was a strong agreement between the three curation teams in classifying genes into evidence tiers. Furthermore, we observed a strong correlation between the quantitative evidence scores for each gene among different groups (ICC=0.924, p = 0.00041) supporting reliability of the method.

Conclusions: These findings question the validity of some genes reported to be associated with BrS, implying that the practical utility of including such genes on clinical genetic testing panels may need to be revised.
Familial hypercholesterolemia (FH) is among the most common monogenic disorders encountered in clinical practice, and is characterized by extreme LDL cholesterol levels and premature cardiovascular disease. In recent times FH has moved toward the forefront of precision medicine, as increasing numbers of patients worldwide are routinely offered genetic testing as a central part of diagnosis. Particularly, targeted next-generation sequencing panels have proven to be efficient platforms for facilitating this effort. However, determining the pathogenicity of detected DNA variants remains a major challenge; this process may be rudimentary and often will differ from one diagnostic laboratory to the next, leading to inconsistencies in variant classification. To improve accuracy, concordance and standardization across laboratories, there has been an increasing effort to implement the recent variant interpretation guidelines proposed by the American College of Medical Genetics (ACMG). Preliminary analyses using these criteria suggest that up to 40% of more than 1900 putative FH mutations in the \textit{LDLR} gene deposited over the past 20 years can be classified by today's standards as "variants of unknown significance". Thus to even further optimize curation and better classify such variants, the ClinGen FH Expert panel has discussed a number of appropriate modifications to existing ACMG criteria. Here we propose a consensus set of ACMG-adapted guidelines specific for FH variant curation that include 1) explicit statements differentiating between \textit{LDLR}, \textit{APOB}, and \textit{PCSK9} genes, 2) an emphasis on functional study evidence types, 3) alteration of population data frequency thresholds, 4) appropriate use and interpretation of computational tools, and 5) addition of criteria for large-scale copy number variants. Furthermore, expert opinion and consensus has further helped generate several sub-classifications and criterion weighting changes. We anticipate that the incorporation of such evidence-based reasoning will enable more accurate assessment of sequence variants in FH research and clinical care.
20. Optimization of the ACMG-AMP Criteria for CDH1 Variant Classification

Rachid Karam¹*, Intan Schrader ²*, Kristy Lee³#, Carla Oliveira⁴, Maegan Roberts ⁵, Keith Nykamp⁶, Tina Pesaran⁷, Rajarshi Ghosh⁷, Joana Figueiredo⁴, Yuya Kobayashi⁸, Pardeep Kaurah⁷, Kate Krempey⁷, Thomas Slavin⁶, Liying Zhang⁹, Elizabeth Chao¹⁰, James Ford¹¹, Amanda Spurdle¹², Sean Tavtigian¹³, Fatima Carneiro⁴, David Huntsman², Sharon Plon⁷.

¹ Ambry Genetics, Aliso Viejo, CA, USA
² British Columbia Cancer Agency, Vancouver, Canada
³ UNC Chapel Hill, NC, USA
⁴ University of Porto, Porto, Portugal
⁵ GeneDx Laboratories, Gaithersburg, MD, USA
⁶ Invitae, CA, USA
⁷ Baylor College of Medicine, Houston, TX, USA
⁸ City of Hope, Los Angeles, CA, USA
⁹ Memorial Sloan Kettering Cancer Center, New York, NY, USA
¹⁰ University of California, Irvine, CA, USA
¹¹ Stanford University School of Medicine, Stanford, CA, USA
¹² QIMR Berghofer Medical Research Institute, Australia
¹³ Huntsman Cancer Institute, University of Utah School of Medicine, UT, USA
*Chairs. #Coordinator.

CDH1 mutations cause Hereditary Diffuse Gastric & Lobular Breast Cancer (HDGC – OMIM 137215). Heterozygotes for a germline CDH1 truncating pathogenic variant present an estimated cumulative risk of developing gastric cancer 70% (95% CI, 59%-80%) for males and 56% (95% CI, 44%-69%) for females. Women also have a 42% (95% CI, 23%-68%) risk of developing lobular breast cancer. CDH1 is a clinically actionable gene. Current recommendations include, for women, consideration of breast MRI starting at 30 years of age and risk-reducing mastectomy. Prophylactic gastrectomy is recommended for both genders between the age of 20 and 30 (annual upper endoscopy following the dedicated Cambridge protocol with a minimum of 30 biopsies for people who do not undergo gastrectomy). Therefore, it is imperative to detect and correctly classify variants in CDH1 according to their pathogenicity. With this in mind, the ClinGen Hereditary Cancer Working Group selected CDH1, together with PTEN and TP53, as high priorities genes and created a CDH1 working group (WG). The CDH1 WG is an expert panel assembled from CDH1 experts encompassing clinicians, scientists and clinical laboratory diagnosticians, with a primary goal to optimize the 2015 ACMG/AMP Variant Interpretation Guidelines specific to CDH1.

We will present the CDH1 WG rule specifications to the ACMG/AMP criteria, which after systematic evaluation with a series of test variants will be finalized and incorporated into the group’s curation process, as part of the CDH1 ClinGen Expert Panel application. The group recommendations for pathogenic and benign criteria optimizations for CDH1 include: standalone and strong benign CDH1 specific allele frequency cutoffs; gene specifications for splicing; recommendations for the use of computational and functional evidence; standardization of acceptable HDGC diagnostic criteria and minimal clinical information. These criteria will be validated by comparing clinical, population, computational, and functional data, among multiple institutions participating in the CDH1 WG. Validated gene-specific optimizations of the ACMG/AMP 2015 guidelines will be described.
21. Analysis of Germline Variants of Uncertain Significance from Paired Tumor-Germline Samples in a Pediatric Genomic Sequencing Research Study

Chimene Kesserwan*1, Dale Hedges*2, Scott Newman2, Kayla Hamilton1, Rose McGee1, Emily Quinn1, Regina Nuccio1, Jessica Valdez1, Scott Foy2, Joy Nakitandwe2, Lynn Harrison1, Annastasia Ouma1, Stacy Hines-Dowell1, Sheila Shurtleff3, Elizabeth Azzato3, David Ellison3, James Downing3, Jinghui Zhang3, Kim E. Nichols1

1Department of Oncology, Division of Cancer Predisposition, St. Jude Children’s Research Hospital, Memphis, TN
2Department of Computational Biology, St. Jude Children’s Research Hospital, Memphis, TN
3Department of Pathology, St. Jude Children’s Research Hospital, Memphis, TN

*equal contribution

Expanding implementation of germline genomic sequencing is increasing the number of Variants of Uncertain Significance (VUS), posing challenges for providers and patients.

Genomes for Kids is a research study addressing the feasibility of integrated clinical genomic analysis using whole genome (WGS), whole exome (WES) and transcriptome sequencing (RNAseq) of tumor samples and WGS and WES of germline samples from pediatric oncology patients. Here, we sought to determine the frequency of germline VUS and examine for concordance with ClinVar submissions. Variant interpretation was performed following ACMG guidelines.

Germline samples from 248 patients were analyzed for mutations in 63 cancer predisposition genes. We identified 152 VUS in 116 patients (47%). No ClinVar entry was available for 62 (41%) VUS. The remaining 90 variants were submitted as: benign/likely benign (B/LB; n = 6), conflicting interpretation of pathogenicity (CIP; n = 26) and VUS (n = 58). Each variant with CIP received 2-9 submissions (mean 4.5/variant). All CIP were called a VUS at least once; to this end we determined our VUS calls to be concordant with ClinVar in 84/90 (93%) of cases. The remaining ClinVar calls were listed as B/LB, except for 2 that were LP: TP53 (R161T) and CDKN2A (I49T). We classified these as VUS due to insufficient available evidence to justify an LP call. We utilized tumor sequence data to guide interpretation of 2 VUS identified in our study: BAP1 (c.256-3C>A) and APC (K150R). For the BAP1 variant, an ultimate classification of LP was established after the variant was demonstrated to negatively impact splicing in the tumor. Conversely, for the APC variant, an external laboratory’s prior classification of LP, based on splice prediction, was called into question by direct assessment of somatic tissue. RNAseq data revealed intact splicing, supporting our interpretation of this variant as a VUS.

To enhance interpretation of germline genomic variants, it is imperative that: 1) variants submitted to ClinVar be systematically curated by an expert panel and 2) investigators utilize somatic data, when available, to provide information about potential downstream effects. Current germline variant interpretation guidelines should be adapted to integrate somatic data. Although our case examples are few, the distinction between receiving a germline report of VUS versus LP cannot be understated given the differences in clinical implications between these results.
ClinVar Supports Multiple Approaches to Access to Variant Interpretations for the ACMG 59 Genes


ClinVar is the public archive at NCBI for interpretations of genetic variants relative to a disease or phenotype. The database includes >452,000 interpretations for >309,000 variants. Interpretations in ClinVar have been provided by nearly 700 submitters, including clinical testing laboratories, research laboratories, OMIM and GeneReviews, and expert panels. The American College of Medical Genetics and Genomics has published recommendations for the reporting of incidental findings in the exons of certain genes. The most recent recommendation notes 59 genes for which incidental findings in clinical testing should be reported (PMID 27854360). ClinVar maintains a page summarizing the ACMG 59 genes along with links to variants in these genes that have been reported to ClinVar as pathogenic or likely pathogenic (https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/). ClinVar has 14,396 variants in the ACMG 59 genes with aggregate interpretations of pathogenic or likely pathogenic. The top ten ACMG genes for number of pathogenic or likely pathogenic variants in ClinVar are BRCA2, BRCA1, LDLR, MLH1, MSH2, FBN1, MYBPC3, COL3A1, MYH7, and APC. Seven ACMG genes are in the overall top ten in ClinVar for number of pathogenic/likely pathogenic variants: BRCA2, BRCA1, APC, TSC2, LDLR, MSH2, and MSH6. ClinVar flags variants in the ACMG 59 genes on the website and in downloadable files. Web searches can be limited to variants in the ACMG 59 genes with an indexed property. Variants in these genes are flagged in ClinVar's XML by a comment on the MeasureRelationship element, and the tab-delimited report variant_summary.txt has a flag for variants in the ACMG genes. Information about variants observed as secondary findings may be included in submissions to ClinVar as part of the evidence for the variant interpretation. The field “Secondary finding” may be used to indicate when the submitted variant was found in a particular individual as an incidental finding. The submitter indicates whether the individual with the variant was affected with the interpreted condition or not; the individual’s age at testing may also be provided if relevant to the age of onset for the disease. Additional details may be included as a free text comment. For example, the submitter may indicate that a variant was detected by genomic screening and while the individual is asymptomatic for the interpreted condition, several family members have features of the disease.
23. Validity Assessment of Genes Commonly Found on Clinically Available Hereditary Breast and Ovarian Cancer Susceptibility Sequencing Panels

Kristy Lee†, Bryce A. Seifert†, Hermela Shimelis‡, Rajarshi Ghosh†, Natalie J. Carter‡, Stephanie B. Crowley†, Kurston Doonanco‡, Sharisse Jimenez†, Jonathan S. Berg†, Sharon E. Plon§, Fergus J. Couch¶

1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Department of Pediatrics-Oncology, Baylor College of Medicine, Houston, TX.; 4) GeneDx Laboratory, Gaithersburg, MD

The National Comprehensive Cancer Network recommends multi-gene test panels when a patient’s personal and/or family history is suggestive of more than one hereditary cancer syndrome. The guidelines warn clinicians that many of the genes on hereditary breast cancer (BC) and ovarian cancer (OC) susceptibility panels have not been systematically examined for their strength of association with disease and do not routinely have medical management recommendations. Also, it can be difficult to know if a particular gene is associated with BC, OC or both. To address these questions, 42 genes commonly offered on these test panels were selected for evaluation using the Clinical Genome Resource (ClinGen) Clinical Validity Framework.

The framework was used to assess the strength of evidence between a gene and BC or OC in order to assign one of the following classifications: Definitive, Strong, Moderate, Limited, No Reported Evidence, Conflicting Evidence or Refuted Evidence. Evidence evaluated included: case-control reports illustrating clinical significance, time elapsed since the first clinical publication and functional data implicating the gene’s association with disease.

Of 16 genes evaluated for BC susceptibility to date, 4 genes had “Definitive”, 0 had “Strong”, 1 had “Moderate”, and 5 genes had “Limited” evidence. Conversely, 3 had “No Clinical Evidence”, 2 had “Disputed” and 1 had “Refuted” evidence. Conversely, 3 had “No Clinical Evidence”, 2 had “Disputed” and 1 had “Refuted” evidence. For hereditary OC susceptibility, 2 had “Definitive”, 0 had “Strong”, 1 had “Disputed”, whereas 12 genes had “No Reported Evidence” of disease association. CHEK1, GEN1, RAD50 and RINT1 had “Limited”, “Disputed” and/or “No Clinical Evidence” for both BC and OC susceptibility and are available on hereditary cancer panels.

The use of a standardized clinical validity assessment should provide important information to a number of stakeholders in clinical genetics including clinicians in choosing the right test panel for patients and identify which genes are associated with which cancer risk, hereditary cancer experts prioritizing genes for development of medical management guidelines and diagnostic laboratories assessing the clinical relevance of genes included on their test panels with an overall goal of improving the care of patients at risk for hereditary breast and ovarian cancer. (NIH 4U01HG007437)
24. Predicting the Functional Impact of Kv7.1 Variants of Unknown Significance

Bian Li1,2, Jeffrey L. Mendenhall1,2, Brett M. Kroncke2,3,4, Keenan C. Taylor2,3, Hui Huang2,3, Derek Smith6, Carlos G. Vanoye5, Jeffrey D. Blume6, Alfred L. George, Jr., MD5, Charles R. Sanders2,3,4, Jens Meiler1,2

1. Department of Chemistry, Vanderbilt University, Nashville, TN 37232, USA; 2. Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA; 3. Department of Biochemistry, Vanderbilt University, Nashville, TN, 37232, USA; 4. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, 37232, USA; 5. Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA; 6. Department of Biostatistics, Vanderbilt University, Nashville, Tennessee 37240, USA

Background: An emerging standard-of-care for long QT syndrome (LQTS) employs clinical genetic testing to identify variants that may guide treatment. However, interpreting results from genetic testing is complicated by the presence of variants of unknown significance (VUS) for which there is inadequate experimental or clinical evidence to determine their pathogenicity with a high degree of certainty.

Methods and Results: In this study, we curated from the literature a high-quality set of 107 functionally characterized Kv7.1 variants. Based on this dataset, we first showed there exists disagreement about variant annotation between functional studies and clinical studies. This observation suggests that translation of functional abnormality to clinical disease manifestation be done with caution. To suggest evidence at the functional level, we trained a neural network, Q1VarPred, specifically for predicting the functional impact of Kv7.1 variants using this experimentally validated dataset. The choice of two specific predictive features were based on detailed analysis on sequence conservation to KCNQ1 function. The estimated predictive performance of Q1VarPred in terms of Matthew’s correlation coefficient and area under the receiver operating characteristic curve were 0.581 and 0.884, respectively. When compared with five non-specific prediction methods, Q1VarPred has an overall better performance.

Conclusions: Distinguishing disease-causing variants from rare benign variants is of critical importance in interpreting the results of clinical genetic testing in LQTS. Although a plethora of tools are available for making predictions over a genome-wide scale, none has sufficient accuracy to warrant clinical application. The performance of Q1VarPred suggests a potential future direction for developing informatics tools that help clinicians to diminish uncertainty raised by novel Kv7.1 variants.
Controversy exists regarding the consistency of variant classifications both within and outside of ClinVar, and recently published studies have come to very different conclusions on that topic. We performed a systematic analysis of data in ClinVar both to understand the nature of agreement and disagreement among submitters, and to understand factors that could lead to differing conclusions in these studies.

We compared ClinVar classifications as of Oct. 2016 both (a) on an actionability basis, i.e. whether the variant would (P or LP) or would not (VUS, LB or B) suggest a change in care for a patient; and (b) on a pathogenicity basis, where VUS were distinct from B/LB. This abstract focuses on actionability although we will present both analyses. We particularly examined variants for which there was consensus (2/3 submitters in agreement) but not unanimity, and we cataloged properties of the outlier interpretations.

Overall, ClinVar classifications are highly concordant: 96.7% of variants reach consensus and 94.1% have complete agreement. Submissions from non-clinical lab sources are 6-fold more likely to be outliers compared to clinical lab submissions. In those hereditary cancer genes listed in the NCCN guidelines, consensus was even higher (98.8%) and non-clinical interpretations were 16-fold more likely to be outliers. Concordance varied by clinical area, with cancer and pediatric condition genes the most concordant, cardiology the least, and neurology and metabolic disease in between. Old classifications were up to 5-fold more likely to be outliers compared to recent ones, depending on age. Notably, low penetrance variants were often discordant: only 78.2% reached consensus and 49.2% complete agreement.

The effect of source, age clinical area and penetrance on discordance is perhaps not surprising. However, from examining recently published studies of interpretation concordance in detail, these factors appear to explain many of the differing conclusions. The data show this to be particularly true in the hotly debated BRCA1 and BRCA2 genes. Failure to take such factors into account may, in some cases, be deliberate, although as ClinVar rapidly increases in use, interface changes to help non-expert users avoid such pitfalls may be increasingly important. The ClinGen community can play an important role in educating clinicians and payers in both the proper and problematic uses of these vital public resources.
26. Text Mining PubMed to Improve the Prioritization, Curation, and Integration of Knowledge for Clinically Relevant Variants

Chih-Hsuan Wei¹, Lon Phan¹, Juliana Feltz¹, Rama Maiti¹, Tim Hefferon¹, and Zhiyong Lu¹

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

Contact: Zhiyong.Lu@nih.gov

Understanding the associations of genomic variants with diseases and conditions and assessing their clinical significance is critical for genomic research precision medicine. Despite significant efforts in expert curation, information about most of the 154 million dbSNP Build 149 reference variants (RS) remains unknown. On the contrary, a wealth of human knowledge about the variant biological function and disease impact is buried in unstructured literature data. Previous studies have attempted to harvest and unlock such information with text-mining techniques but are of limited use because their mutation extraction results are not standardized or integrated with curated data.

We describe a novel text-mining method to extract variant mentions in the literature and subsequently normalize them to corresponding standardized dbSNP RS numbers, which are unique identifiers with aggregated genomic information such as associated gene, clinical significance, and allele frequency. Our method, in benchmarking results, demonstrates a high accuracy of ~90% in F-measure and compared favorably to the state of the art.

Next, we applied our approach to the entire PubMed and validated the results by verifying that each text-mined SNV-gene pair matched the dbSNP annotation based on genomic position, and by analyzing variants curated in ClinVar. We then determined which text-mined variants and genes are novel records and information not found in dbSNP or ClinVar. Our analysis reveals 41,889 RS numbers and 9,151 genes not found in ClinVar. Moreover, our results include 12,462 rare variants (MAF ≤ 0.01) in 3,849 genes which are presumed to be deleterious and are not frequently found in the general population.

Furthermore, to demonstrate the utility of our approach in assisting manual variant curation, we prioritized and ranked text-mined RS results and manually curated a subset of 10 ultra-rare variants (MAF ≤ 0.001) novel to ClinVar, including several ACMG genes, with the functional class missense, frameshift or nonsense variations. Our manual analysis showed all 10 variants to be described in the literature as associated with diseases or cancer. These findings demonstrate that our approach can identify high impact variants from publications and that our results can be combined with dbSNP and ClinVar data to prioritize and rank the variants by functional consequence, allele frequency, gene annotation, and clinical significance for further manual or automated analysis and interpretation of effects on biological functions and diseases to enrich our current knowledge.
To our knowledge, we are the first to develop and apply an automatic mutation detection and normalization method for extracting genomic variant information in literature and integrating such data with an existing curated knowledge-base. Our large-scale analysis shows that combining automatically extracted information from PubMed articles with existing variant database annotations can significantly aid human efforts in curating and prioritizing variants in genomic research.

Availability: Our PubMed-scale text-mined data is publicly available to the entire scientific community through multiple channels. The complete data can be downloaded from our ftp website. Alternatively, we have incorporated our results in PubTator, an NCBI text-mining web service for visualizing and retrieving pre-computed biomedical entities and relations in the entire literature, where data of individual articles can be retrieved through web APIs. Finally, our mutation extraction and normalization software tool, tmVar, is already made open source. We are also integrating these computed data into dbSNP.
27. Simulating the Functional Impact of Somatic Mutation on ALK Resistance to Tyrosine Kinase Inhibitors
Matthew D. McCoy¹, Keeshaloy Thompson¹, Subha Madhavan¹
Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington DC, USA

Multiple sequencing efforts have identified drug resistant somatic missense mutations in anaplastic lymphoma kinase (ALK), which has led to widespread consideration of sequence variation during treatment, development of alternative therapies, and interpretation/design of clinical trials. Deregulated expression of ALK has been identified in a number of cancers, acting as an oncogene in most anaplastic large cell lymphomas to regulate cellular proliferation and differentiation. To mitigate the molecular signaling pathways controlled by ALK phosphorylation, a class of drugs known as tyrosine kinase inhibitors are used to compete for binding with ATP. Crizotinib was the first of these drugs to be widely used as a targeted therapy for ALK positive cancers, but was not universally effective and subject to selective tumor evolution pressures leading to drug resistance. Somatic variation which occurs in the ATP binding pocket disrupts the function of the inhibitor, but the success of second line interventions using alternative tyrosine kinase inhibitors like alectinib and ceritinib indicate drug interactions are differentially impacted. Binding is influenced by both the chemical composition and structure of the small molecule therapy, as well as the environment created by the interacting amino acid side chains. Because testing every somatic variant against every molecular therapy is not feasible in an experimental setting, computational binding simulations provide a powerful tool to not only understand the mechanism of drug resistance, but also to predict or rank the potential effectiveness of personalized therapy options. Beginning with the crystal structure of the ALK tyrosine kinase domain complexed with ADP, eleven tyrosine kinase inhibitor molecules (as well as the natural ligands ATP and ADP) were docked to seven drug resistant somatic variants of ALK using AutoDock Vina. Side chain flexibility of fourteen amino acids (including somatic variant loci) lining the ATP binding pocket was explicitly modeled and the docking results were structurally clustered to identify the most stable conformations for each of the 104 variant/drug combinations. Comparison to experimental data provides context for interpreting the results and making predictions where data does not exist, and shows the potential to aid decision making when selecting a personalized therapy regimen.
28. Combining Protein and Genome Annotations for Deciphering Functional Effects of Variation

Peter McGarvey PhD\textsuperscript{1,3}, Andrew Nightingale PhD\textsuperscript{3}, Hongzhan Huang PhD\textsuperscript{2,3}, Maria Martin PhD\textsuperscript{3}, Cathy Wu PhD\textsuperscript{2,3}, Subha Madhavan PhD\textsuperscript{1} Alex Bateman PhD\textsuperscript{3} and the UniProt Consortium\textsuperscript{3}

1 Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington DC
2 Department of Computer and Information Science, University of Delaware, Newark, DE
3 European Bioinformatics Institute, Hinxton, UK; Swiss Institute of Bioinformatics, Geneva, Switzerland; Protein Information Resource, Washington DC, USA.

Genomic variants cause deleterious effects through many mechanisms, including aberrant gene transcription and splicing, disruption of translation, and altered protein structure and function. Understanding the potential effects of non-synonymous SNPs on protein function is key for clinical interpretation, but this information is not readily available. Standard NGS annotation tools identify simple issues like translation stops and frameshift mutations, but for missense mutations connecting these variations to more subtle effects like the potential disruption of enzymatic active sites, protein-binding sites, or post-translational modifications is not easily achieved. UniProtKB represents decades of effort in capturing protein functional information and phenotypes or diseases associated with variants via literature-based and semi-automated expert curation. Included in this information are sequence positional features such as enzyme active sites; modified residues; binding domains; protein isoforms; glycosylation sites; protein variations and more.

The locations of 98,237 human reference proteins including isoforms and sequence variations from UniProtKB are mapped to GRCh38. Over 26 annotation types (e.g. active sites, modified residues, domains and natural variants) with associated curated information are aligned. Public track hubs for the Ensembl and UCSC genome browsers were created. Specific biological examples in disease-related genes and proteins are shown illustrating the utility of combining protein and genome annotations for the functional interpretation of variants. A larger comparison of UniProtKB features and variants that collocate with SNP data from ClinVar shows that currently 50% of UniProtKB disease associated variants exist in ClinVar and 35% of ClinVar’s SNPs with harmful assertions are present as variants in UniProtKB. There is general agreement in annotation between variations with 83% agreement in what is a "harmful" variant in UniProt and ClinVar and 85% agreement in "benign" variants. That some annotations are discordant between databases is not surprising as standards, methods and levels of evidence employed by protein curators and medical geneticists are different and evolving. The combination of gene and protein annotation can assist medical geneticists with the functional interpretation of variation while variant curation at the genome level helps protein curators refine critical structural and functional domains.
The progression of the ClinGen gene clinical validity classification over time

Jennifer McGlaughon, Jennifer Goldstein, Jonathan Berg
University of North Carolina at Chapel Hill, NC, USA

The objective of the NIH-funded Clinical Genome Resource (ClinGen) is to build a central resource that defines the clinical relevance of genes and variants for use in medicine and research. The gene curation working group at ClinGen has developed a framework to semi-quantitatively define the clinical validity of a gene-disease relationship (Strande et al., 2017). As a resource for use by clinicians and clinical laboratories, it is important for the gene curation activities of ClinGen to reflect up-to-date information for gene-disease relationships. There are many factors that can influence the clinical validity classification of a gene over time. Understanding these factors will provide insight into determining how often the classification of a gene should be re-evaluated by the gene curation teams. In order to explore these factors, we chose a diverse selection of curated gene-disease pairs and, using the ClinGen Gene Curation framework, simulated clinical validity assessments at one-year intervals from the time of first publication. Literature was identified as it accumulated and the classification was determined at each time point. The following information was collected to determine what variables impact how the clinical validity classification changes over time: year of first report of a gene-disease relationship, journal of publication, technologies used (linkage analysis, Whole Exome Sequencing, Next-Gen Sequencing, etc.), disease frequency, and the disease clinical domain. Additionally, variables that influence whether a "limited" gene-disease classification advances to "moderate" or "strong" are being explored. This retrospective analysis of clinical validity classifications over time will be useful in determining the most acceptable timeframe for re-evaluating gene-disease relationships as more information becomes available in the literature.
30. The development of interactive gene curation training modules for ClinGen biocurators

Jennifer McGlaughon¹, Jennifer Goldstein¹, Erin Rooney Riggs², Marina DiStefano³, Raj Ghosh⁴, Courtney Thaxton¹, Jonathan Berg¹

¹University of North Carolina at Chapel Hill, NC, USA ²Autism and Development Institute, Geisinger Health System, Lewisburg, PA, USA ³Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA ⁴Baylor College of Medicine, Houston, TX, USA

The objective of the NIH-funded Clinical Genome Resource (ClinGen) is to build a central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. The ClinGen gene curation working group has developed a framework for curators to semi-quantitatively define the clinical validity of a gene-disease relationship (Strande et al., 2017). As the ClinGen knowledge base is a resource for use by clinicians and clinical laboratories, it is important that there is consistency and accuracy among the more than 50 biocurators involved in gene curation projects across more than 25 institutions. The ClinGen biocurator working group was formed with the goal of providing education, training and support for all biocurators. To support this mission, interactive gene curation training modules were developed to introduce new biocurators to the gene curation process. We will present a live demo of these interactive training modules. The objective of the training modules is to present the biocurator with a variety of genetic and/or experimental evidence for a gene-disease relationship and prompt the biocurator to decide how the evidence should be assessed using the ClinGen semi-quantitative framework. For example, in one training module, biocurators are presented with evidence from a research paper that proposes a novel association between KLHL24 and epidermolysis bullosa (He et al., 2016). This paper was chosen due to the quantity and diversity of evidence (genetic, segregation, experimental) for the proposed gene-disease relationship. Biocurators are introduced to each piece of evidence, followed by a multiple choice question to assess how they would score the evidence based on the ClinGen clinical validity framework. Additional interactive training modules are being developed to demonstrate additional concepts with varying levels of complexity, including inheritance pattern differences and varying types of evidence. These modules will serve as a starting point for new biocurators, as continuing education for current biocurators, and will promote consistency in evidence interpretation for a gene-disease relationship across the ClinGen network.
31. Gene-Specific Criteria for PTEN Variant Curation

Jessica Mester¹, Rajarshi Ghosh², Tina Pesaran³, Rachid Karam³, Robert Huether⁴, Kathleen Hruska¹, Helio Costa⁵, Katherine Lachlan⁶, Joanne Ngeow⁷, Jill Barnholtz-Sloan⁸, Liying Zhang⁹, Kaitlin Sesock¹⁰, Sharon Plon², Madhuri Hegde¹¹, Charis Eng⁸¹⁰

¹: GeneDx Laboratories, Gaithersburg, MD
²: Baylor College of Medicine, Houston, TX
³: Ambry Genetics, Aliso Viejo, CA
⁴: Tempus, Chicago, IL
⁵: Stanford University School of Medicine, Stanford, CA
⁶: Princess Anne Hospital, Southampton, UK
⁷: National Cancer Centre Singapore, Singapore
⁸: Case Comprehensive Cancer Center, Cleveland, OH
⁹: Memorial Sloan Kettering Cancer Center, New York, NY
¹⁰: Cleveland Clinic Foundation, Cleveland, OH
¹¹: Emory Genetics Laboratory, Atlanta, GA

The PTEN Expert Panel was the first ClinGen group assembled within the Hereditary Cancer domain, with a primary goal to tailor the 2015 ACMG/AMP Variant Interpretation Guidelines specific to PTEN. The group, which meets monthly, includes clinicians, researchers, and diagnostic laboratory members with a special interest in and experience with PTEN. We present our finalized benign and draft pathogenic criteria, which after evaluating with a test variant curation set will be finalized and incorporated into the group's curation process as part of the ClinGen Expert Panel application. The group developed and tested draft benign criteria on a set of 15 variants classified as benign/likely benign (BEN/LBEN) per multiple ClinVar submitters, and were advised that a concordance of 80% would prove acceptable. Using these criteria the group arrived at a BEN/LBEN classification for 13 (86.7%). The two variants not meeting criteria for BEN/LBEN included a promoter variant, c.-1311T>C, and a synonymous variant c.75G>A (p.L25L). Although c.-1311T>C is present in 23/1618 (1.4%) East Asian alleles per gnomAD, the allele denominator in this subpopulation is below the recommended 2,000 minimum threshold, leaving us unable to apply criteria that would otherwise lead to a BEN classification. PTEN c.75G>A is near a predicted U12-dependent splice donor for which in silico tools are not available. The group has also drafted pathogenic criteria. Optimizations to the ACMG/AMP pathogenic criteria specific to PTEN include phenotype criteria for PP4, boundaries for use of PVS1 vs. PM4 for truncating variants not predicted to undergo nonsense-mediated decay, functional domain definitions, functional assays qualifying for use of PS3 or supporting level evidence, numbers of meioses and families for segregation criteria, and use of a strong criterion for more than one de novo occurrence without maternity/paternity confirmed. These criteria will be tested on a set of 15 variants classified as pathogenic/likely pathogenic per multiple ClinVar submitters. Should a concordance of 80% again be achieved, the criteria will be considered final and be incorporated along with the curation process into the expert panel application. These criteria will be compared with gene-specific optimizations developed by other ClinGen expert panels to identify common and unique modifications to the ACMG/AMP guidelines.
32. New Developments and Features of the My Cancer Genome Precision Cancer Medicine Knowledge Resource


My Cancer Genome (MCG) is a knowledgebase, website (www.mycancergenome.org), and set of tools that provides information about molecular alterations in cancer and allows physicians to match patients to therapies and clinical trials. MCG is in the midst of a period of transformative change. This abstract will outline a few of the changes and set the stage for new developments to come.

Over several years, MCG has been working to create a new knowledge management system (KMS). The KMS includes the ability to define disease, alteration, and other vocabularies; curate therapeutic and prognostic assertions; curate clinical trials, including biomarker eligibility criteria; and classify references such as medical literature, clinical practice guidelines, and drug labels.

Vocabulary management includes different features depending on the term type. For example, the disease ontology includes the ability to define parent-child relationships and mappings to other ontologies. The gene and alteration vocabularies include validation against reference resources and enforcement of HGVS nomenclature. Specific alterations roll up to region-based alterations; sequence and copy number variants, rearrangements, karyotypes, and protein expression are included. In addition, an override feature provides the flexibility to add alterations that the KMS is unable to parse.

An important piece of the KMS is the clinical trial curation system. All new and changed cancer clinical trial documents from clinicaltrials.gov are retrieved nightly. Text mining is used to highlight disease and biomarker keywords. Manual curation is then used to extract accurate diagnosis and molecular biomarker eligibility criteria for each biomarker-driven trial. Analyses of trial matching to large basket trials like NCI-MATCH and ASCO’s TAPUR study will be presented.

A reference manager incorporating our levels of evidence classification system for somatic alterations in cancer is in development. This system permits capture of evidence supporting therapeutic and prognostic assertions in a way that enables automated updating of the level of evidence supporting specific therapeutic and prognostic assertions. Clinical trial outcomes will also be curated in the reference manager.

Finally, one of the driving principles of MCG over the last seven years has been to focus on its target audiences’ needs. MCG dissemination strategy will be reviewed, and the completely redesigned MCG website will be presented.
Multiple bioinformatic tools have been created that can perform structural variant (SV) detection (e.g. LUMPY, MANTRA, DELLY). However, these tools often produce divergent variant calls making it very difficult to reconcile the resulting variants into a single, accurate set. To assure confidence in these variants, further steps of validation are necessary which typically require expert manual curation and can be difficult and time-intensive to reduce false positives. Here we present a novel method to visualize sequencing read alignments to aid in variant curation.

Our software algorithm, GRAPHITE (https://github.com/dillon/graphite), takes as input a set of variant calls from one or more detection tools and applies a novel “variant adjudication” procedure to discard false positive, while keeping true negatives. This is accomplished by constructing a graph from these variants (the Variant Graph) where the reference and alternate alleles are represented as different branches. GRAPHITE then applies a graph mapping algorithm (GSSW, a graph extension of the Smith-Waterman alignment algorithm) to re-map reads contributing to the candidate alleles. Reads are assigned to their respective branches on the graph based on their alignment scores produced by the mapping algorithm. Candidate variants not confirmed by re-mapping are then discarded from further analysis. This results in a call set that is highly specific due to the algorithm and highly sensitive by starting with a call set from previous SV caller tools.

Visualization of the variant call set is a key method when confirming or rejecting novel candidate alleles. Current visualization techniques require the analyst to verify SV breakpoints by closely tracking deviations from the reference which is cumbersome and can lead to errors. GRAPHITE’s output has been formatted to visualize sequencing read alignments against a Variant Graph, rather than a single reference string, which can be intuitively displayed by the popular IGV (Integrative Genome Viewer) alignment viewer program. With this format, each branch can be visually inspected with the appropriate reads realigned. This view greatly simplifies the effort needed to confirm the presence of SV’s which can lead to a more accurate diagnosis.
34. Recommendations for Classification, Interpretation and Reporting of Sequence Variants in Oncology


The increasing use of sequencing technology in the oncology space has given rise to an evolving need for a robust, yet adaptable framework for tumor variant classification and reporting. At Human Longevity Inc. (HLI), a health intelligence company, we provide germline and tumor DNA sequencing to identify actionable targets that may be driving the specific cancer. The goal is to identify potential for targeted therapy for on-label and off-label use of an approved drug, and to identify relevant clinical trial options. Accurate classification and interpretation of variants is critical for the above clinical decision making process. A major distinction between germline and tumor variant classification is that, the latter is driven by clinical utility of the alterations and is determined by several factors. Unlike germline variants which have established guidelines, somatic variant guidelines are still evolving, and the current framework is inadequate to address the expanding definitions of clinical actionability. The goal of this presentation is to delineate the practical challenges facing somatic variant classification and reporting, and how we have addressed the issue. The nature of cancer as a disease, the role of gene-variant-disease-drug association in driving tumor response, secondary mutations, regulatory and professional guidelines, and experiment therapy options are some of the factors that drive variant classification. Another important component of tumor variant classification is the organization of the contents of the report, which is key to avoiding ambiguity and presenting interpretations with maximum clarity. In this presentation, we will describe tumor variant classification and reporting at HLI. We will specifically address the complexities of somatic variant interpretations and areas that lead to ambiguity in reporting, caveats in interpreting FDA labels and NCCN guidelines, ranking evidences from published studies, reporting therapies for multiple driver mutations, and application of disease ontology.
Using an integrated approach to interpret CNVs, AOH and Sequence Variants: Unmasking Compound Heterozygous Alterations

Andrea O’Hara, Alessio Venier, Zhiwei Che, Shalini Verma, Soheil Shams

BioDiscovery, Inc., 715 N. Douglas St., El Segundo, CA 90245

Given the separate technologies used to detect CNVs, AOH, and Sequence Variants (SNP Microarrays for CNVs and AOH, and NGS for Sequence Variants), this data is commonly reviewed independently; this segregation is further exaggerated by the separate expertise in Cytogenetics and Molecular Genetics. As these fields are now merging, the importance of an integrated view of all the data has become even more apparent. Here we present a complex case using a new system that can integrate data from array and NGS platforms to create a single genomic view of structural and sequence changes in an individual sample. A family trio was subjected to whole exome sequencing (WES) and SNP microarray analysis to identify the causative variant(s) for the proband affected with a complex phenotype including severe global developmental delay, infantile spasms and limb dystonia. High-density SNP microarray analysis of the proband showed no remarkable primary causative copy number alterations alone, while whole exome sequencing of the proband yielded >177,000 variants. Filtering of sequence variants to exclude variants of poor quality, low read depth and common polymorphisms reduced this to >20,000 variants. After refinement to exclude noncoding and synonymous variants, among others, 388 sequence variants remained. Incorporation of phenotypic information reduced this to 88 variants. De novo and inherited events were then evaluated, which included the possibility of an inherited compound heterozygous event. Applying advanced filtering using integration of data modalities, only the sequence variants in regions of copy number alteration or stretches of homozygosity were displayed. This reduced the total count of variants for review to only six alterations. A single compound heterozygous aberration was easily observed which identified a heterozygous (single copy) microdeletion, inherited by the mother, in a gene with a secondary in-frame deletion, inherited by the father; this gene is associated with both epilepsy and mental retardation. While originally detected by the CNV analysis, it was initially dismissed as an autosomal recessive gene. This case demonstrates the utility of a single system to identify complex aberrations detected using multiple technologies. An integrated view of CNVs, AOH and sequence variants assures a rapid detection of compound events and the integration with standardized phenotype terms in the form of HPO IDs facilitates the event classification.
36. NIAID Clinical Genomics Tools

Andrew J. Oler¹, Sandhya Xirasagar¹, Zhiwen Li¹, Xi Cheng¹, David Kaminsky¹, Morgan Similuk², Yu
Zhang³, Joyce Johnson², Celine Hong⁴, Joshua D. Milner⁵

¹Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational
Biology, ²Laboratory of Immunology, ³Laboratory of Host Defenses, ⁴Division of Intramural Research,
⁵Laboratory of Allergic Diseases
National Institute of Allergy and Infectious Diseases (NIAID), NIH, Bethesda, MD

The National Institute of Allergy and Infectious Diseases (NIAID) Intramural Clinical Genomics Program
was started in 2014 with the goal of using state of the art genomics technologies to improve diagnosis
and treatment of rare or novel immunological disorders. To facilitate their goal, we have developed an
integrative, scalable, collaboration-engendering secure web application, Genomic Research Integration
System (GRIS). GRIS enables the capture of patient data with standardized phenotype and genotype
data in a family- or individual-based context and importantly, genetic analysis utilizing high quality
annotations. The system leverages existing open source tools, ontologies, and genomic annotation
databases. A customized version of PhenoTips with a user-friendly pedigree editor interface and Human
Phenotype Ontology (HPO) has been implemented to enable patient data capture. Standardized analysis
pipelines use open source tools including BWA, GATK/Queue, and GEMINI to process sequencing and
variant data and perform quality control at various stages of data processing. The resulting variant report,
enriched with genomic annotations from ClinVar, OMIM, ExAC, gnomAD, dbNSFP and other high quality
annotation databases are displayed in an interactive web-based tool, GeminiViewer, to aid researchers in
identifying candidate causal variants in Mendelian disorders.
37. ClinGen Allele Registry supports scalable integration of data and growth of knowledge about genetic variants

Piotr Pawliczek¹, Ronak Y Patel¹, Lillian R Ashmore¹, Andrew R. Jackson¹, Neethu Shah¹, Chris Bizon⁴, Tristan Nelson⁵, Steven Harrison², Bradford Powell⁶, Robert Freimuth⁴, Larry Babb⁷, Sharon E Plon¹, Aleksandar Milosavljevic¹ on behalf of the ClinGen Resource

1. Baylor College of Medicine, Houston, TX
2. Laboratory of Molecular Medicine, Partners HealthCare Personalized Medicine
3. Renaissance Computing Institute, Chapel Hill, North Carolina
4. Mayo Clinic Rochester Minnesota
5. Geisinger autism and developmental medicine, Lewisburg, PA 17837
6. University of North Carolina, Department of Genetics, Chapel Hill, North Carolina
7. GeneInsight a Sunquest Company, Boston, MA

To help organize a growing data ecosystem around genetic variants, ClinGen investigators have developed the ClinGen Allele Registry (services and documentation are available at http://reg.clinicalgenome.org). As a key component of the federated ecosystem, the Registry enables the linking of information about genetic variants by providing unique identifiers for genetic variants from publicly available databases as well as novel variants identified through research or clinical testing. The variants are automatically mapped across known reference sequences and to identifiers from major variant databases using an in-memory sequence-based index. Integrated validation and normalization facilitate unambiguous resolution of variant identity. The identifiers are provided both via web interface and programmatically (via stable and well-documented APIs) in the form of dereferenceable Uniform Resource Identifiers (URIs). The Registry services are the first to implement on-demand high-bandwidth registration of genomic variants using the current Linked Data standards including JSON-LD. The Registry houses dereferenceable identifiers for more than 600 million alleles. These include identifiers from gnomAD, ExAC, dbSNP, MyVariant.info, ClinVar and other key sources. Conceptual and resource model documentation for Alleles is available at http://datamodel.clinicalgenome.org/development/allele/.

We present the features of the next major Registry release and the key use cases. A new on-demand high-bandwidth registration service for variants defined using HGVS expressions or VCF files enables rapid registration of millions variants per request. We demonstrate the functionality of the newly redesigned Allele Registry user interface, the process for registering new variants and the linking of information about them. We demonstrate collation of evidence about indels that have highly diverse representations yet have common canonical identifiers. We also demonstrate collation of evidence between the nucleotide and amino acid levels. We highlight programmatic use of the Registry via the APIs by ClinGen applications, including the Pathogenicity Calculator and Variant Curation Interface.
We discuss two use cases demonstrating the utility of the Registry to collate information about alleles from different sources using APIs and Linked Data technologies. In one use case, we simulate two sources of information about \textit{BRCA1} variants, both sources producing JSON-LD responses to CAid queries. In another use case, we present results and limitations of our initial efforts to integrate information from model organisms to help resolve human variants of unknown significance. In summary, the use cases demonstrate data interoperability enabled by Registry services and illustrate the features of the next Registry release that will further help organize the growing ecosystem of variant data and knowledge.

This work was funded by the NIH National Human Genome Research Institute (NHGRI) grants U01 HG007436-01 and U41HG006834.

* Both authors contributed equally to the work
Germline pathogenic variants in TP53 cause Li-Fraumeni Syndrome (LFS), an autosomal dominant cancer predisposition syndrome associated with high risk of malignancy, including soft tissue sarcoma, osteosarcoma, premenopausal breast cancer, brain tumors, adrenocortical carcinoma, and leukemia. Individuals with germline TP53 pathogenic variants often develop LFS-associated cancers in early childhood or early adulthood and are at increased risk for multiple primary cancers. It is estimated that up to 80% of individuals meeting classic clinical criteria for LFS have a detectable TP53 pathogenic variant, most of which are missense variants. In addition, the frequency of de novo TP53 mutations in LFS is around 7-20%. At the somatic level, mutations in TP53 are frequent in the majority of cancers. Germline pathogenic variants in TP53 are clinically actionable prompting the National Comprehensive Cancer Network to publish guidelines for screening recommendations and counseling about risk-reduction strategies for individuals with TP53 germline mutations and their at-risk relatives. Accurate and consistent classification of variants in TP53 across clinical and research laboratories are therefore very important for patient care.

The ClinGen TP53 Expert panel was formed under the umbrella of the Hereditary Cancer Domain and tasked with the goal to optimize the 2015 ACMG-AMP Variant Interpretation Guidelines for clinical interpretation of variants identified in TP53. Members of the panel consist of clinicians, researchers, genetic counselors, statisticians, structural biologists and diagnostic laboratory members with expertise and experience in TP53-associated pathology and TP53 variant classification. The panel meets monthly and has created three working groups: Computational/Predictive working group, Phenotype/Segregation/De novo working group and the Functional Data working group. The working groups have been reviewing and optimizing the ACMG-AMP guidelines applicable to TP53 variants and presenting to the larger expert committee for comment and discussion. Here we present the optimized pathogenic and benign evidence criteria including recommendations on LFS-associated phenotypic criteria, segregation, de novo events, population frequency cut-offs, in silico predictors, functional evidence and
splicing evidence. These criteria will be validated on a set of established benign and pathogenic variants as an initial test of these optimizations.
39. Representing Variant Interpretations and Supporting Evidence in a Data Model with Defined Provenance

Bradford C. Powell MD-PhD(1), Matthew Brush PhD(2) Marina DiStefano PhD(3), Robert R. Freimuth PhD(4), Rajarshi Ghosh PhD(5), Melissa A. Haendel PhD(2), Steven Harrison PhD(6), Tristan Nelson(7), Heidi L. Rehm, PhD(3,6), Shawn Rynearson(8), Natasha Strande PhD(1), Chris Bizon PhD(9), Larry Babb(10), on behalf of the ClinGen Consortium Data Model Working Group

(1) University of North Carolina at Chapel Hill, Chapel Hill NC
(2) Oregon Health and Science University, Department of Medical Informatics and Clinical Epidemiology, Portland, OR
(3) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge MA
(4) Mayo Clinic, Department of Health Sciences Research, Rochester, MN
(5) Baylor College of Medicine, Department of Pediatrics-Oncology, Houston, TX
(6) Harvard Medical School, Boston, MA
(7) Geisinger Autism and Developmental Medicine Institute, PA
(8) University of Utah, Department of Biomedical Informatics, Salt Lake City UT
(9) Renaissance Computing Institute, University of North Carolina at Chapel Hill, NC
(10) GeneInsight a Sunquest Company, Boston MA

Sharing of the structured clinical interpretation of genetic sequence variants is an important approach to solving the data challenges inherent in clinical genomic sequencing. A statement of pathogenicity alone or with only unstructured metadata does not provide sufficient detail for reanalysis of the diverse sources of data that inform expert interpretation or adjudication of discordant interpretations.

The Data Model Working Group (DMWG) of the Clinical Genome Resource (ClinGen) develops representations of data relevant to clinical genomics to aid in the collection and exchange of these data within ClinGen and among the broader genetics community. The ClinGen Interpretation Model encompasses representations of the overall assertion along with supporting evidence and a structured description of how evidence was obtained and combined.

Development of the ClinGen Interpretation Model is guided by concrete examples of application of the ACMG-AMP variant interpretation guidelines, many derived from the experience of the CSER consortium’s comparison of variant interpretations among its sites. We have formally described over 95 examples documenting interpretation criteria using over 280 captured data attributes used in one or more of the ACMG-AMP criteria. The ability to accurately and efficiently represent these examples was a key goal of the data model, ensuring that the data model satisfies its primary use case: to enable the transparent communication of semantically rich and structured clinically-relevant evidence underlying variant interpretations.

Our model aligns with the Scientific Evidence and Provenance Information Ontology (SEPIO), a generalizable model for linking scientific assertions to underlying evidence in development by the Monarch Initiative. SEPIO will provide a unified and flexible representation of evidence for assertions with integration across diverse biomedical databases. Alignment with SEPIO will enable long-term interoperability of our data model with data sources that conform to the SEPIO standard. The use of community ontologies and Linked Data principles enables semantic queries and extensibility of the types of data that can be represented. We are developing tools to transform evidence-supported interpretations produced in the ClinGen Variant Curation Interface into the ClinGen Interpretation Model, and to submit these to ClinVar.

Structured data examples and documentation are available at http://datamodel.clinicalgenome.org.
40. MACE2K: Molecular and Clinical Extraction to Knowledge, a tool for Precision Medicine

Shruti Rao1, A. S. M. Ashique Mahmood2, K. Vijay-Shanker2, Vishakha Sharma1, Simina Boca1,4, Karen Ross5, Peter McGarvey1,5, Robert Beckman1,4, Cathy Wu2,3,5, Subha Madhavan1,4
1) Innovation Center For Biomedical Informatics, Georgetown University, Washington D.C, U.S.A; 2) Department of Computer and Information Science, University of Delaware, Newark, DE, U.S.A; 3) Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, U.S.A; 4) Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington D.C, U.S.A; 5) Protein Information Resource, Georgetown University Medical Center, Washington D.C, U.S.A.

The velocity, variety and volume of data from biomedical literature makes it challenging for oncologists to collect and review pertinent information that can support personalized treatment for their patients. Moreover, it takes a lot of time and effort to manually extract relevant information from literature about the clinical actionability of these biomarkers. In order to help clinicians and clinical researchers assess the overall evidence associated with biomarkers that predict response to cancer therapies, we developed a software suite for precision medicine called MACE2K.

We first developed a natural language processing (NLP) system that can detect different entities and relations such as cancer types, gene/protein names, molecular anomalies, therapies and disease outcome terms from Medline abstracts. For example, in the following sentence from an abstract by Wang et. al (PMID: 21370501) “No correlation was observed between gene expression of RRM1 and that of BRCA1 (P>0.05), but there was a strong correlation between the expression of RRM1 and the response to chemotherapy (P=0.003)”, the NLP system will indicate the relation between the gene RRM1 and the “response to chemotherapy” with the mnemonic GE-ASS-RO, where GE is gene expression, ASS is association and RO is response/outcome. This tool relies on the syntactic nature of the sentences coupled with various textual features for the extraction of relations between genomic anomalies and drug response. Our system achieved precision, recall and F-measure of up to 0.95, 0.84 and 0.89, respectively from the evaluation of different annotated datasets, both created in-house and obtained externally from PharmGKB. To date we have processed 37,243 abstracts for 50 different gene-drug combinations. We then developed a ranking algorithm that uses the NLP output to determine the relevance of the abstract to precision medicine. The NLP tool produces output in JSON, CSV and HTML formats to facilitate data exchange and integration of text mining results end user interfaces. All the data will be stored in a database designed for molecular NLP results from a variety of tools. Cognitive systems analysis methods will be applied to optimize user interface design.

The structuring, organization and analysis of dispersed public data and associated metadata from biomarker driven studies into MACE2K will enable researchers to readily generate hypotheses for new precision medicine based clinical trials.
Recently the Clinical Genome Resource (ClinGen) Somatic Working Group (WG) developed a standardized data format for curation of somatic variants: Minimum Variant Level Data (MVLD). We are actively collaborating with cancer databases to implement MVLD, and engaged in multiple curation efforts to disseminate MVLD. Curation Efforts: The Somatic WG has developed a strong collaboration with Clinical Interpretations of Variants in Cancer (CIViC) knowledgebase, a distributed, asynchronous, community-supported curation platform with a growing body of active curators. The Somatic WG identified a list of ~30 high-priority cancer genes and associated variants to curate in CIViC. Thus far, we have 3 active curators with 43 evidence items added to CIViC. In addition, we are developing a pediatric cancer sub-group that will focus curation on a list of ~19 genes currently not curated for pediatric cancers in CIViC. MVLD Engagement: We developed a scriptable conversion of CIViC curation to a ClinVar submission, which uses elements of MVLD to standardize that conversion process. CIViC has adopted the HGVS formatting to comply with MVLD, and ClinVar is implementing NCI codes to comply with MVLD. MVLD will add handling for the Association of Molecular Pathology interpretation guidelines, as well as Disease Ontology (DO). We are also collaborating with Clinical Sequencing Exploratory Research (CSER) to assist in data transformation using MVLD elements to convert data into a ClinVar submission. Standardizing Somatic Curation and Developing Somatic Expert Panels: We are developing standard operating procedures (SOPs) for the ClinGen somatic curation process, which incorporate elements of the ClinGen germline variant curation SOP where possible. The SOP considers nuances in interpreting predictive and prognostic actionability of somatic variants, drug response and resistance implications, as well as patient outcomes (such as progression free survival, disease free survival, overall survival) in clinical studies. The process includes steps for repeatability and transparency, such as history of literature searches, publications reviewed/excluded, dates of curation work, and time-tracking, as well as multiple-curator review with a cancer expert final review. We are currently reviewing the expert panel application process developed by ClinGen and will actively seek to develop a process for expert panels in somatic cancer.
Human Genome Variation Society (HGVS) nomenclature is a set of recommendations for describing biological sequence variants. Evolution of the nomenclature may lead to discrepancies in its application across different systems. Tools that manipulate HGVS-formatted variants may differ in the completeness or interpretation of the recommendations, potentially resulting in variants that are incorrectly formatted. This has the potential to lead to misinterpretation of clinically significant variants. Currently, there is no way to easily assess the accuracy of tools that either manipulate the nomenclature of, or predict the consequences of genomic variants. The goal of this project is to develop an evaluation framework which will use test-cases to assess the capabilities and measure the accuracy of HGVS tools. This framework: hgvs-eval will be made up of 3 components: 1) a command-line script that runs a set of verified test cases through each tool; 2) a REST API which allows for standardized and language agnostic access to the framework; and 3) a website which will display up-to-date results of tests. Test cases will assess features of HGVS-formating tools like parsing a variant, validating that the variant is correctly represented in HGVS nomenclature, normalizing a variant by 3’ shifting and projecting of a sequence variant from one type of sequence to another. An evaluation of select tools will be performed and results will be displayed via a website to allow for community-wide utilization and continual assessments and updates of results. This project was part of the GA4GH 2016 hackathon. Source code for this project is available at https://github.com/biocommons/hgvs-eval.
Maximizing the personal, research, and clinical value of genomic information will require that clinicians, researchers, and testing laboratories exchange data about sequence variation reliably. The Variation Modeling Collaboration (VMC; https://github.com/ga4gh/vmc) is a partnership of representatives from the ClinGen, NCBI ClinVar, FHIR Genomics, GA4GH, HGVS, and HL7 Clinical Genomics Working Group communities. Our goal is to propose and encourage the adoption of a data model with clear semantics for the computational representation of variation on any type of sequence. The VMC model will be interoperable with other existing formats and enable the reliable exchange of sequence variation among data providers and consumers.

The VMC draft specification precisely defines three essential genetic states — allele, haplotype, and genotype — and provides conceptual data models for each. Allele represents a contiguous sequence or sequence change with respect to a reference sequence. Haplotype represents alleles occurring on the same molecular sequence (that is, "in phase" or "in cis"). Genotype represents a list of haplotypes, thereby providing a representation of Alleles at any number of sites and observed with various specified ploidy. Alleles, haplotypes, and genotypes may be defined on DNA, RNA, and protein sequences.

In addition to its core models, the VMC draft specification includes additional guidance to standardize data exchange in "real world" uses. As such, it includes models to represent human-readable identifiers for instances of genetic variation, such as references to ClinVar variants or names for haplotypes. It also provides models to represent multiple notions of relationships between concepts, such as alleles related by normalization or translation. Additionally, the model specifies and recommends an algorithm for constructing distributed and globally-unique identifiers for genetic states. Importantly, this algorithm enables two parties to share variation data without prior agreement about an assigned identifier since the identifier is computed based on the data itself. Finally, the draft specification provides technical guidance for transmitting models using common serialization technologies.

The VMC model will improve the fidelity of exchanging sequence variation data and improve interoperability among existing standards.
GenomeConnect, the Clinical Genome Resource patient registry, was created to provide patients a secure mechanism for sharing genotype and phenotype information. Participants fill out surveys about their health, then upload their genetic testing reports; with their consent, this information is de-identified and shared with ClinGen approved databases, such as NCBI’s ClinVar. To date, GenomeConnect has submitted 361 sequence variants from 149 patients to ClinVar. Of these variants, 58.7% (212/361) have not previously been reported to ClinVar, demonstrating the importance of GenomeConnect as a data source. The remaining 41.3% of variants (149) have been previously submitted to ClinVar from a clinical laboratory; 61% (92/149) of these are from the same reporting laboratory as the patient’s report. In 10.9% of these cases (10/92), we identified discrepancies in variant interpretation between the result reported to the GenomeConnect participant at the time of testing and the laboratory’s current interpretation in ClinVar. Although many laboratories attempt to inform clinicians about updated classifications, this information may not always be communicated to the patient. Realizing that GenomeConnect could serve as a liaison to relay this potentially medically relevant information, we surveyed participants to determine their preferences for receiving such information. Of the 137 consented participants that completed the survey (response rate 19.6%), 99% indicated that they want to receive information about updated variant interpretations from GenomeConnect, and a process for providing these updates is now under development. As part of our submission to ClinVar, we also identified conflicting interpretations between different clinical laboratories. Of the 149 submitted variants with interpretations already in ClinVar, 32.9% conflicted with other laboratory submissions. In these instances, while we will not relay information back to participants, GenomeConnect will encourage laboratories to address these discrepancies through the ClinGen variant discrepancy resolution process. By engaging participants in genomic data-sharing efforts, GenomeConnect is able to contribute information to the public knowledge base that may not have otherwise been available; this process benefits both our participants and the genetics community.
Assessing the Clinical Validity of Genes Implicated in Hereditary Colorectal Cancer and Polyposis Using the ClinGen Framework

Bryce A. Seifert¹, Sarah Jackson², Jennifer McLaughon¹, Deborah Ritter³, Ryan J. Schmidt⁴, Sharisse Jimenez⁵, Kristy Lee⁶, Sharon E. Plon³, Kenneth Offit⁵, Jonathan S. Berg¹, Zsofia K. Stadler⁵, Liying Zhang⁵, Marc S. Greenblatt⁷, Matthew J. Ferber³

¹ University of North Carolina at Chapel Hill, Chapel Hill, NC
² GeneDx, Gaithersburg, MD
³ Baylor College of Medicine, Houston, TX
⁴ Brigham and Women's Hospital and Harvard Medical School, Boston, MA
⁵ Memorial Sloan Kettering Cancer Center, New York, NY
⁶ Mayo Clinic, Rochester, MN
⁷ University of Vermont, Burlington, VT

Recent advances in genomic technologies have elevated the number of genetic diagnostic panels for hereditary colorectal cancer and polyposis. However, many genes on these panels have not been systematically examined for their strength of association with these diseases. The Clinical Genome Resource (ClinGen) is an NIH-funded program for creating a publicly available resource that assesses the clinical relevance of genes and variants within specific diseases. As part of this effort, the ClinGen Clinical Validity framework provides a standardized systematic evaluation of literature evidence to assign strength to a gene’s association with disease into one of the following clinical validity classifications: Definitive, Strong, Moderate, Limited, No Reported Evidence, or Conflicting Evidence Reported.

Using this framework, the Hereditary Colorectal Cancer and Polyposis Gene Curation Team evaluated the strength of gene-disease associations in both diseases. We developed a comprehensive list of 35 genes found on current clinical testing panels, and assessed the evidence implicating these genes in hereditary colorectal cancer and polyposis. Two curators did preliminary classification, followed by a group discussion of the evidence by an expert panel to determine the final clinical validity classification. To date, final classifications have been assigned for 16/35 (45.7%) of genes. For hereditary colorectal cancer, one gene had “Definitive” evidence (POLE), one had “Strong” evidence (POLD1), three had “Moderate” evidence (NTHL1, AXIN2, and MLH3), and nine genes currently on panels had “Limited” evidence (BARD1, CDH1, EPHX1, GALNT12, RPS20, EXO1, CDKN1B, STK11, and CHEK2). For hereditary polyposis, one gene had “Strong” evidence (GREM1) and one had “Moderate” evidence (MSH3). Genes well-accepted for Lynch syndrome and polyposis have not yet been curated.

The benefit of an “expert panel” approach was exemplified by curating POLD1 for the Hereditary colorectal cancer phenotype; the original designation of “Moderate” was increased to “Strong” after team discussion. Overall, this team approach has highlighted our ability to tailor the clinical validity framework for gene-disease associations associated with polyposis phenotypes that progress to colorectal adenocarcinomas. Ultimately, these curations will benefit clinicians and aid diagnostic laboratories in assessing the clinical relevance of genes included on marketed hereditary cancer genetic testing panels.
Ambry Genetics performs comprehensive genetic testing on multiple hereditary conditions. Of these the lab performs testing on neuro developmental cases where, sequence enrichment of the targeted coding exons and adjacent intronic nucleotides is carried out by a bait-capture methodology using long biotinylated oligonucleotide probes. This is followed by polymerase chain reaction and next-generation sequencing of up to 196 genes associated with intellectual disability, autism spectrum disorders and epilepsy. Gross deletion/duplication analysis for all genes is performed utilizing a targeted chromosomal microarray. Similarly, for constitutional testing, Genomic deoxyribonucleic acid (gDNA) is labeled and hybridized to an oligonucleotide array with more than 1.9 million copy number probes and nearly 750,000 SNP probes used for genotyping and copy number analysis. This array allows for detection of loss of copy number (deletion), gain of copy number (duplication) or regions of homozygosity. Negative results can be reflexed to customizable exome sequencing of up to 500 genes. For both of these tests, CNV, AOH, and SV data is generated and visualized using different platforms. In this presentation, we describe how all of these different data modalities can be brought together to provide a much-improved clinical utility of genetic testing over a single test approach.
47. “n=1” cases in GeneMatcher: Reporting alterations in truly novel disease genes

Deepali N. Shinde¹, Taylor Cain¹, Wendy Alcaraz¹, Zöe Powis¹

Ambry Genetics, Inc., Aliso Viejo, CA 92656

Disease gene discovery efforts have gained momentum with the use of whole exome sequencing (WES) for identifying pathogenic alterations in novel candidate genes. These efforts were aided to a large extent by data sharing and collaborations between referring physicians, diagnostic and academic laboratories. Since 2011, we have reported alterations in candidate genes in 6% (131/2223) of patients referred to our laboratory for diagnostic WES (DES). Fully vetted candidate gene scoring criteria were used for evaluating the evidence to support the reporting of such alterations. More than 50% of our reported findings have since been corroborated by publications from independent research groups. In March 2016, we began submitting candidate gene alterations to GeneMatcher and have matched with submissions spanning 97 genes. Matches have not been established yet for alterations in 13 genes for which there is strong evidence from candidate gene scoring criteria. Here we present a case of a stillborn fetus delivered at 27 weeks gestation. Fetal MRI identified brain abnormalities including enlarged ventricles, deficient cerebellum and dysgenesis of the corpus callosum. Additional findings on autopsy included complete absence of muscle tissue, lung hypoplasia, contractures, micrognathia, cleft palate, and dysmorphic features. DES identified compound heterozygous alterations in the candidate gene NMNAT2, c.695G>A (p.R232Q) and c.403dupC (p.Q135Pfs*44). Both alterations were classified as likely pathogenic using Ambry’s variant classification scheme. Candidate gene scoring criteria estimated strong evidence for the clinical association of biallelic alterations in this gene with our patient’s phenotype based on strong phenotypic overlap with the homozygous Nmnat2 knockout mouse. Since no other patients with NMNAT2 alterations had been reported in literature, we submitted our case to GeneMatcher and have found two cases, one of which did not match the phenotype of our patient or the mouse model. We have not established contact with the second case yet. Alterations in such presumably essential genes are usually observed in single families (n=1 cases) and it may take several years to identify another family with alterations in these genes. With more widespread use of data sharing resources such as GeneMatcher, we hope to uncover additional alterations in such genes in order to establish a stronger link to disease for improved diagnoses and patient care in the future.
Exome Sequencing has become a powerful diagnostic healthcare tool, and Expanded Secondary Findings (ESF) offers healthy individuals the opportunity to learn more about their medical risks and carrier status. To ensure the clinical relevance of ESF panel content at a diagnostic laboratory, the list of 1227 genes was updated using 1. recently published clinical validity criteria (Smith, et al., 2017) and 2. comparison with published scores. Genes were originally selected for the ESF panel due to reported evidence for a role in disease in various public databases in 2013. Using revised criteria, genes were ranked by their Mendelian disease association with the highest clinical validity: 654 genes were Definitive, 200 were Strong, 320 were Moderate, 40 were Limited, 5 were Risk Association Alleles, and 8 were Disputed. Generally, only gene-disease associations with a score of Moderate or above are considered a characterized genetic etiology. Therefore, genes scored below Moderate were subsequently removed from the ESF panel, leaving an updated panel of 1175 genes.

The most extensive published list of clinical validity scores for human genes is from the BabySeq Project, based on a preliminary ClinGen clinical validity classification framework (Ceyhan-Birsoy, et al., 2017). To evaluate our concordance with these published scores, the newly-scored gene content of the ESF panel was compared to the curated gene list published for newborn genomic sequencing in BabySeq. Of the 878 genes scored for both the BabySeq project and the ESF panel, 64% (561 genes) had identical clinical validity categories. Discrepant scores were identified for genes for which different diseases had been evaluated, eg HCN4 was Limited for Brugada Syndrome (BabySeq) but Strong for Sick Sinus syndrome (ESF). There were 19 genes with a differing characterization status between lists, including WNK1-Hereditary Sensory Neuropathy, Limited (BabySeq) vs Strong (ESF); and CACNA1S Malignant Hyperthermia, Limited (BabySeq) vs Moderate (ESF). Overall, comparison of clinical validity scores in ESF to BabySeq shows good concordance between curators and scoring systems for matching gene-disease pairs. Ultimately, both clinical validity evaluation and comparison to published guidelines contribute to selection of genes that are most appropriate to report, but the specific disease used for analysis and reporting must be carefully curated to ensure accurate medical reporting for unaffected individuals.

REFERENCES


GeneMatcher, a review of the database, its connections to other databases and the users’ comments.

N. Sobreira¹, F. Schiettecatte², D. Valle¹, A. Hamosh¹ ¹) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine (JHUSOM), Baltimore, MD 21205; ²) FS Consulting, Salem, MA

In clinical and research sequencing centers, the causal gene cannot be identified in more than half of the individuals sequenced. One of the reasons for this relatively low yield is inadequate communication between clinicians and basic scientists with knowledge of particular genes, proteins or biological systems. To facilitate such communication and improve the search for patients or model organisms with variants in specific candidate genes we have also been adding capabilities to GeneMatcher (www.genematcher.org).

In GeneMatcher matches can be made based on gene name, OMIM® number, genomic location and/or on phenotypic features. Recently, we have also started to allow patient participation in the database. As of today, we have 26 patients participating in the database. GeneMatcher is also part of the Matchmaker Exchange (MME) (http://matchmakerexchange.org/), which has developed an Application Programing Interface that was implemented in August 2015 and allows GeneMatcher users to submit their data to query PhenomeCentral, DECIPHER, MyGene2, Broad Matchbox and the Australian Genomics Health Alliance. As of April 1st 2017, 3,170 individuals from 63 countries have created an account in GeneMatcher and 6,578 genes were submitted. The 18,039 matches involving 2,506 genes have enabled collaborations and the description of novel Mendelian phenotypes and novel Mendelian genes such as SPATA5, HNRNPK, TELO2, RSPRY1, HIVEP2, CHAMP1 and others described in more than 30 publications. A recent user survey showed that 86 users out of 145 responding have identified a novel gene/phenotype association by using GeneMatcher and 42 have published at least 1 paper that included data acquired from a match in GeneMatcher. Seventy-five of 140 users identified themselves as health care providers and 107 as researchers. Forty-eight health care providers said they identified the responsible gene for at least one family under their care by using GeneMatcher and 7 of these families had been under diagnostic investigation for more than 9 years. As of April 28th 2017, OMIM described 3,734 genes with phenotype-causing mutation leaving >15,000 genes without a phenotype association and now more than 4000 of them are in GeneMatcher as a candidate gene for a phenotype and about 60% of them have not had a match yet. In addition, 2,225 users have not had a match yet. These numbers suggest numerous possibilities for upcoming matches that will uncover novel Mendelian genes and phenotypes.
The dilemma of categorizing and classifying scientific entities has been debated for centuries, and throughout this time several entities have cycled through phenotypic lumping and splitting as scientific knowledge and technologies have advanced. In medical genetics, nosology — or the classification of disease — has historically approached classifying disorders based on phenotypes, either lumping similar phenotypes into a broader disease entity (i.e. syndrome) or splitting out isolated phenotypes. However, with progressive advancements in medical genetics and genomics, nosology has been turned on end. With next-generation and whole exome sequencing technologies, we are now able to identify the underlying genetic etiologies of diseases that may have previously been “lumped” based on phenotypic similarities; these etiologies are often distinct, and systematic reclassification and re-categorization of disease entities may be necessary. The Clinical Genome Resource (ClinGen) is a resource that defines the clinical relevance of genes and variants for use in precision medicine, and has developed a system to classify the strength of a gene:disease relationship through biocuration. One of many challenges faced by biocurators is how to curate genes associated with several phenotypic entities: when to lump or when to split a gene:disease curation given the evidence present in the literature? To assist biocurators and establish consistency within the consortium, ClinGen has formed a task group to assemble criteria to address the lumping and splitting predicament. Our preliminary criteria include: (1) defining the disease entity(ies) for a particular gene:disease relationship; (2) establishing the molecular mechanism(s) underlying the associated entities; (3) discerning the phenotype and variable expressivity (including interfamilial and intrafamilial differences). We also formulated a pre-curation process that allows for an initial assessment of the associated disease entities of a gene. As part of this process, biocurators and domain experts would preliminarily bin genes based on one of three possible phenotypic etiologies: isolated phenotype, complex phenotype, or syndromic phenotype. This pre-curation will ultimately facilitate and streamline lumping and splitting based on our criteria. Here, we will outline the pre-curation process, our guidelines, as well as specific examples to illustrate the process of determining when to “lump” and when to “split.”
Challenges and solutions in implementing the ACMG/AMP variant classification guidelines in a large reference laboratory

Scott Topper, Keith Nykamp, Michael Anderson, John Garcia, Blanca Herrera, Yuan-yuan Ho, Yuya Kobayashi, Nila Patil, Janita Thusberg, Marjorie Westbrook

The number of variants requiring classification has grown dramatically in the NGS era, due to the routine analysis of larger parts of the genome and to a massive increase in test volume from an increasingly diverse patient population. The intellectual and operational challenges of accurately and consistently classifying variants are felt acutely in high-throughput reference laboratories, where these challenges must be met by a large team and a supervised distribution of labor. New demands are put upon laboratory directors to ensure the clinical validity of their test results and to maintain effective and meaningful supervision of their teams.

To support this, we aimed to develop a rigorous, comprehensive and fully auditable variant classification protocol, derived from the ACMG/AMP’s 2015 Interpretation of Sequence Variants guideline. We implemented this framework into our clinical testing workflow and established a working group to evaluate its effectiveness and to refine our procedures on an ongoing basis. Through the classification of thousands of variants, we challenged the ISV guideline, identifying cases for which strict adherence to the framework led to what we perceived to be poorly supported classifications, or when uncertainty or disagreement arose about the correct application of the rules. We worked to identify the genetic or conceptual issues underlying the discrepancies, and to refine the criteria, their valuations and their inter-dependencies. The resultant framework, “Sherloc”, is a comprehensive refinement of the ACMG-AMP guidelines. We consider ongoing development of this protocol to be an essential part of laboratory process quality improvement.

We will describe our translation of the ISV guideline to a point-based system which recognizes a hierarchical relationship between evidence types, and which includes a dramatic expansion of the ISV rule set to establish discrete criteria for related but subtly different genetic data. We include protections against double counting evidence, heuristics for evaluating the strength of certain evidence types, and the development of decision trees to support consistent application of the rule set. We also describe an efficient classification workflow and details of our software implementation to ensure auditability and to provide a platform for refinement of classifications over time. A detailed description of these methods is currently in press, and will be available to the ClinGen community.
Structured gene-condition information to support accurate clinical variant classification

Scott Topper, Michelle Hogue, Rita Quintana, Carolina Pardo, Yuya Kobayashi, Keith Nykamp, Michael Anderson, John Garcia, Marjorie Westbrook, Michelle Sommargren, Emily Hare

The interpretation of genetic variants depends on a rigorous approach to evaluating evidence and a body of specific knowledge about the genes in which the variants are found. This contextual knowledge includes a set of functional and clinical considerations that directly inform how variant-specific data are evaluated. Although invaluable collaborative resources exist to aggregate this information (OMIM, Medgen, GeneReviews), a critical and structured approach to the evaluation of this data is a core responsibility of practitioners of laboratory medicine, especially as the breadth of genes and diseases under consideration increases. We present here a minimal set of core information, systematic heuristics for evaluating the underlying data, and a structured approach to capturing complex gene-mechanism-disease information. These include:

- A rigorous method for evaluating the hypothesis that loss-of-function is an established mechanism of disease; essential for the interpretation of premature termination variants.
- A categorical heuristic for the assessment of penetrance, severity and age-of-onset, and the development of a gestalt quality based on these considerations; important for making effective use of case-report and population data.
- An evaluation of the severity of the complete null phenotype, required for consistent evaluation of homozygous and hemizygous data from databases and case reports.
- A simple method for the evaluation of the strength of the gene-condition relationship that is necessarily harmonized with the structured approach to variant classification; important for the effective communication of results.

The data structure employed captures the complex, many-to-many relationships between genes and conditions, and the sometimes fluid and evolving nature of these relationships. Its basic unit is tripartite, and comprised of a “gene” object and a “condition” object that are linked by a “gene-related condition details” object. The “gene” object contains information about the molecular definition of the analyzed region, technical issues that affect analytic sensitivity, and clinical considerations applied to all variants in the gene regardless of specific clinical correlations. The condition object contains the clinical descriptions and associated phenotypic features of the general clinical condition, the range of clinical subtypes, and epidemiological information. The gene–condition details object contains an evaluation of the strength of the gene–condition relationship, molecular details of the gene–condition relationship, and clinical details of the gene-specific condition. This structure supports complex networks of many genes and many conditions and provides a rigorous foundation for consistent variant classification.
53. Adding Value to the ClinVar Database through Non-Laboratory Clinical Provider Variant Submissions

Karen E. Wain¹, Thomas D. Challman¹, Brenda Finucane¹, Andrea Seeley², Christa Lese Martin¹

Geisinger Health System, Autism & Developmental Medicine Institute, Danville, PA¹; Geisinger Health System, Department of Pediatrics, Danville, PA²

The availability of clinical exome and genome sequencing tests has reinforced the need for data-sharing and curation efforts to improve the interpretation of clinically-reported genomic variants. NCBI’s ClinVar database facilitates these efforts as a repository for clinical assertions about genomic variants’ associations with disease and for the evidence supporting those assertions. Most variants are submitted from clinical laboratories, which may lack appreciation of a patient’s clinical presentation or follow-up data obtained after a variant is reported. Therefore, clinical providers can contribute to ClinVar by submitting variants with their own assertions and supporting evidence. The Geisinger Autism & Developmental Medicine Institute genetics team routinely reviews the clinical significance of all variants obtained through clinical genetic testing, using currently published ACMG guidelines. We have submitted our interpretations of three sequence variants previously submitted to ClinVar and two variants which were not yet submitted as examples of the value of clinical provider ClinVar submissions. To our knowledge, this is the first such submission from a neurodevelopmental clinic. Five variants were identified by whole exome sequencing in five patients with developmental or neurological phenotypes, including two compound heterozygote siblings. Four variants were originally reported as likely pathogenic (SCNA11A, c.1187T>C; SALL1, c.1108_1109delGT; SGSH, c.1063G>A; SGSH, c.673T>C) and one was a variant of uncertain significance (OPA1, c.113_130del18). All four likely pathogenic variants were upgraded to pathogenic by our clinical assessment and the OPA1 variant remained as uncertain. Several factors, unavailable at the time of testing, contributed to our assertions: clinical correlation in a proband or relative, familial segregation data, subsequent laboratory testing, and research-based functional studies. Additionally, literature-based data was assessed that was not specifically detailed in one clinical report. Our early experience demonstrates that clinical provider ClinVar submissions are valuable contributions to curation efforts by adding new variants to ClinVar, providing clinical expertise and clinically-nuanced phenotype descriptions, and outlining additional supporting evidence. Given our preliminary success, we are developing resources as part of the ClinGen project to facilitate variant and phenotypic submissions by clinical providers.
54. Defining the pediatric actionability of genetic conditions for utility in newborn screening

Kathleen Wallace, Stephanie Crowley, Daniela DeCristo, Kate Foreman, Laura Milko, Lonna Mollison, Julianne O'Daniel, Bradford Powell, Cynthia Powell, Jonathan Berg

Department of Genetics, Department of Pediatrics and Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Newborn screening (NBS) has significantly improved public health outcomes by ascertaining medically intervenable conditions before symptoms present. Next-generation sequencing (NGS) has the potential to dramatically increase the number of genetic conditions detected presymptomatically through NBS. Several challenges exist, however, including knowledge gaps, technical logistics, and ethical, legal and social issues (ELSI).

The North Carolina Newborn Exome Sequencing for Universal Screening (NC NEXUS) project is investigating NGS as a potential component of NBS. For the project, we have implemented a modified version of a previously published semi-quantitative metric (SQM) (Berg et al., 2016), with an additional age-based metric, to determine the pediatric actionability of curated gene-condition pairs and subsequently “bin” those pairs into age-actionability categories. The updated SQM consists of five quantitative criteria including severity, likelihood, efficacy, acceptability and knowledge base as well as categorized metrics of age of onset or intervention. As an ongoing process, we compare the results of our SQM with protocols and gene lists produced by the ClinGen Actionability Working Group, the BabySeq project (Ceyhan-Birsoy 2017), and the Recommended Universal Screening Panel.

Currently, 736 gene-condition pairs have been curated and assigned to one of four categories in the first iteration of our curated gene-conditions list. The core ‘NGS-NBS’ panel consists of 450 gene-condition pairs that have been determined to be actionable in childhood and will be evaluated in all study participants. Parents will be randomly assigned to a control group or to a decision group where they will be asked to decide whether to also receive results for non-medically actionable, childhood-onset conditions (240 gene-conditions pairs) and/or medically-actionability, adult-onset conditions (25 gene-disease pairs). Adult-onset non-medically actionable conditions will not be evaluated in any participants. The application of the semi-quantitative metric to categorize NGS results will help to simplify parental decision-making and allow us to better understand the impact of applying NGS to newborn screening.
A quantitative and disease- and gene- specific approach to variant interpretation improves the yield of genetic testing

Roddy Walsh\textsuperscript{1,2}, Nicky Whiffin\textsuperscript{1,2,3}, Francesco Mazzarotto\textsuperscript{1}, Risha Govind\textsuperscript{1,2}, Alicja E Wilk\textsuperscript{1,2}, William Midwinter\textsuperscript{1,2}, Rachel Buchan\textsuperscript{1,2}, Paul JR Barton\textsuperscript{1,2}, James S Ware\textsuperscript{1,2,3}, Stuart A Cook\textsuperscript{1,3,4,5}

1. National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK
2. Cardiovascular Research Centre, Royal Brompton and Harefield NHS Foundation Trust and Imperial College London, London SW3 6NP, UK
3. MRC London Institute of Medical Sciences, Imperial College London, London W12 0NN, UK
4. National Heart Research Institute Singapore, National Heart Centre Singapore, 168752, Singapore
5. Division of Cardiovascular & Metabolic Disorders, Duke-National University of Singapore, 169857, Singapore

Background: The accurate interpretation of variants is a critical and often limiting step in the adoption of genetic testing for Mendelian diseases. Guidelines proposed by the American College of Medical Genetics aim to standardise methods for variant assessment in clinical genetic testing and avoid the preponderance of false positive associations that blight the research literature. However these guidelines are by necessity both a generic approach for all Mendelian diseases and conservative in assigning pathogenicity, particularly for missense variants that require segregation data in tested or previously published pedigrees to confirm pathogenicity.

Methods: Here, focusing on genes associated with hypertrophic cardiomyopathy (HCM), we use large sequencing datasets to identify variant classes with a high probability of pathogenicity, allowing us to increase the yield from genetic testing even when limited evidence is available for the specific detected variants. By utilising strict thresholds on population frequency and comparing the frequency of such rare variants in population (ExAC) and large HCM cohorts, we can define the etiological fraction (EF) as a measure of the prior likelihood of pathogenicity (dependent on gene and variant type).

Results: While truncating variants in MYBPC3 (responsible for about 10% of HCM cases) have a high and informative EF of $>0.99$, the EF of the more common non-truncating variants ranges from 0.70 to 0.95. These variants can however be further discriminated through methods such as functional prediction algorithms or the clustering of pathogenic variants in specific gene regions. We demonstrate that for genes where clustering of pathogenic variants is observed (MYH7, TNNT2, TNNI3, FLNC), this is a significantly better discriminator than even consensus predictions from functional algorithms. An EF of $\geq0.95$ was identified for variant classes that account for around 70% of causative variants in HCM - if this threshold was deemed sufficient evidence for a likely pathogenic classification (in the absence of other supporting evidence), the yield for HCM clinical genetic testing would be substantially increased.

Conclusion: In summary, we demonstrate a quantitative approach to variant interpretation that can identify variant classes with a high prior probability of pathogenicity thus increasing the yield of genetic testing and highlight the importance of developing disease and gene specific approaches to variant assessment.
CardioClassifier – semi-automated and interactive disease-specific variant interpretation according to the ACMG/AMP guidelines

Nicola Whiffin1,2,3, Roddy Walsh2,3, Elizabeth Edwards2,3, Risha Govind2,3, Mian Ahmad2,3, Matthew Edwards3, Xiaolei Zhang2,3, Rachel Buchan2,3, William Midwinter2,3, Hanna Najgebauer2,3, Alicja E Wilk2,3, Deborah J Morris-Rosendahl2,3, Paul JR Barton2,3, James S Ware1,2,3 and Stuart A Cook1,2,3

1. London Institute of Medical Sciences, Medical Research Council, London, UK
2. National Heart and Lung Institute, Imperial College London, UK
3. Royal Brompton and Harefield NHS Foundation Trust, London, UK

The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) consensus guidelines for variant interpretation outline a set of criteria to assess a variant against and the weighting that each should be given. These guidelines are intentionally broad in scope, to allow adoption across a spectrum of different disease phenotypes. The framework requires substantial additional expert refinement before it can be implemented for a specific disease or gene.

We have created CardioClassifier, a semi-automated web-interface to support consistent and rapid variant interpretation according to the ACMG/AMP guidelines, for a range of inherited cardiac conditions (ICCs). CardioClassifier draws upon deeply annotated and expertly curated disease- and gene-specific knowledge, in addition to multiple large pre-compiled data-sources, to automatically assign ACMG/AMP rules relating to 17 computational data-types. Results are displayed on an interactive web-interface that allows users to refine the classification and add additional case-level and functional data.

We have tested CardioClassifier on a series of gold-standard orthogonally curated variants and show excellent concordance with manually curated interpretations. Comparison of CardioClassifier to a non-disease-specific tool showed markedly improved performance; with more than twice as many pathogenic variants correctly identified as clinically actionable (i.e. Pathogenic or Likely Pathogenic; 71.2% versus 29.2%; n=219), using computational data alone. Finally, the rate of potentially clinically relevant variants (those identified as Pathogenic or Likely Pathogenic in addition to the case rate of VUSs over the rate in healthy volunteers) identified by CardioClassifier in 327 hypertrophic cardiomyopathy (HCM) cases, at 33.7%, is similar to previous reports.

To further increase the utility of CardioClassifier, we have developed a ‘knowledge base’ of manually curated annotations for 120 variants associated with the most common cardiomyopathies (HCM, DCM and ARVC). In addition, we intend to continuously update and adapt CardioClassifier to incorporate new knowledge and gene-disease specific guidance, as it becomes available. This includes harmonising our rule parameterisation with recommendations released by the ClinGen Cardiovascular Domain Working Group.
The American College of Medical Genetics and Genomics (ACMG) has published recommendations for reporting secondary findings from clinical genome and exome sequencing. This list recommends 59 genes to be analyzed for disease-causing variants and returned to patients who do not opt out of receiving these findings. In parallel with the efforts of the ACMG, ClinGen’s Actionability Working Group (AWG) has developed a framework to create evidence-based summary reports and apply a semi-quantitative metric to systematically assess the clinical actionability of gene-disorder pairs by evaluating potential interventions that could be undertaken in the context of secondary findings. Domains considered by the AWG include: 1) the nature of the health threat; 2) likelihood (penetrance); 3) effectiveness of interventions to prevent harm; and 4) the risk/burden of the intervention to the individual. To date the AWG has applied this process to 122 gene-disorder pairs (including the 59 ACMG recommended genes) and scored 284 outcome-specific interventions. The distribution of scores associated with genes included in the ACMG recommendations is higher for these domains of actionability compared with scores associated with genes not included in the ACMG recommendations. Nevertheless, the AWG framework has identified additional gene-disorder pairs not currently included in the ACMG recommendations that receive high scores across multiple domains of clinical actionability. Across all topics evaluated by the AWG, common features of situations with higher actionability scores include cardiovascular conditions that can lead to sudden death, and disorders for which non-invasive imaging or monitoring effectively reduces morbidity or mortality. Factors that frequently negatively impact actionability scores include limited documentation of the likelihood of the condition or effectiveness of the intervention, or interventions that are burdensome or invasive. In addition to evaluating findings in unaffected individuals, a subset of topics have been identified that may guide accurate diagnosis of individuals with clinical signs or symptoms (e.g., identification of a HNF1A variant in an individual previously diagnosed with type II diabetes mellitus). We will provide illustrative examples of the spectrum of scores that are encountered by the AWG. The continued work by the AWG may help to identify topics that could be considered by stakeholders for return as secondary findings.

Source of Funding: NCI Contract No. HHSN261200800001E
Increased sharing of clinically interpreted variants has revealed a significant degree of discordance in variant classification. The joint ACMG/AMP recommendations for variant interpretation have made considerable progress towards standardization, but apply to a wide range of Mendelian scenarios. In order to adapt these guidelines to specific disease areas, ClinGen has formed working groups with expertise in specific clinical domains. The Cardiovascular Domain Working Group (CDWG) established a ClinGen-approved Expert Panel for inherited cardiomyopathies (CMP-EP), which selected a well-established gene for inherited cardiomyopathies, myosin heavy chain 7 (MYH7), to develop a framework for adapting these guidelines to inherited cardiomyopathies. The CMP-EP finalized its framework by interacting with ClinGen’s processes and oversight mechanisms to ensure harmonization of approaches taken by various Expert Panels.

CMP-EP review of the ACMG/AMP criteria yielded consensus revisions specific to MYH7-associated inherited cardiomyopathies. Of the original ACMG/AMP criteria, 32% (9/28) were deemed not applicable to MYH7 and 39% (11/28) required gene- and/or disease-specific adjustments. Additional modifications reflected implementation of quantitative frameworks for benign frequency thresholds and adding additional evidence weight layers to rules reflecting segregation and case-control data. To test the revisions, two expert reviewers independently scored 60 MYH7 variants, which yielded a concordance of 93% (56/60) with 3% (2/60) representing discrepancies that may impact medical management. Internal clinical and laboratory data from >6,000 probands impacted rule application in 24/60 variants (40%) and changed the classification of 10/24 variants (42%), half of which would impact patient management. Despite specific adjustments, clinical judgment remained important and was used to override classifications for 2 variants. Interaction with the Sequence Variant Interpretation (SVI) working group led to final adjustments ensuring compatibility with the intent of the parent rule framework as well as across other expert panel efforts.

This work demonstrates that expert-driven adaptation of the ACMG/AMP criteria improves its specificity but also highlights remaining challenges. Use errors usually reflect complexities of highly heterogeneous diseases, emphasizing the importance of expert involvement in this process.