

# Pathogenicity Interpretation in the Age of Precision Medicine: The 2015 Annual Scientific Meeting of the Human Genome Variation Society

William S. Oetting,<sup>1\*</sup> Steven E. Brenner,<sup>2</sup> Anthony J. Brookes,<sup>3</sup> Marc S. Greenblatt,<sup>4</sup> Reece K. Hart,<sup>5</sup> Rachel Karchin,<sup>6</sup> Shamil R. Sunyaev,<sup>7</sup> and Peter E. Taschner<sup>8,9</sup>

<sup>1</sup>Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, Minnesota; <sup>2</sup>Department of Plant and Microbial Biology, University of California, Berkeley, California; <sup>3</sup>University of Leicester, Leicester, United Kingdom; <sup>4</sup>Department of Medicine, University of Vermont, Burlington, Vermont; <sup>5</sup>Invitae, San Francisco, California; <sup>6</sup>Departments of Biomedical Engineering/Oncology and Institute for Computational Medicine, Johns Hopkins University, Baltimore, Maryland; <sup>7</sup>Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; <sup>8</sup>Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; <sup>9</sup>Generade Centre of Expertise Genomics and University of Applied Sciences, Leiden, The Netherlands

Received 10 December 2015; accepted revised manuscript 10 January 2016.

Published online 21 January 2016 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22958

Hum Mutat 37:406–411, 2016. © 2016 Wiley Periodicals, Inc.

## Introduction

There is now a convergence of two modes of genetic testing, that of testing a few candidate genes at a time based on suspicion of a specific genetic disease, and that of genomic testing, especially when a candidate gene(s) is not suspected or known. Both aim to interpret pathogenicity of identified genetic variants. The 2015 annual scientific meeting of the Human Genome Variation Society (HGVS; <http://www.hgvs.org>) was held on the 6th of October in Baltimore, MD, with the theme of “Pathogenicity Interpretation in the Age of Precision Medicine.” The HGVS is focusing attention on advancing the field of variant interpretation. Progress will require both new developments in analytical methods (in vitro, in silico, statistical, and other methods) and cooperation among scientific, clinical, and regulatory participants in developing and maintaining standards for all areas of pathogenicity assessment, variant nomenclature, and annotation. This year's meeting covered all of these areas.

## In Remembrance of Professor Richard G. H. Cotton, BAgSc, PhD, DSc, AM

The HGVS annual meeting was opened by Marc Greenblatt of the University of Vermont, beginning with a remembrance of Dr. Richard G. H. Cotton of the Genomic Disorders Research Centre, University of Melbourne and St. Vincent's Hospital, Melbourne, Australia, who passed away suddenly from a stroke in June, 2015. Professor Cotton was the driving force behind the creation of the HGVS and served as its founding President (2001–2008). He then led the creation of the Human Variome Project (HVP; <http://www.humanvariomeproject.org>), which continued his efforts, spanning more than two decades, to better understand genetic

variation and its impact on human health. His early research was instrumental for many scientific advancements, including the development of mutation detection techniques and the planning and execution of the fundamental experiments that led to the production of monoclonal antibodies. He worked tirelessly to bring together individuals of similar interest to facilitate the collection and study of genetic variation data to improve human health. Professor Cotton was influential in the professional development of scores of people in the genetics world, through the uncommonly special combination of sharp intellect, warm unpretentious personality, and high integrity. It is through the combination of these attributes that he accomplished so much, and he will be greatly missed.

## Computational Methods for the Functional Interpretation of Variants

The first scientific session was chaired by Marc Greenblatt. For the first talk, Carlos Bustamante of the Department of Genetics, Stanford University, spoke on “Development of robust pathogenicity predictors and functional validation.” Pathogenicity prediction for genetic variants is still an imperfect science. There is a significant amount of variation in the accuracy between different prediction methods of pathogenicity, though machine learning methods work better than individual prediction models. In the case of cystic fibrosis (CF), thousands of pathogenic variants within the coding region of the CFTR protein have been identified in CF patients in multiple populations, but only a small percentage of these variants have been unambiguously shown to be functional. There are multiple mechanisms of pathogenicity for a genetic variant such as reduced protein stability or altered function, increasing the difficulty of determining pathogenicity. Solely protein structural-based predictors do not seem to be the answer. To create more accurate algorithms, high-quality training sets of known functional and benign variants need to be created to test and improve programs that can predict the functional consequence of a variant. Ultimately, these algorithms may need to be gene specific to accurately predict pathogenicity. A computational tool trained to predict pathogenicity for CFTR missense variants was presented. Features considered by the computational tool include the outputs from available pathogenicity prediction algorithms (based on evolutionary conservation, genomic context, etc.) supplemented with protein structure and allele frequency data.

\*Correspondence to: William S. Oetting, Department of Experimental and Clinical Pharmacology, 7-115 Weaver-Densford Hall, 308 Harvard Street S.E., Minneapolis, MN 55455. E-mail: oetti001@umn.edu

Although this tool is more accurate than existing predictors alone, a significant question is whether it can be improved by including functional measurements made *in vitro*. Work is now underway to perform functional validation of variants in *CFTR*, which will be added to the predictor as features to increase its accuracy.

The next talk was by Rachel Karchin of the Johns Hopkins Institute of Computational Medicine and Biomedical Engineering/Oncology faculty, who spoke on “Towards increasing the clinical relevance of computational methods to predict the consequence of human genetic variation.” *In silico* prediction of variant pathogenicity has the potential to be clinically important but there are weaknesses in the current paradigm. Independent benchmarking of current methods reveals that most do not reliably reproduce their reported sensitivities and specificities, resulting in clinicians remaining skeptical of their utility. Most predictive models oversimplify variant analysis by limiting variants to two classes, normal or disease producing. But the disease impact of variants is not “yes or no” but exists on a continuum. The disease phenotypes to be predicted exist several steps removed from the specific effect of a single genetic variant. Analysis may be more effective by focusing on endophenotypes, which are measurable components unseen by the unaided eye along the pathway between disease and distal genotype. There are several advantages of using endophenotypes for *in silico* pathogenicity variant prediction including: (1) they are closer to the gene/protein function than more complex disease phenotypes, and (2) the effect size of variants are more directly measurable. CF was presented as one example where the effect of a variant on specific CFTR protein function (endophenotype) was more accurately predicted using *in silico* prediction models than the presence or absence of CF itself.

Matthew D. Rasmussen, Director of Software Engineering at Counsyl, continued on the topic of *in silico* analysis for predicting functionality of variants in his talk “Semi-supervised learning for clinical variant interpretation.” Interpretation of variants is becoming the most important part of large-scale DNA sequence interpretation but is not scalable to the same degree that DNA sequencing is. New computational methods promise to remove this bottleneck. Semisupervised clustering of mutations is a high-performance method for classifying novel missense variants. In creating the method, a training set with known functional variants was created. It was felt that nonsense and frameshift variants led to overfitting the data and were excluded. Instead, a training set of variants including those with a high MAF as benign variants and a number of simulated variants as unlabeled variants were used. Once the model was created, it was validated using a set of known variants. It was found that the method worked better than other prediction methods including SIFT and PolyPhen2. The source code can be found at <https://github.com/counsyl/sscm>.

Comparison of variant impact prediction models is difficult when individuals evaluate their methods with different test sets of variants. For reliable evaluation, a set of variants created independently of any single laboratory experiment is needed, in which definitive answers have been determined but are unavailable to predictors. This has been the approach of the Critical Assessment of Genome Interpretation (CAGI). Steven Brenner of the Department of Plant and Microbial Biology, University of California, Berkeley, gave an update on CAGI in his talk “Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction.” The fourth international experiment is now in progress and will run through December 2015. One test set from previous years consisted of 84 experimentally tested variants within the cystathionine-beta-synthase (*CBS*) gene to be used as a test for model comparison. This and other test sets can be

found at <https://genomeinterpretation.org>. The CAGI results will be presented in March of 2016.

Steven Brenner also presented on the “Diagnostic Role of Exome Sequencing in Immune Deficiency Disorders.” A protocol was created to study individuals with undiagnosed immune disorders using whole-exome sequencing (WES). In one example, two unrelated infant girls had screened positive for severe combined immunodeficiency (SCID), but actually had some T-cells and no variants in known SCID-related genes. After WES, the functional variant was identified in the ataxia telangiectasia (*ATM*) gene, which can have an immune phenotype. In a second example, Brenner showed how a custom analysis pipeline identified a variant that proved to be the first heterozygous cause of SCID. These cases show the importance of WES in cases with no obvious diagnosis. Individuals who use WES as a clinical diagnostic test must be prepared to accept unusual and unexpected results as shown in the above two examples.

As stated previously, an important resource for creating, testing, and validating variant prediction algorithms is a test set of known functional variants. To this end, Gary Saunders, from EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, presented “The European Variation Archive: assessing data quality, integration with LOVD and addition of clinically relevant data” as a resource of annotated variants. The European Variation Archive (EVA) is an open-access database of all types of genetic variation data from multiple species. For human variants, the EVA contains over 250 human variant call files (VCF) that describe in excess of 125 million unique variants, from over 100,000 samples. Variants stored at EVA are annotated using a variety of methods, including the Variant Effect Predictor from Ensembl (<http://www.ensembl.org/info/docs/tools/vep>), statistics calculated both within and between studies, and EVA provided tools to allow users to mine these data, filtering on any combination of species, methodology, variant type, phenotype, consequence, or allele frequency. Results from queries can be downloaded in a variety of formats including VCF and CSV. Additionally, EVA provides a comprehensive RESTful Web service, to allow programmatic access, and hence the integration of these data with other resources such as Ensembl (<http://www.ensembl.org>) and Uniprot. In addition, and in collaboration with ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar>), EVA also provides access to the largest collection of open-access clinically relevant variants available, worldwide. The EVA resource can be found at <http://www.ebi.ac.uk/eva>.

The proper description of genetic variants is becoming increasingly important in the age of next-generation sequencing. Peter E. Taschner from the Genade Center of Expertise Genomics in Leiden, The Netherlands, talked about this in his presentation “Description Extractor: Automated HGVS-recommended sequence variant description.” To properly describe variants, you need to understand the current nomenclature. This can be difficult because some variants are complex, especially if multiple variants are being called in phase to create haplotypes or when rearrangements have occurred. The Description Extractor is part of the open source Mutalyzer suite of programs that will name variants using HGVS-recommended nomenclature and is recommended for variant description validation by several scientific journals including *Human Mutation*. It creates relatively small HGVS descriptions from large DNA sequences, even in the formation of haplotypes. It is efficient in terms of computational speed even when comparing whole chromosome sequences. Curators can use it to compare changes between different reference sequences. The Description Extractor can be found at <https://mutalyzer.nl/description-extractor>.

The final talk of this session was by Dan Richards, Vice-President of Biomedical Informatics, QIAGEN Bioinformatics, Redwood City,

CA, who spoke on “Genome-scale ACMG pathogenicity classification using comprehensive curated clinical evidence and data.” Accurate classification of variants is critical to move patient variants into the clinic. The American College of Medical Genetics (ACMG) and the Association for Molecular Pathology (AMP) have established standards for the assessment of clinically relevant genetic variants. QIAGEN Clinical Insight has created a database of over 10 million curated biomedical variants to help in identifying the level of functionality of disease-associated variants. The data were annotated using published and private biomedical knowledge, the Ingenuity knowledge database, and other sources of information associated with variants to classify them as either benign, likely benign, uncertain significance, likely pathogenic, and pathogenic as stated in ACMG guidelines. Currently, the application can be used in oncology for somatic and hereditary cancer testing.

## Genetic Variant Nomenclature

The second scientific session, chaired by Steven Brenner, focused on variant nomenclature. Next-generation sequencing methods can identify millions of variants in a single genomic sequence. A shared nomenclature for describing these variants is critical, especially when archiving variants into a database that may be accessed for many decades or when variants are combined from different studies for analysis. The session was organized by the HGVS in order to engage the community in a discussion about variant nomenclature in general, and the HGVS Nomenclature Recommendations (<http://www.hgvs.org/mutnomen/>) specifically. The session began with presentations by Johan den Dunnen, Deanna Church, and Jean McGowan-Jordan, followed by a panel session with the speakers, as well as Fiona Cunningham, and Reece Hart.

Johan T. den Dunnen of the Leiden University Medical Center, The Netherlands, chair of the HVP sequence variant description working group (<http://goo.gl/pG3zaS>) and coauthor of the initial publication of the nomenclature recommendations, initiated the session by providing a historical perspective of the HGVS recommendations, highlighting key principles, for example, the requirement to specify a reference sequence and discussing some of the challenges in the nomenclature. A core principle of the HGVS recommendations is to encourage the reporting of observed variants rather than derived variants. This principle led to the recommendation that derived variants be reported in parentheses, a proposal that is often ignored in practice and elicited discussion in the panel session. Although the nomenclature has gradually grown into a stable standard used worldwide, there are always ongoing discussions. The working group maintains the standard by generating proposals that address contemporary needs and soliciting feedback from the community (<http://www.hgvs.org/mutnomen/comments.html>). Examples of recent proposals are support for noncoding transcripts and a pending proposal to harmonize variant representations with the International System for Human Cytogenetic Nomenclature (ISCN). In addition, Dr. den Dunnen maintains a Facebook page (<https://www.facebook.com/HGVSmutnomen/>) to respond to proposal feedback and answer general questions about nomenclature usage.

The second speaker in this session was Deanna M. Church of Personalis Inc. in Menlo Park, CA, who spoke on “Variant representation in a world of change.” There are a number of variables that can affect variant naming. For example, using different reference assemblies can alter the name, as shown in the variant rs776746 (CYP3A5\*3) where using the GRCh37 assembly gives the reference allele as “C” but the GRCh38 assembly gives the reference allele

as “T.” Additionally, the sequence can also be affected by how the data is derived. The same variant can be represented in different ways within a VCF. Even after normalization, data within the VCF context can be different compared with HGVS standards. The VCF standard specifies nucleotide numbering being right shifted with respect to the genome, whereas the HGVS standard is left-shifted with respect to the transcript. Not only do we need to know which reference transcript is used because of differences between alternatively spliced transcripts, but we also need to know the exact alignment used as different software can produce different cDNA to genome alignments. Because of this, we need to move beyond pure location-based representations and instead create a naming standard to allow for a world with multiple assemblies/references. We will also need tools that can run locally and quickly to unambiguously correctly name variants using the proper reference sequence and transcript.

Jean McGowan-Jordan, Chair of the International System for Human Cytogenetic Nomenclature (ISCN), from the Department of Genetics at the Children’s Hospital of Eastern Ontario and University of Ottawa, Ottawa, Canada, spoke on “Working towards a combined ISCN and HGVS standard for the description of chromosome rearrangements.” The ISCN is responsible for creating a set of standards to describe chromosomal alterations. Their scope has been expanded to include FISH- and microarray-derived analysis. A standard nomenclature has been created that can describe any chromosome rearrangement identified by standard techniques. There is now a need for nomenclature for chromosomal changes assayed using DNA sequencing. The proposed nomenclature is both ISCN-like and HGVS-like. The ISCN-like portion appears first, beginning with “seq” and includes the genome build. The HGVS-like portion uses the existing HGVS standards along with additional recommendations including autosomes listed first followed by sex chromosomes, multiple breakpoints in one chromosome named consecutively from pter to qter, breakpoint junctions being labeled with a double colon (::), and each element in a separate line. For example, a deletion of a chromosomal region from nucleotide 89555676 to 100352080 on the X chromosome, based on GRCh37 genome assembly (hg19), would be:

```
seq[GRCh37] del(X)(q21.31q22.2)
chrX:g.89555676_100352080del
```

A balanced translocation between chromosome 2 and chromosome 11 would be described as:

```
seq[GRCh37] t(2;11)(p25.1;p15.2)
g.[chr2:pter_8247756::chr11:15825273_cen_qter]
g.[chr11:pter_15825272::chr2:8247757_cen_qter]
```

A duplication for a segment on chromosome 8 would be:

```
seq[GRCh37] dup(8)(q24.21q24.21)
chr8:g.128746677_1287491dup
```

This nomenclature will allow for a more accurate designation of chromosome changes based on DNA sequence data and a more accurate means to compare these changes between patients and their clinical outcomes.

## Genetic Variant Nomenclature Panel Session

These three session speakers were then invited to become part of a panel discussion along with Fiona Cunningham, the Variation Annotation Coordinator of EMBL-EBI and Reece Hart of Invitae. The discussion, entitled “Evolution of nomenclature systems to create a standard that incorporates traditional HGVS nomenclature

and genomic systems” was moderated by Steven Brenner. Fiona Cunningham started off the panel discussion speaking on the importance of using a clearly specified reference sequence and version to remove ambiguity in reporting variants. The existence of different transcript isoforms can often produce many possible descriptors for a single variant and so specifying the transcript and reference sequence version is critical. She argued for the versioning of the recommendations, narrowing the flexibility of the recommendations and enforcing correct usage, especially for publishing variants. She requested a simplification of HGVS nomenclature, removing redundancy, and the extensive flexibility, which exists for historical reasons. Choosing a restricted subset of the current guidelines would result in a simplified set of required components (e.g., either the one of the long: *p.Arg97Glyfs\*26*, *p.Arg97GlyfsTer26*, or the short form *p.Arg97fs* versions).

Reece Hart, research and engineering fellow at Invitae and member of the sequence variant description working group, continued the discussion with three lessons that stemmed from building the HGVS (<https://bitbucket.org/biocommons/hgvs>) and Universal Transcript Archive (UTA; <https://bitbucket.org/biocommons/uta>) packages used in a clinical diagnostics setting. First, variant mapping necessitates access to lots of data that are difficult, and sometimes impossible, to obtain. Projecting genomic variants onto transcripts requires sequence alignments, which are currently implicit and mutable concepts at NCBI, UCSC, and Ensembl. Because NCBI exposes alignments between current transcripts and current assemblies only, readers of papers often find that the data necessary to interpret reported variants are not available. This gap led to the development of the UTA, which stores snapshots of several database sources with alignments to multiple assemblies. The message he presented is that major data sources, particularly the NCBI, need to improve the management and accessibility of data. Second, funding organizations should more strongly encourage collaboration on shared tools in order to improve interoperability and exchange of information. Third, maximizing the value of genomic data requires sharing variants reliably, which necessitates a single representation for variants. Multiple standards significantly impede data sharing. Two “HGVS-like” syntaxes were proposed recently by groups that had not contacted (and may not have been aware of) the sequence variant description committee. The sequence variant description working group needs to modernize by looking to contemporary tools to communicate, educate, document, and solicit feedback from the community so that data standards balance the needs for sharing with needs for innovation.

The session was then opened for discussion. A summary of the major discussion points were as follows:

- HGVS recommendations need modernization in administration, documentation, and code. In particular, the recommendations should adopt the language of specifications in order to minimize ambiguity and generally become stricter. The working group was encouraged to use conventional tools to document, gather feedback, and version the specification.
- There should only be one sequence variant standard nomenclature. The community should discourage splinter nomenclatures.
- Variant nomenclature should strictly require reporting the assayed variant, and permit reporting-derived variants as appropriate. For genome and exome sequencing, this would mean always reporting the assayed variant on a genomic reference and optionally reporting the inferred transcript and protein variant. A vote of the audience on this found nearly universal support. In cases of transcript sequencing, the transcript-based variant should be required and the genomic variant would be optional.

- Inferred variants are an area without consistency. The HGVS standard states that protein variants inferred from transcript variants should be denoted with parentheses. At least three related issues were discussed. First, this rule is poorly adopted in practice. Second, if inference is denoted by parentheses, then that rule should also apply to transcript variants inferred from genomic variants, though this is not clearly recommended. Third, the notion of inference is a kind of annotation that some feel should be kept distinct from a minimal notion of a variant.
- Alignments, particularly those between genome and transcript sequences, should be managed as immutable, citable concepts. The projection of a genomic variant onto a transcript, and the converse operation, requires an alignment between those two sequences. Currently, such alignments are only implied. Because alignments for the same versioned transcript may vary across data sources (e.g., NCBI and UCSC) or even within a single source (e.g., when alignments are updated), the projections between genome and transcript variants may change over time. Dr. Church pointed out that the UTA database provides this for the hgvs package and encouraged a similar effort with government support.
- The community should collectively articulate needs to funding agencies in order to improve standards and a common software suite for variant manipulation. A global, shared facility for immutable sequences, transcripts, and alignments is a specific near-term need.
- HGVS and the working group should do more to educate the community about the complexities of variant representation and manipulation. In particular, many clinical users seem to be unaware of the complexity and uncertainty associated with projecting variants between sequences and do not use correct and unambiguous HGVS syntax.
- An evaluation suite would help the community compare feature differences among packages that manipulate sequence variants. Currently, there are at least six software packages that manipulate HGVS-formatted variants: Mutalyzer, hgvs, pyhgvs, the Variant Effect Predictor, snpEff, and Alamut. An evaluation suite would provide an objective comparison of the features of these tools.
- A variant validation tool would likely decrease errors in variants reported in scientific literature. As envisioned during the session, the community might create a prepublication variant validation tool that journal editors could require authors to use prior to manuscript acceptance.

## Pathogenicity Prediction of Genetic Variants

The third scientific session was chaired by Christophe Bérout of the Aix Marseille University, France. As noted previously, prediction programs for determining the effect of variation on protein function are not always accurate. The best way to determine the functionality of genetic variants is to study the variant protein in a biological system. In the next four talks, functional assays were presented as a way to accurately determine the consequences of genetic variation in disease-associated genes. Harry Ostrer of the Department of Pathology at the Albert Einstein College of Medicine, Bronx, New York spoke on “Predicting the pathogenicity of genetic variants in the DNA double-strand break repair pathway.” There is a great demand for rapid and inexpensive methods that can assess the effects of genetic variants. This becomes increasingly important with the use of panel genetic sequencing where many variants of unknown pathogenicity can be identified. In the case of cancer, there are multiple pathways that when altered, result in neoplasia

including double-stranded break repair (DSB), mismatch repair, cyclin checkpoint proteins, and PI3K/AKT pathways, making a pathway-focused approach a much needed improvement when trying to determine the functional genetic variants associated with a phenotype or disease. In this report, a functional assay for the Fanconi Anemia (FA)-BRCA DSB repair pathway was created. In this analysis, eight *BRCA1* variants, 21 *BRCA2* or FA variants, and 14 benign variants were recreated in cultured cells and analyzed. Cells are initially treated to create DSB, and by using flow cytometry, the levels of different proteins in the DSB pathway are analyzed using labeled antibodies. Nuclear localization of BRCA1 and BRCA2 could also be determined. In this analysis, functionality of each tested variant was accurately determined. Multiple proteins and pathways can be analyzed using several antibodies labeled with different fluorescent tags. It is possible that this method of flow variant analysis can be used on circulating cells from an individual as a test for cancer risk.

The second talk on this topic was by Fergus J. Couch, Department of Pathology, Mayo Clinic, Rochester, MN, who spoke on "Functional assays for assessment of variants of uncertain significance (VUS) in breast and ovarian cancer predisposition genes." Many individuals receiving genetic predisposition gene testing have been found to have VUS in the *BRCA2* gene. Unfortunately, most of these VUS cannot be evaluated for clinical relevance by family-based studies because of limited availability of information from families with these rare variants. In contrast, standardized functional assays may offer a useful option for evaluating VUS. The Couch laboratory has adapted a cell-based direct repeat green fluorescent protein (DR-GFP) reporter assay of homologous recombination (HR) DNA repair for assessment of the impact of VUS on BRCA2 HR activity based on reconstitution of BRCA2-deficient cells with wild-type and mutant forms of BRCA2. In this test, mutated GFP is normally repaired using homologous recombination after generation of a unique restriction enzyme-dependent DNA break. The sensitivity and specificity for known pathogenic variants was found to be 100%. An additional 207 VUS were tested and 79 displayed loss of activity, suggesting that approximately 30% of VUS in the DNA-binding domain of BRCA2 may be pathogenic.

The third talk on this topic was by Nicholas Katsanis of the Center for Human Disease Modeling, Duke University, Durham, NC, entitled "Modeling Structural Defects in Neonates and Young Children." He highlighted the fact that our current ability to interpret the functional effect of genetic variation accurately is modest, especially in the case of complex phenotypes, as well as in establishing a direction of effect. Zebrafish embryos can provide a cost-effective experimental model to assess the pathogenicity of human genetic variation and examples were highlighted in which this model was used to test variants associated with a number of clinical phenotypes, such as gut mobility defects; craniofacial anomalies; cardiac malformations; vascular integrity; glaucoma; renal atrophy/cysts; muscular dystrophy; and changes in head size. This paradigm was used to seed the Taskforce for Neonatal Genomics (TFNG), a synthetic clinical/research unit at Duke University that recruits neonates and young children with suspected genetic defects and models their mutations in zebrafish embryos. The TFNG recruits from nine specialty pediatric clinics were analyzed using WES, followed by systematic evaluation of all candidate pathogenic alleles under a rare, recessive, or de novo paradigm. Aggregate data from the first 100 families indicated a potential diagnostic rate >60%. The last part of the talk focused on alleles predicted to be benign because of evolutionary constraints. Specifically, it reported how in a significant number of

cases, human missense alleles that cause disease can be found in the genomes of other species without apparent pathogenesis. Recent data were presented on the phenomenon of cis-complementation, in which a second rare allele in the same sequence could rescue the pathogenic effect of the primary variant, and thus the evolutionary tolerance for some of these alleles. As a highlight, a case of a child was presented with microcephaly and other complex neurodevelopmental phenotypes, with a de novo missense variant in *BTG2* that was deemed benign because the mutant allele was present in >50 nonprimate vertebrates. Functional studies in zebrafish embryos showed the existence of two compensatory variants in those species that had been lost in humans, thereby exposing the infant to the deleterious consequences of the mutant alleles. It was discussed how this phenomenon can create false negatives in computational predictions and the importance of genomic context on allele impact/function.

The fourth talk on this topic was by Frederick P. Roth of the Donnelly Centre, University of Toronto and Mt Sinai Hospital, Toronto, Canada, who spoke on "Systematically identifying pathogenic human variants using yeast." Sequencing the human genome creates a large set of variants with most having an unknown functionality. The vision of this work is to create a lookup table for every possible missense variant in disease-associated loci before these variants are ever seen in the clinic. Using baker's yeast as a model system provides a very inexpensive way to carry out functional analysis on large pools of mutants. For this model, the first step is to identify homologous yeast genes of human genes associated with disease phenotypes. The next step is to determine whether the human gene rescues the yeast when the yeast homologue is mutated. A temperature sensitive (ts) mutation in the yeast gene is then created and the human rescue gene is mutated with the variant to be tested. The yeast is then grown under permissive and nonpermissive temperatures. Yeast growth under nonpermissive temperatures shows that the mutation is not functional. In testing known functional variants, it was found that the yeast functional assay outperformed in silico prediction programs. Using the power of yeast genetics, large-scale experiments are possible to test all possible variants in a gene. Additionally, this technique can be used to identify somatic variants associated with cancer and may be modifiable to use with cell lines.

A related talk was given by Gérald Le Gac of the Department of Molecular Genetics, University Hospital, Brest, France, entitled "Rare non-synonymous variations in the human ferroportin iron transporter gene (haemochromatosis type 4): the quest for causal mutations." Causal variants of the ferroportin gene (*SLC40A1*) with phenotypic consequence fall into two functional categories (loss- vs. gain-of-function) underlying two distinct clinical entities (Hemochromatosis type 4A, also called Ferroportin Disease, vs. Hemochromatosis 4B). Both result in iron overload. The vast majority of *SLC40A1* alterations are missense variants. Even the most commonly detected *SLC40A1* alleles are relatively uncommon. This situation can pose challenges for the interpretation of molecular testing for patients with iron overload because detection of novel *SLC40A1* variations of unknown significance may occur relatively frequently. Dr. Le Gac presented the results of an integrated approach collecting genetic and phenotypic data from 44 suspected hemochromatosis type 4 patients, with comprehensive protein structure (3D model) and functional annotations. Causality was demonstrated for 10 missense variants, showing a clear dichotomy between the two hemochromatosis-type 4 subtypes.

Two subgroups of loss-of-function variants were distinguished: one impairing cell-surface expression and one altering only iron egress. A new gain-of-function mutation was identified, and the degradation of ferroportin on hepcidin binding was shown to probably depend on the integrity of a large extracellular loop outside of the hepcidin-binding domain. In addition, eight missense variants, shown to be pathogenic in popular computational tools, were found to be benign after functional analysis using this system. A combination of in silico and direct functional analysis is needed to accurately determine the functional effects of genetic variants and their association with disease.

## **Acknowledgments**

The Scientific Program Committee would like to thank Rania Horaitis, Heather Howard, and Timothy Smith for their professional help in running the HGVS Annual Meeting. This year's annual meeting of the HGVS was chaired by Marc S. Greenblatt, Steven E. Brenner, and Christophe Bérout. The authors would like to thank the speakers for their help in the preparation of this report. This meeting was run in partnership with The Human Variome Project. Dr. Brenner's participation is supported by NIH R01 AI105776 and NIH U41 HG007346.