Development of a Comprehensive Gene-/Disease-Specific Analysis and Knowledgebase to Enhance the Efficiency and Accuracy of Sequence Variant Interpretation and Clinical Reporting.

Dandan He, PhD[2], Yingjun Jiang, PhD[2], Yuandong Peng, PhD[2], Yaran Wen, PhD[2], Jun Yang, PhD[2], Ying Zhou, PhD[2], John Lattier, PhD[2], Linyan Meng, PhD[1,2], Xi Luo, PhD[1,2]

1. Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, USA.
2. Baylor Genetics Laboratory, Houston, TX, USA.

P730

INTRODUCTION

With advances in high-throughput genomic sequencing technologies, clinical molecular laboratories are evaluating a rapidly increasing number of novel sequence variants in genes associated with genetic disorders. The evidence and knowledge of gene-disease relationships and variant pathogenicity of the disease-causing genes is also accumulating swiftly. To enhance the efficiency and accuracy of sequence variant interpretation and clinical reporting, a comprehensive gene-/disease-specific analysis and knowledgebase is being developed. Here, the progress on developing the framework of this knowledgebase and its utility will be discussed.

METHOD

The framework of this knowledgebase is composed of four modules: 1) The gene and disease information, for example gene/disease descriptions, mechanisms, penetrance, inheritance patterns, age of onset, and alternative transcripts. 2) Lumping and splitting of the disease phenotypes to define the proper disease for variant curation and reporting in each clinical test. 3) Gene-disease validity curation, focusing on establishing the strength of relationships of gene-disease pairs, utilizing the evidence-based framework developed by ClinGen. 4) Gene-/disease-specific rules for variant curation, which include but not limited to clinical diagnostic criteria and test ranges of biochemical values, population allele frequency cutoffs, validated functional assays, mutational hotspots/coldspots, functional domains and critical motifs, and defining rules and cutoffs for 5' and 3' nonsense, frameshift, or initiation codon variants.

RESULTS

The on-going work of the initial phase of this project aims to evaluate over 500 genes that are frequently seen in diagnostic testing such as carrier screening, prenatal sequencing, hereditary cancer gene panels, and the genes on the ACMG secondary finding list. The disease characteristics and gene features' analysis have been done on 202 gene-disease pairs. 79% of those genes have lumped or single entry of disease phenotypes and 21% of them have split diseases. The gene-disease clinical validity has been established as definitive for 90% of those genes. About three quarter of the genes have Loss-of-Function as the disease mechanism and 12% are Gain-of-Function. Multiple inheritance patterns are included with 46% autosomal recessive and 34% autosomal dominant. Among those genes, we have developed gene-specific variant curation rules for 46 genes which have shown tremendous benefit for the efficiency and accuracy of our daily variant curation activities. For example, the 5' rules define all the start codons and 5' LoF variants as likely pathogenic by default for about 61% of the genes analyzed. The 3' rules for establishing the threshold for likely pathogenic variants are based on nonsense mediated decay prediction, percentage of truncating protein, and the most 3' end well-established P/LP variant.

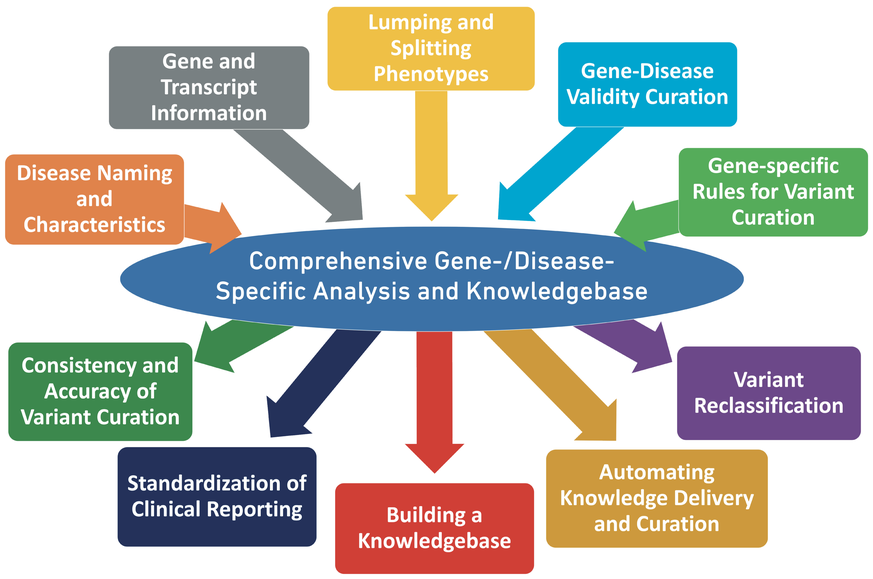

Figure 1. Scope and purpose of the project

Disease Characteristics	Gene Features	Gene-/Disease- Specifications
• Lumping and Splitting Phenotypes • Disease Naming • Disease Description • Gene-Disease Clinical Validity • Disease Mechanism • Inheritance Pattern • Age of onset • Penetrance • Disease Notes	• Gene Name • Alternative Gene Names • Gene Description • Major Transcript • Alternative/Secondary Transcripts • Legacy Codon Naming • Gene-specific Rules for Variant Curation • Gene Notes	• Clinical Instruction ➢ Phenotypic Criteria ➢ Biochemical Ranges • Population Instruction ➢ Allele Frequency Thresholds ➢ Most common P/LP Variant ➢ Penetrance and Age of Onset • Experimental Instruction ➢ Functional Assays and Thresholds • Theoretical Instruction ➢ Rules and Cutoffs for LoF variants at 5' end of the gene ➢ Rules and Cutoffs for LoF variants at 3' end of the gene ➢ Hotspots/Coldspots/Promotor/Domains

Table 1. Key Contents included in the Analysis and Knowledgebase

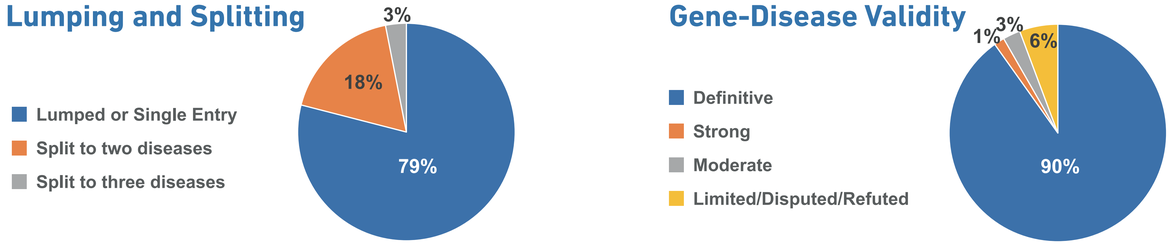

Figure 2. Analysis of gene (n=162) and disease (n=202) features

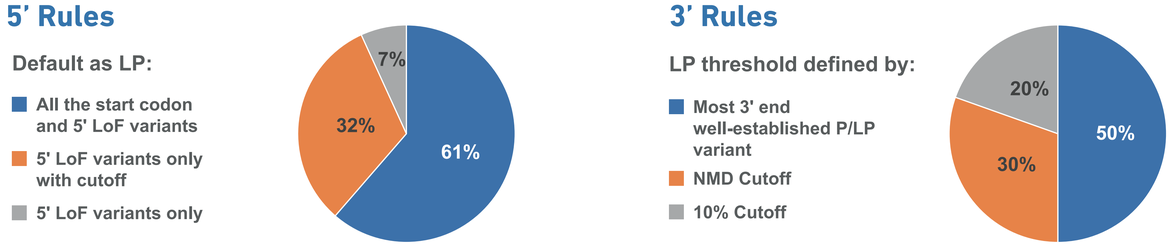

Figure 3. The development of 5' and 3' rules for a pilot of 46 genes

CONCLUSIONS

A comprehensive gene-/disease-specific analysis and knowledgebase to ensure the accuracy, completeness, and accessibility of genetic information is useful for advancing our understanding of genes, their roles in biological processes, and their implications for human disease. Additionally, the goal of this effort may also assist in developing internal automated systems for the efficient delivery of curated knowledge to our data reviewers.

FINANCIAL DISCLOSURE

Does not have any relevant disclosures.

The joint venture of Department of Molecular and Human Genetics at Baylor College of Medicine (BCM) and Baylor Genetics Laboratories (BG) derives revenue from the clinical sequencing offered at BG and the authors who are BCM members or BG employees are indicated in the affiliation section.