Understanding Next Generation Sequencing Results
An Update on Elements of Variant Interpretation and Classification

Karen Vikstrom[1], Brianne Kirkpatrick[2], Michael Anderson[1], Keith Nykamp[1], Scott Topper[1], Martin Powers[1]
[1]Invitae Corporation, San Francisco, CA, [2]Geisinger Health System, Danville, PA

New Standards for the Interpretation of Sequence Variants

Background: With the growing acceptance and utilization of next generation sequencing (NGS) techniques for diagnostic purposes, the complexity and uncertainty of the findings are also expected to increase. Understanding the process by which a diagnostic laboratory analyzes and classifies sequence variants has clinical utility for genetic counselors, as patients need to be informed of the rationale behind variant interpretations, particularly in cases where a variant of unknown significance is found and may be reclassified in the future. The authors present an overview of the standard, evidence-based practice of scoring and classifying variants, as proposed by the most recent ACMG guidelines (March 2013). Examples of sequence variant interpretations, across all proposed classifications (i.e. Pathogenic, Likely Pathogenic, Uncertain Significance (VUS), Likely Benign and Benign) will be presented as a way to highlight important aspects of the interpretation process.

Methods: The PubMed database was searched for the literature published from 2007 to August 2014, looking for publications that concern a systematic approach to variant classification of genetic variants. Additional relevant resources were identified by reviewing webinars and presentations created by various laboratories. Information and updates from the ACMG Work Group that convened in 2013 was obtained through personal communication with an Invitae Laboratory Director.

The American College of Medical Genetics clinical laboratory standards for next-generation Sequencing (Practice Guidelines, Sept 2013) were also reviewed and included; see the website for more information: https://www.acmg.net/.

Nomenclature for Sequence Variants

ACMG/HGVS Recommendations: Due to increasing confusion over the clinical implications of commonly used terms such as "mutation" and "polymorphism", a uniform nomenclature has been proposed by the ACMG and the Human Genome Variation Society (HGVS). The terms "sequence variant" or "sequence alterations" are recommended in place of "mutation" and "polymorphism". Additionally, a 5-tier system describing the variant is recommended.

The clinical significance of a sequence change falls along a gradient, which is captured by the 5-tier classification system:

Pathogenic ⇔ Likely Pathogenic ⇔ Uncertain Significance ⇔ Likely Benign ⇔ Benign

Pathogenic (P): This sequence change is expected to directly contribute to the development of disease, but is not necessarily fully penetrant or sufficient to cause disease on its own. Additional evidence is not expected to alter the classification of this sequence change.
Likely Pathogenic (LP): This sequence change is very likely to contribute to the development of disease, however the scientific evidence is currently insufficient to prove this conclusively. Additional evidence is expected to confirm this assertion of pathogenicity, however, occasionally additional evidence may demonstrate that this sequence change has little or no clinical significance.
Uncertain Significance (VUS): There is not enough information at this time to support a more definitive classification of this sequence change.
Likely Benign (LB): This sequence change is not expected to have a major effect on disease, however the scientific evidence is currently insufficient to prove this conclusively. Additional evidence will likely confirm this assertion, but occasionally additional evidence may indicate a clinically important effect on disease for this sequence change.
Benign (B): This sequence change does not have a major effect on disease. Although some apparently benign changes may confer low increases or decreases in disease risk, this classification system only pertains to variants with likely high (>50%) penetrance.
One additional classification has been developed at Invitae:
Pathogenic (low penetrance). This sequence change is commonly accepted as a contributing factor of disease. However, the penetrance of this particular change is sufficiently low (<25%) that it is often seen in individuals without disease. As a result, the predictive value of this information may not be very high.

Why Are There Discrepancies Among Labs?

Different labs score and weigh the evidence in slightly different ways. For example, one lab may assign more weight to segregation data and less weight to functional/predictive data, resulting in a different classification. Currently, there is a trend towards standardization of methodologies and towards sharing variant information in publicly available databases such as ClinVar. The authors support the sharing of data, as it promotes the advancement of genomic medicine and therapies, and is a great benefit to patients and researchers.

Logic and Evidence

Variant interpretation involves a formal assessment of the available evidence. The ACMG guidelines have been updated to be more granular and descriptive of the evidence used to classify variants. A scoring system that rates the evidence as "Very Strong", "Strong", "Moderate", and "Supporting" was also proposed. For example, PVS = Pathogenic Very Strong, PM= Pathogenic Moderate, PS = Pathogenic Supporting, etc.

Examples of the resources used to evaluate and score a variant are listed below. Often, multiple lines of evidence are used to support a pathogenic/benign call. Variants without enough supporting evidence to fall into the P/LP or LB/B range fall into the category of Uncertain Significance.

The process starts with an evaluation of the evidence listed in the top three boxes. If a variant is found in the patient databases or in a literature search, then segregation and functional data is reviewed. If none exist, then predictive data is included in the report, but this is a much weaker type of evidence and often doesn't influence classification.

Lines of Evidence

Type of Variant
- Protein Truncating
  - Frameshift
  - Nonsense
  - Exonic deletions
- Consensus Splice Site
- Missense
  In-frame gain or loss
- Silent
  non-conserved intronic

Population Frequency
- **Databases**
  dbSNP
  NHLBI GO ESP
  1000 Genomes
  ClinSeq Project
- **Benign**
  - Recessive > 3 %
  - Dominant > 1.5%
- **Likely Benign**
  - Recessive >1 %
  - Dominant > 0.5%

Patient Information
- **Databases**
  ClinVar
  OMIM
  HGMD
  LOVD
  Decipher
  ClinVitae
- Literature Search

Segregation
- De novo
- Multiple **affected** family members and/or unrelated individuals
- Multiple **unaffected** family members and/or unrelated individuals
- Co-occurrence with causative variant (dominant only)

Functional
- Is there experimental evidence that this sequence change impacts protein function?
- Protein or Splicing effect?
- Is it deleterious or neutral?
- Is the evidence strong or weak?

Predictive
- Evolutionary Conservation
- Computational Algorithms
- Functional Domain?
- Mutation Hotspot?
- Essential amino acid?

Example of a Conflicting Variant Call

The Invitae variant interpretation system is a modified point-based system that assigns a pre-determined number of points for any evidence type, and calculates an interpretation based on the points allotted. The ACMG system assigns weight with qualitative descriptions (e.g., PM= Pathogenic Moderate). Below is an example of the two systems of evidence in a side-by-side comparison. Note, the same slight discrepancy could happen between a LB>VUS <LP.

Variant: CDH1 NM_004360.3:c.2343A>T (p.Glu781Asp)	
ACMG	Invitae (points assigned)
PS: Well-established functional studies	Protein Function Disrupted (2.5)
PM2: Absent from population	Novel with expected phenotype (2.5)
PP4: Expected Phenotype	
PP1: Co-segregation	Weak segregation (1.0)
PP3: Computational predictions	All protein predictors deleterious (0.5)
Likely Pathogenic (1 Strong + 1 Mod + 3 Sup) Need 1 more Sup for Pathogenic	**Pathogenic (6.5)**