Illumina Clinical Services Laboratory Variant Classification

Summary

Variant classification in the Illumina Clinical Services Laboratory (ICSL) is performed in accordance with established guidelines (Richards et al. 2015), and utilizes an integrated evaluation of knowledge regarding the gene and disease relationship, frequency information obtained from population databases, and other relevant evidence gleaned from the literature and other resources.

Gene List

- For a predefined list of genes associated with Mendelian disorders, information regarding the gene and associated disease is held in an internally curated list that details the disease name, gene symbol, transcript, inheritance mode, and penetrance and prevalence estimates. The information in the list is curated from four sources (GeneReviews, Genetics Home Reference, Orphanet, and Online Mendelian Inheritance in Man (OMIM)) and is updated regularly.
- For genes outside of the predefined list in which a variant of interest is selected for curation, a manual gene curation is pursued to establish the strength of evidence supporting a relationship with Mendelian disease. The information is curated using the sources listed above as well as ClinVar, ClinGen, and primary literature.

Variant Curation

- Variants that are found in HGMD (prior to June 1, 2018) are subjected to a manual curation and classified based on variant frequency and evidence from the literature.
- Population frequency information is gathered from:
    - 1000 Genomes Project
    - NHLBI GO Exome Sequencing Project (ESP)
    - Exome Aggregation Consortium (ExAC)
    - Genome Aggregation Database (gnomAD) (for more recent variants)

  Frequency information also may be interrogated from our internal database of whole genome sequences.
- Literature searches are performed in PubMed, PubMed Central (PMC), Google Scholar, Google, HGMD (prior to June 1, 2018) and ClinVar for each variant using the gene name, cDNA change, amino acid change, and rsID, utilizing alternative nomenclature as available.
- Additional resources are consulted as required depending on the variant. For example, the effect of potentially truncating or elongating variants is confirmed using Mutalyzer.
- Variants that are not found in HGMD (prior to June 1, 2018) are subject to classification using an automated classification scoring system. Utilizing variant allele frequency, disease prevalence and penetrance estimates, and inheritance mode, an automated score is calculated to assess if the variant is too frequent to cause the disease in question. Based on the score and internal cut-off values, a variant classified as benign or likely benign is not subjected to manual review. Variants that cannot be ruled out on the basis of autoscore are subjected to a literature search as outlined

above. If no literature is found, the variant is classified as a variant of unknown significance. If literature is found, the variant is subjected to the manual curation process outlined above. Details of the autoscore calculation are given below.

Variant Classification

Variants are classified based on assessment and concordance of the available evidence. Each variant classification is subject to professional review. Our classification system has been developed from the recommendations of the American College of Medical Genetics (ACMG) for variant classification and reporting (Richards et al. 2015), with consideration that many variants are detected in ostensibly healthy adults and with the addition of a sixth category termed VUS-Suspicious (VUS-S). This category was developed for variants of unknown significance that have limited evidence for pathogenicity but are deemed noteworthy for reporting, bringing attention to variants that are on the border between unknown significance and likely pathogenic. Variants classified as VUS-S are submitted as VUS in ClinVar, with language distinguishing the VUS-S classification found in the evidence summary.

The classification criteria for each of the six categories are given below. For the clinically significant classifications, these take the form of minimum criteria, with the following assumptions:

- Individuals in the literature carrying the variant have the disease / phenotype the variant is being classified for.
- The methodology of the study under curation is thought to be robust and able to detect the types of variants under curation.
- Genetic heterogeneity is always taken in to account.
- The variant frequency is higher in cases than in controls and not too frequent to rule the variant out of being disease-causing.
- All functional studies used in classification are consistent, strong and relevant to the disease mechanism.
- Loss of function should be established as a mechanism of disease for all null variants.

For these criteria, null variants are defined as stop-gained, stop-lost, frameshifts, canonical splice variants (+/- 1 or 2 of the acceptor or donor site), or initiation codon variants. Other variant types are defined as missense variants, splice region, 3’ UTR, 5’UTR, intronic and synonymous variants.

Pathogenic

- For recessive conditions:
    - For a null variant:
        - The variant is reported in the literature in three or more unrelated homozygotes or compound heterozygotes with functional evidence or other evidence suggesting a deleterious effect of the variant on gene expression or function consistent with the mechanism of disease.
          OR
        - In the absence of functional data, the variant is found in an at least five individuals affected with the specific condition.
    - For other variant types:
        - The variant is reported in the literature in four or more unrelated homozygotes or compound heterozygotes with functional evidence or other evidence suggesting a deleterious effect of the variant on gene expression or function consistent with the mechanism of disease.

OR
        - In the absence of functional data, the variant is found in at least seven individuals affected with the specific condition.
- For dominant conditions:
    - For a null variant:
        - The variant is reported in the literature in four or more unrelated heterozygotes in at least two studies with or without functional evidence but with evidence of segregation of the variant with disease and a family history data.
    - For other variant types:
        - The variant is reported in the literature in five or more unrelated heterozygotes, in at least two separate studies with strong functional evidence
        - OR
        - In the absence of functional data, the variant is reported in the literature in five or more unrelated heterozygotes, in at least two separate studies with evidence of segregation
          OR
        - In the absence of functional data, the variant is reported in the literature in seven or more unrelated heterozygotes, in at least two separate studies with family history data.

Likely Pathogenic

- For recessive conditions:
    - For a null variant:
        - The variant is reported in the literature in at least two unrelated homozygotes or compound heterozygotes.
    - For other variant types:
        - The variant is reported in the literature in two or more unrelated homozygotes or compound heterozygotes with functional evidence or other evidence suggesting a deleterious effect of the variant on gene expression or function consistent with the mechanism of disease.
          OR
        - The variant is found in three unrelated homozygotes or compound heterozygotes without functional evidence but with moderate additional evidence.*
- For dominant conditions:
    - For a null variant:
        - The variant is reported in the literature in at least two unrelated heterozygotes with or without functional data but with family history data.
    - For other variant types:
        - The variant is reported in the literature in three or more unrelated heterozygotes, in at least two separate studies with strong functional evidence
          OR
        - The variant is reported in the literature in three or more unrelated heterozygotes, in at least two separate studies without functional evidence but with evidence of segregation
          OR
        - If there is no functional evidence, the variant is reported in the literature in five or more unrelated heterozygotes, in at least two separate studies with family history data.

*Moderate additional evidence includes located in a mutational hot spot and/or critical and well-established functional domain without benign variation; for recessive disorders, being detected in trans (with phase confirmed) with a known pathogenic variant in at least one patient; novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.

Variant of Unknown Significance-Suspicious (VUS-Suspicious)

- There is limited evidence that the variant could be causative of disease. The information available is insufficient to categorize the variant as likely pathogenic. This category was added to bring attention to variants that are on the border between unknown significance and likely pathogenic. For example, if the variant is reported in only a single homozygote or compound heterozygote with or without functional evidence for a recessive condition, or in a very limited number of cases for a dominant condition, this evidence is limited and suggestive of pathogenicity but is not conclusive.
- Null variants, including nonsense, frameshift, canonical +/- 1 or 2 splice sites, and initiation codon variants, with no other supporting evidence are considered to be suspicious for pathogenicity and are classified in this category. Additional evidence is needed to classify these variants as likely pathogenic or pathogenic.

Variant of Unknown Significance

- Little or nothing has been reported regarding this variant, or the reported evidence in the literature is incomplete and/or contradictory.
- The evidence could be contradictory within the literature or between the literature and other available evidence (e.g., allele frequency).

Likely Benign

- The variant is reported in the literature in a similar number of cases and controls if control data are available.
- The variant does not segregate with disease within a family.
- Variant frequency is higher than expected in the general population based on inheritance mode and disease prevalence and penetrance estimates.
- The variant may be non-conserved and / or predicted to be well-tolerated.
- Functional evidence or other evidence suggests no deleterious effect of the variant on gene expression or function.

Benign

- The variant is not reported in the literature in cases or is reported in a similar number of cases and controls if control data are available.
- Established in the literature as a variant that is not associated with Mendelian disease.
- The variant does not segregate with disease within a family.
- Variant frequency is too high to be causative based on inheritance mode and disease prevalence and penetrance estimates.
- The variant may be non-conserved and / or predicted to be well-tolerated.
- Functional evidence or other evidence suggests no deleterious effect of the variant on gene expression or function.

Sources Used in Variant Classification

Frequency information:
1000 Genomes Project: http://browser.1000genomes.org
NHLBI GO Exome Sequencing Project (ESP): http://evs.gs.washington.edu/EVS/
Exome Aggregation Consortium (ExAC): http://exac.broadinstitute.org/
Genome Aggregation Database (gnomAD): http://gnomad.broadinstitute.org/about

Literature searches:
PubMed: https://www.ncbi.nlm.nih.gov/pubmed
PubMed Central (PMC): https://www.ncbi.nlm.nih.gov/pmc
Google Scholar: https://scholar.google.com/
Google: https://www.google.com/

Gene and disease information:
GeneReviews: https://www.ncbi.nlm.nih.gov/books/NBK1116/
Genetics Home Reference: https://ghr.nlm.nih.gov/
Orphanet: http://www.orpha.net/consor/cgi-bin/index.php?lng=EN
Online Mendelian Inheritance in Man (OMIM): https://www.ncbi.nlm.nih.gov/omim

Disease-specific resources:
Cystic Fibrosis Mutation Database: http://www.genet.sickkids.on.ca/app
Clinical and Functional Translation of CFTR (CFTR2): http://cftr2.org/
Tuberous Sclerosis Project: http://tsc-project.partners.org/
Locus-Specific Mutation Databases: http://www.hgvs.org/locus-specific-mutation-databases
The International Society for Gastrointestinal Hereditary Tumours (InSiGHT): https://www.insight-group.org/variants/databases/
Locus Specific Database list: http://grenada.lumc.nl/LSDB_list/lsdbs

Additional resources:
ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
dbSNP: https://www.ncbi.nlm.nih.gov/SNP/
HUGO Genome Nomenclature Committee (HGNC): http://www.genenames.org/
Human Gene Mutation Database (HGMD): http://www.hgmd.cf.ac.uk/ac/index.php
Human Genome Variation Society: http://www.hgvs.org
Mutalyzer: https://mutalyzer.nl/
NCBI Protein database: https://www.ncbi.nlm.nih.gov/protein
UCSC Genome Browser: http://genome.ucsc.edu
Varsome: https://varsome.com/

Autoscore Calculation

Variants that are not immediately sent for manual curation are automatically categorized and classified according to specific autoscore cut-offs, as listed below. Different equations are used for diseases following an autosomal recessive inheritance pattern, or autosomal dominant or X-linked inheritance. These equations are listed below:

$$AR\ autoscore\ = \mathrm{Log}_{10}\frac{\left[\left(95\%\ \text{CI lower bound of}\ \frac{\mathrm{AC}}{\mathrm{AN}}\right)^{2} * \text{Penetrance}\right]}{\text{Prevalence}}$$

$$AD\ or\ X-linked\ autoscore\ = \mathrm{Log}_{10}\frac{\left[\left(1-\left(1-95\%\ \text{CI lower bound of}\ \frac{\mathrm{AC}}{\mathrm{AN}}\right)^{2}\right) * \text{Penetrance}\right]}{\text{Prevalence}}$$

The values used for AC (allele count) and AN (allele number) are obtained from databases listed in the frequency information section above.

- Score cut-offs
  - ≤ 0.0: Variant of Unknown Significance (VUS)
  - (0.0-1.18): Likely Benign
  - ≥ 1.18: Benign

References

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May; 17(5):405-24