Harmonizing variant interpretation for sequence and copy number changes within individual genes in clinical diagnostic testing

Ali Entezam, Rebecca Truty, Yuya Kobayashi, Dan Schmidt, Rachel Lewis, Amanda Stafford, Darlene Riethmaier, Scott Topper, Keith Nykamp, Swaroop Aradhya

Invitae Corporation, San Francisco, CA

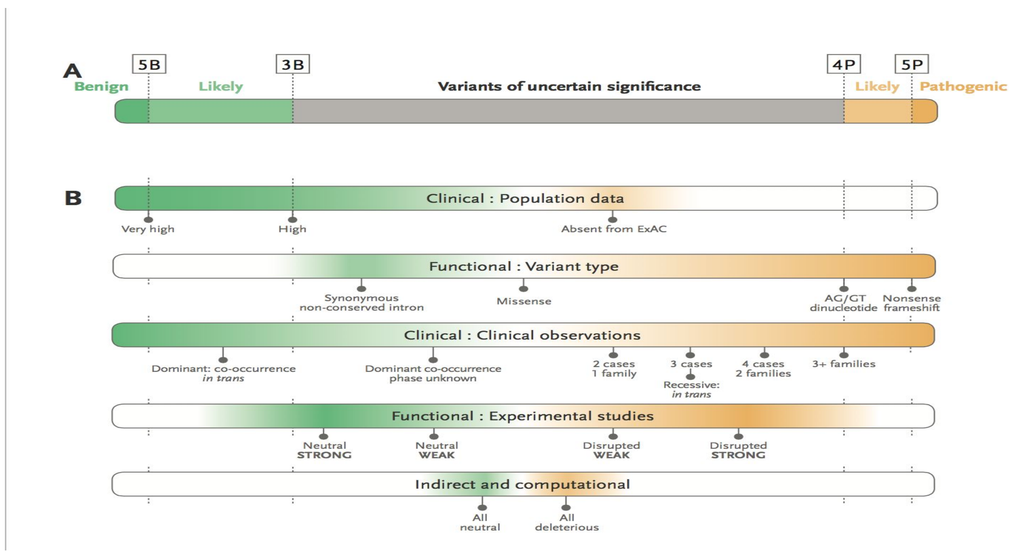
Disclosure statement: All authors are employees and stockholders of Invitae Corporation.

Abstract

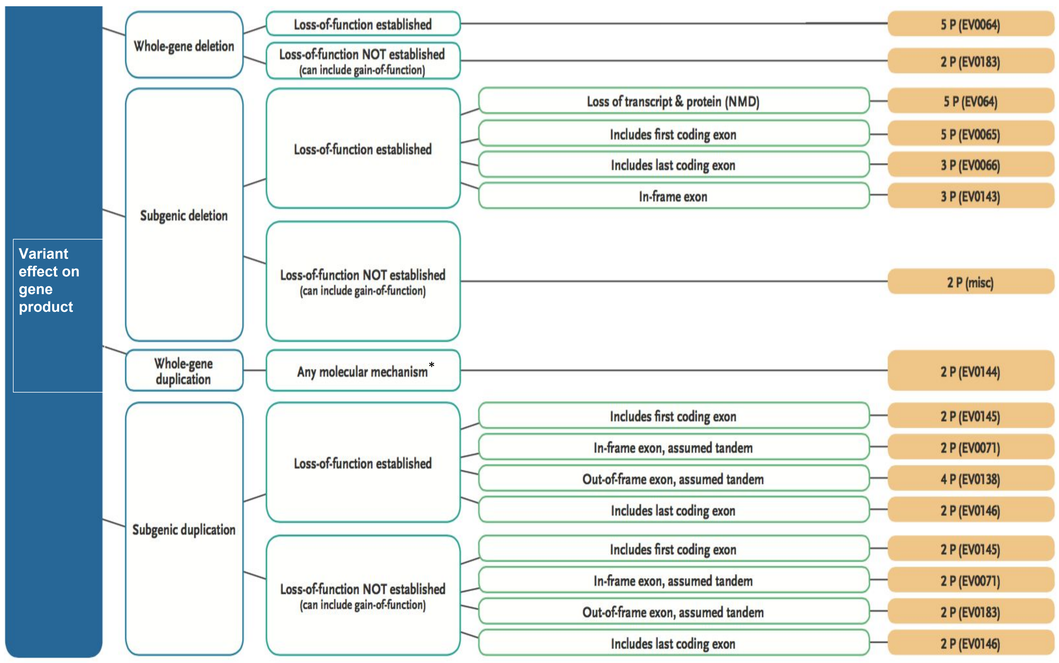
As methods to produce vast amounts of genetic information become more accessible in clinical genetic testing for germline disorders, the establishment of rigorous rubrics for interpreting the significance of that information consistently across large numbers of individuals has emerged as a major challenge. These data include both sequence (single nucleotide) variants (SNVs) and copy number variants (CNVs). Recently updated ACMG guidelines for SNV interpretation are detailed and specific, encouraging more consistent application of evidence among diagnostic laboratories. We recently described our interpretation schema (named "Sherloc") that augments the ACMG guidelines with more granular rules and contexts for applying various lines of evidence. In contrast to SNVs, CNVs require the application of a specific set of concepts to assess evidence of pathogenicity. The ACMG guidelines for interpreting CNVs are pointed at large cytogenetic aberrations extending hundreds of kilobases but not at CNVs that affect all or part of a single gene. The ACMG CNV guidelines also provide useful high-level conceptual guidance but are not as detailed as those available for SNVs. Because many genetic disorders are caused by either pathogenic SNVs or CNVs, we have extended Sherloc to include a framework for interpreting CNVs that impact single genes either as intragenic exonic events or large cytogenetic events that end within a gene. This extension has harmonized the interpretation of the broad spectrum of genetic variation observed in clinical genetic testing. We describe 26 rules in Sherloc that apply to both SNV and CNV interpretation. A subset of these rules applies to large cytogenetic events, including those that impact disease-causing single genes at their boundaries. In a large set of CNVs identified from diagnostic testing, 415 were deletions and 161 were duplications. These CNVs separated into nine categories: partial gene in-frame deletions, partial gene out-of-frame deletions, partial gene deletions including first exon, full gene deletions, partial gene in-frame duplications, partial gene out-of-frame duplications, terminal exon duplications, partial gene duplications including first exon, and full gene duplications. Each of these categories required a specific set of rules to interpret clinical significance, and we have integrated them into Sherloc to complement SNV interpretation and provide examples of SNV and CNV interpretation within individual cases. In summary, we describe explicit interpretation criteria for intragenic or cytogenetic CNVs that take molecular genetic principles into account. Sherloc now includes 123 rules for interpreting SNVs, including 26 rules that apply to CNVs, and we continue to deepen its granularity to ensure the sophisticated and consistent interpretation of clinically observed variants.

Methods

- We use a point system for variant interpretation (Sherloc) that is based on the ACMG guidelines[1,2]:

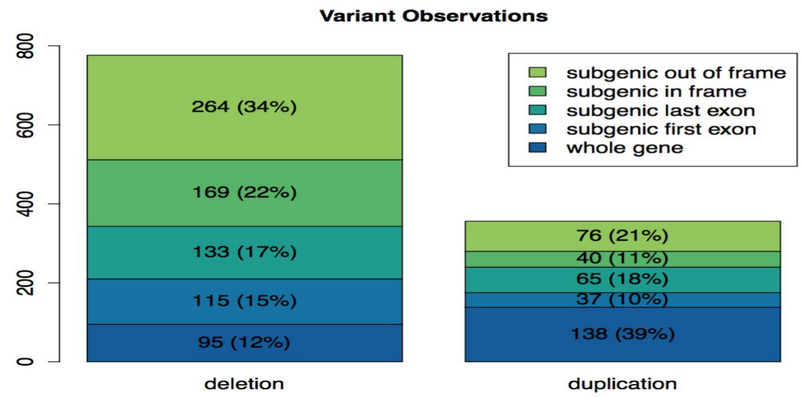


- We defined set of evidence-based criteria that can be applied to intragenic CNV interpretation. The criteria for aberrations that affect all or part of a single gene are used in conjunction with a subset of criteria, such as case reports, segregation, and functional evidence, for interpreting SNVs



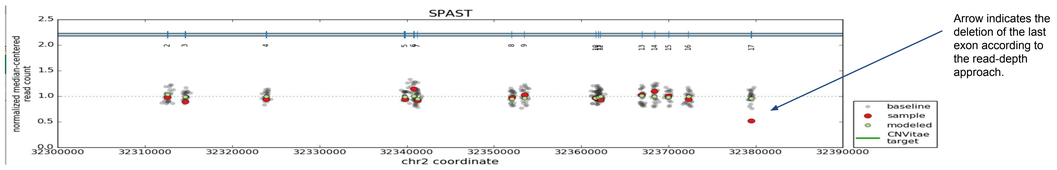
*We recognize that the duplication of a gene with known dosage sensitivity will be pathogenic (e.g., PMP22, PLP1, MECP2); however, we are currently developing rules for establishing dosage sensitivity as a molecular mechanism in the context of our interpretation system.[3]

Results



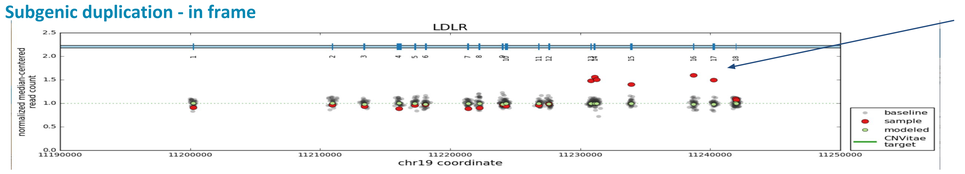
Examples

A. Subgenic deletion - includes last exon



Evidence type	Pathogenic points	Notes	Evidence for the variant
CNV: subgenic deletion - includes last exon	3	This criteria is for a subgenic deletions that encompass the last exon but does not include the initiator codon. These are predicted to result in the expression of a truncated protein and not expected to result in nonsense mediated decay. Consider the number of amino acids deleted from the predicted protein relative to the total size of the normal protein and whether the deleted amino acids might be critical for protein function. If the deleted region encompasses well-documented, loss-of-function (LOF) missense mutations, use in combination with criteria: Disruption of essential amino acid residue to yield a likely pathogenic (LP) interpretation.	LOF established gene
Moderate segregation with disease	2.5	LOD > 1.5 in two or more families. For dominant genes, this will generally require at least six informative individuals. For recessive genes, this will generally require three affected individuals, but two affected individuals and four unaffected siblings will also suffice. This number of individuals assumes high penetrance (>95%) and low possibility of phenocopies. A LOD score should be calculated when possible.	Moderate segregation PMID: 17345589, 17035675, 11134375
Pathogenic	5.5		

B. Subgenic duplication - in frame



Evidence type	Pathogenic points	Notes	Evidence for the variant
CNV: subgenic duplication - in-frame, assumed tandem	2	This criteria is for a subgenic duplication that do not encompass either the first exon or the last exon. Given the small size (and rarity) of these subgenic duplications, we make the assumption that they most likely occur in tandem. Use this criterion if the inclusion of duplicated exons into the spliced transcript may result in an in-frame insertion. The effect of this insertion on protein function is very difficult to predict without experimental evidence.	in-frame duplication
Disruption of essential amino acid residue	1.5	An essential amino acid can be inferred from previous reports of pathogenic (P) (should meet our criteria for LP or P) missense changes at this codon. Consider the mechanism of pathogenicity for the previously reported change. If the mechanism is known to be splicing, do not use this criterion. This criterion can also be used for in-frame indels, exonic deletions that encompass a known pathogenic missense, and truncating mutations that are expected to escape NMD. To use this criteria with a truncating variant that may escape NMD, (1) the disease mechanism must be LOF-established, and (2) the PTC created by the truncating variant must be upstream of the LP/P missense variant.	This duplication includes functional and structural domains whose duplication would interfere with low density lipoprotein receptor function (PMID: 6091915, 2988123).
Weak segregation with disease	1	Three or more informative individuals from one or more families.	Present in three family members tested at Invitae
Likely pathogenic	4.5		

Conclusions

- The ACMG guidelines for interpreting CNVs are pointed at large cytogenetic aberrations extending hundreds of kilobases but not at CNVs that affect all or part of a single gene. The ACMG CNV guidelines also provide useful high-level conceptual guidance but are not as deeply detailed as those available for SNVs.
- We have developed and extended Sherloc rules for interpreting CNVs that impact single genes either as intragenic exonic events or large cytogenetic events that end within a gene. These rules have been integrated into Sherloc to complement SNV interpretation and are applied in conjuction with a subset of criteria that apply to both CNVs and SNVs.
- This extension has harmonized the interpretation of the broad spectrum of genetic variation observed in clinical genetic testing.

References

1. Richards S *et al.*, 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; 17(5):405-24.
2. Invitae. Invitae's method of variant classification. Available at https://www.invitae.com/en/variant-classification/. Accessed March 9, 2017.
3. ClinGen. ClinGen dosage sensitivity map. Available at https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/. Accessed March 10, 2017.