Harmonizing clinical interpretation of intragenic sequence and copy number variants in monogenic disease

INVITAE

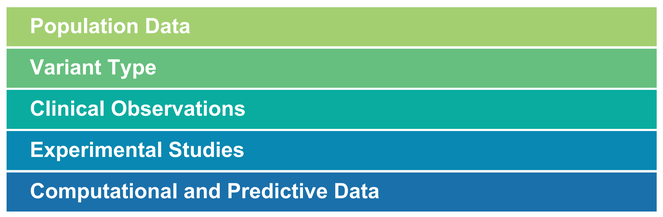
Yuya Kobayashi, Ali Entezam, Rebecca Truty, Rachel Lewis,
Keith Nykamp, Invitae Clinical Genomics Group, Swaroop Aradhya
Invitae, San Francisco, CA

Introduction

- Intragenic copy number variant (CNV) detection is increasingly utilized in clinical genetic testing.
- The 2015 ACMG variant interpretation guidelines provided limited guidance for interpreting single/multi-exon deletions and do not address single/multi-exon duplications.
- Previously, we published the **Sherloc** variant interpretation schema based on the ACMG guidelines. Sherloc was iteratively refined based on the experiences of interpreting thousands of intragenic CNVs from a clinical cohort of over a hundred thousand individuals.
- These refinements have resulted in a unified variant interpretation system that allows for accurate, consistent, and efficient classifications for both sequence variants (SV) and intragenic CNVs.

Methods

- A key feature of Sherloc was the reorganization of weighted evidence-criteria into **categories** based on five **molecular and clinical genetics concepts**:



This allowed for the establishment of **complex relationships** within and between evidence categories.

Example relationships:

- Experimental data **trumps** *in silico* predictions.
- High allele frequency in the general population **modulates** the significance of a variant observations in a patient.
- Co-segregation of a variant with disease is **additive** for each additional family.

- Sherloc was expanded for interpretation of CNVs by:
  - Creation of additional evidence-criteria for **different types of CNVs**, and weighted *relative to* previously established criteria of similar concept.

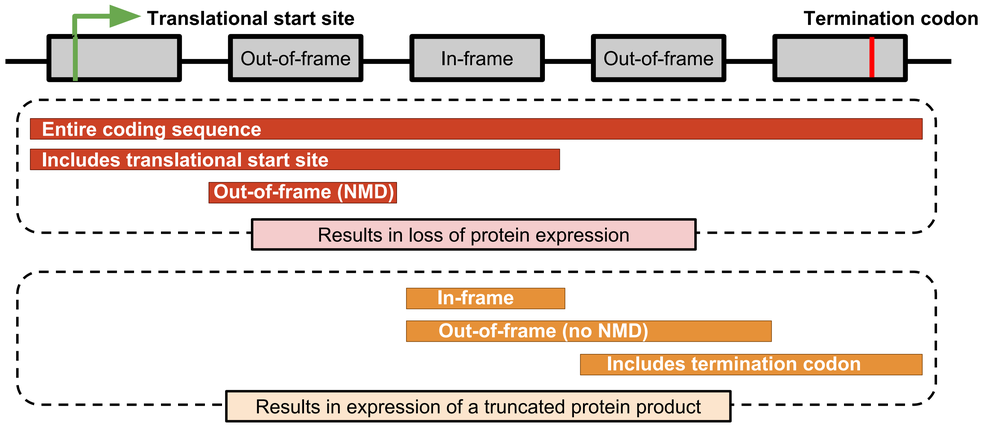
Examples:

A **whole-gene deletion is similar to a nonsense variant that leads to nonsense-mediated decay** since both abolish protein expression. These two criteria should be weighted the **same**.

A **duplication of an out-of-frame exon is similar to a frameshift that leads to nonsense-mediated decay** only if the duplication occurred in a tandem orientation. These two criteria should be weighted **similarly, but adjusted** for the possibility of the duplication event is not in tandem.

  - Inclusion of these CNV-related criteria within the **Variant Type** category to maintain established relationships with criteria in other categories.

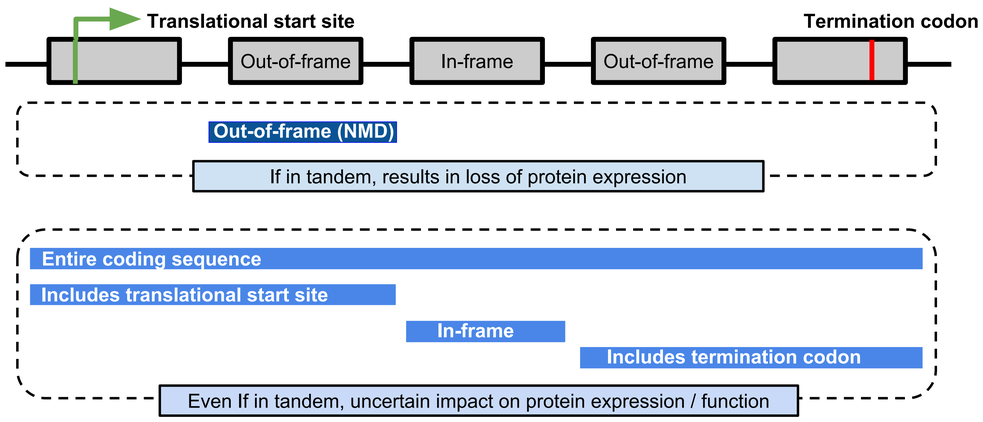
Types of Copy Number Loss Variants (Deletions)



Variant Type	Mechanism of disease is loss-of-function	Sherloc Score*
Whole gene deletion	**Yes**	**5 pts**
	No	2 pts
Include translational start site	**Yes**	**5 pts**
	No	2 pts
Out-of-frame exon(s), does not include penultimate coding exon	**Yes**	**5 pts**
	No	2 pts
Out-of-frame(s), includes penultimate coding exon	**Yes**	3 pts
	No	0 pts
In-frame exon(s)	**Yes**	3 pts
	No	0 pts
Includes termination codon	**Yes**	3 pts
	No	0 pts

* Sherloc score of 5 points corresponds to a classification of **Pathogenic** without requiring additional supporting evidence. A score of 4 points corresponds to a classification of **Likely Pathogenic.**

Types of Copy Number Gain Variants (Duplications)

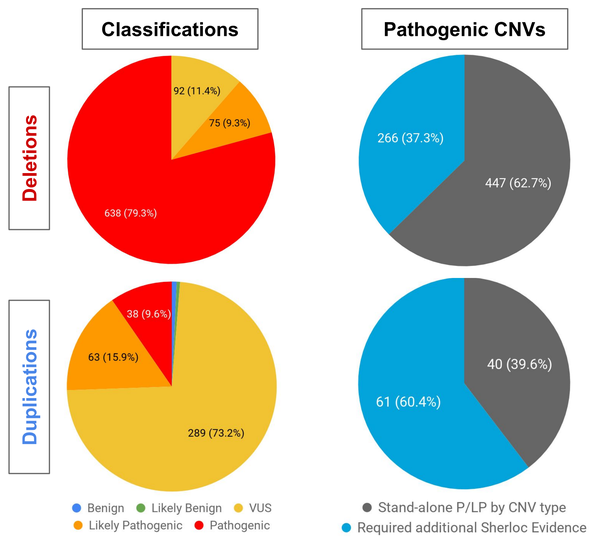


Variant Type	Mechanism of disease is loss-of-function	Sherloc Score*
Whole gene duplication	**Yes**	2 pts
	No	0 pts
Include translational start site	**Yes**	2 pts
	No	0 pts
Out-of-frame exon(s), does not include penultimate coding exon	**Yes**	**4 pts**
	No	0 pts
In-frame exon(s)	**Yes**	2 pts
	No	0 pts
Includes termination codon	**Yes**	2 pts
	No	0 pts

* Sherloc score of 5 points corresponds to a classification of **Pathogenic** without requiring additional supporting evidence. A score of 4 points corresponds to a classification of **Likely Pathogenic.**

The Impact of a Unified Interpretation System on a Clinical Cohort

- In a clinical cohort of 143,515 individuals, we observed 805 *unique* deletion CNVs and 395 *unique* duplication CNVs.
- 68% (814/1200) of them reached a classification of Pathogenic or Likely pathogenic (P/LP).
- Of these P/LP variants, 40% (327/814) would have been classified as a VUS had it not been for supporting evidence from other Sherloc categories that were originally created for SV interpretations, such as segregation and clinical observations.



Example: MLH1 Deletion (Exon 16-19)

Variant Type: **CNV Deletion that includes the termination codon** in a gene where molecular mechanism of disease has been established as **LOF**.

This is expected to result in a truncated MLH1 protein product.

Sherloc Score: 3 pts

Additional Sherloc evidence:

- **Segregation** in a family (1 pt)
- A single amino acid deletion within this region (p.Lys618del) has been classified as pathogenic, indicating that this CNV disrupts an **essential codon** (1.5 pts)

Final Classification:

Pathogenic (5.5 pts)

CONCLUSIONS

- The organizational structure of Sherloc, defined by molecular and clinical genetics concepts, allowed for seamless expansion of CNV-type evidence criteria, without perturbing any other component of this variant interpretation schema.
- The ability to apply previously-established Sherloc criteria from the other four categories allowed for a large fraction of variants to reach a classification of Pathogenic or Likely Pathogenic. This was particularly critical for duplication CNVs.

Disclosures: Authors are stockholders and employees of Invitae.