Integration of the ACMG ISV Guidelines into Clinical Variant Analysis and Interpretation Processes: A Model for Diagnostic Standardization

Keith Nykamp, Yuya Kobayashi, Shan Yang, Stephen Lincoln, Stephanie Gandomi, Michael Anderson, Martin Powers, Scott Topper and The Invitae Clinical Genomics Group

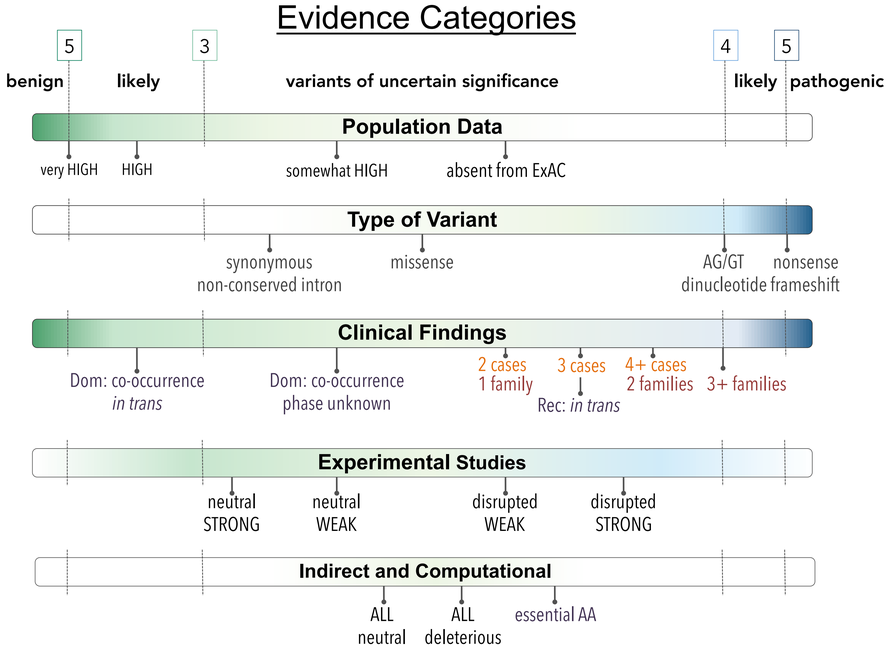
Introduction

With increasing use of high-throughput sequencing and multi-gene panels for diagnostic purposes, there is growing concern about the potential for inconsistent variant classifications among clinical labs. The American College of Medical Genetics and Genomics (ACMG) recently drafted new standards and guidelines for the interpretation of sequence variants (ISV) to address this concern (Richards et al., Genet Med 2015). Draft versions of these guidelines were shared broadly with the clinical genetics community, feedback was incorporated, and an evidence-based checklist for interpreting Mendelian disease variants has now been published. This checklist represents a major step towards evidence assessment standardization and variant classification consistency. However many ISV criteria are quite expansive which could result in inconsistency in their application. To date, a rigorous study has not been published to examine the impact of these new guidelines on variant classification and clinical reporting. It also remains to be seen whether the new evidence checklist results in increased interpretation concordance between clinical laboratories. We considered these issues when validating our own laboratory's classification procedures which are based on the new ISV guidelines.

A Score-Based System for Interpretation

We developed, validated and implemented a score-based variant classification system based on the ISV guidelines. This implementation, "Sherloc", adheres to the guidelines as published, while adding refinements to ensure consistent use and accurate conclusions when applied broadly. This implementation adds greater resolution to broad ISV rules that encompass various use cases, adds dependencies between rules to reduce redundant use of evidence, and incorporates detailed usage notes for various evidence types.

Evidence Categories



Defining Principles

1. Five major categories of evidence are considered in a hierarchical approach.
2. A point score is given to evidence (EV) criteria within each category.
   - A null variant (i.e. nonsense, frameshift) in LOF gene is 5 points for Pathogenic.
   - Very HIGH MAF (i.e. much greater than expected) is 5 points for Benign.
   - All other EV scores reflect the relative importance along this scale.
   - Miscellaneous EV for Benign or Pathogenic can be applied as necessary.
3. Only the strongest EV is counted if two criteria contribute to the same basic argument (e.g protein function is disrupted, either STRONG or WEAK).
4. Certain EVs can be counted more than once (e.g. single co-occurrences) and others can not (e.g. 4+ case reports).
5. EV scores are calculated in Pathogenic and Benign directions independently.
6. Classification thresholds (P, LP, LB, B) are pre-determined, and point scores above these thresholds receive a "calculated" classification.
7. Calculated classifications are suggestions only. The final interpretation is manually performed by a trained genomic scientist. All variant interpretations are reviewed by a 2nd experienced genomic scientist and an ABMG-certified medical geneticist.

A Consensus-Based Method for Evaluating an Interpretation Schema

To evaluate concordance of this method with community standards, we constructed a consensus of Sherloc and all ClinVar interpretations for variants where at least 2 other diagnostic lab reports were available. Consensus was defined as at least 60% of laboratories (i.e. 2 out of 3 labs, 3 of 4, etc.) exactly agreeing on the 5-class scale. We then conducted an "odd man out" analysis to see how often any lab disagreed with this consensus. This particular methodology was chosen to facilitate an identical comparison for all labs, using the same variants and same consensus interpretations. 814 variants could be compared and 66 were excluded because no consensus was formed.

Consensus Interpretation

Sherloc	Pathogenic	LP	VUS	LB	Benign	Total	#Variants
Pathogenic	99.1%	0.9%	0.0%	0.0%	0.0%	100%	117
LP	50.0%	50.0%	0.0%	0.0%	0.0%	100%	10
VUS	0.0%	0.4%	93.6%	5.6%	0.4%	100%	249
LB	0.0%	0.0%	11.0%	78.0%	11.0%	100%	82
Benign	0.0%	0.0%	2.5%	2.2%	95.2%	100%	356
						Total	814

Sherloc interpretations were often (92.2%) in the consensus majority. When not, almost all of the differences were either in likelihood (i.e. Benign vs. LB or Pathogenic vs. LP) or were VUS vs. Benign/LB cases. The single exception was a Sherloc VUS compared to a consensus likely pathogenic call.

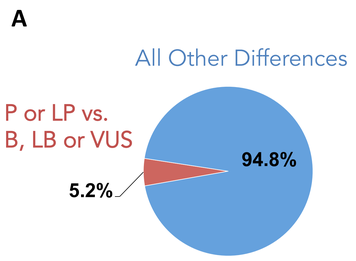
Consensus Interpretation

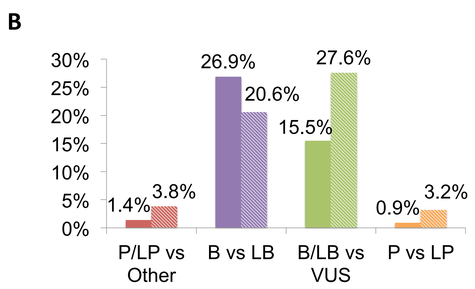
Other Labs	Pathogenic	LP	VUS	LB	Benign	Total	#Interpretations
Pathogenic	97.8%	1.3%	0.6%	0.3%	0.0%	100%	312
LP	13.3%	66.7%	20.0%	0.0%	0.0%	100%	15
VUS	0.6%	0.2%	93.1%	3.0%	3.0%	100%	494
LB	0.4%	0.0%	12.1%	49.8%	37.7%	100%	281
Benign	0.0%	0.0%	2.0%	4.4%	93.6%	100%	798
						Total	1900

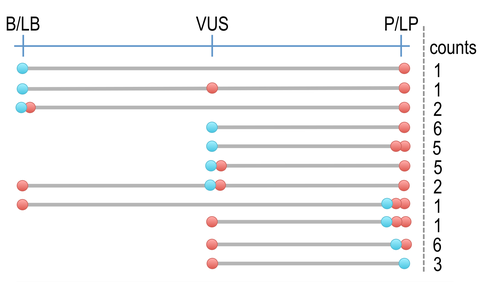
Other laboratories were also often in the consensus but less so (85.7%) and more clinically significant differences were observed. We conclude that Sherloc is at least as concordant with community consensus as are most ClinVar submissions for these variants.

Comparison with Individual Submissions

While a large number of ClinVar records had multiple submissions and a derived consensus interpretation, many did not. We thus also compared Sherloc against individual ClinVar submissions from diagnostic labs to evaluate the type and direction of disagreements that occur. This analysis included 1886 variants of which 623 had one or more disagreements with Sherloc.







Legend

A: Vast majority of disagreements are within VUS, LB and B classifications or 1-step P to LP (blue).

B: Directionality of disagreements: (e.g. Sherloc B vs ClinVar LB = 26.9%, Sherloc LB vs ClinVar B = 20.6%)

C: Type and counts of disagreements within the "P or LP vs. B, LB, or VUS" group.

Re-Evaluation of Discordant Variants

To evaluate these discordant variant classifications, and at the same time assess the robustness of the Sherloc system, we selected 42 variants to be re-interpreted by a different scientist who was blinded to the original interpretation. Importantly, these 42 variants are strongly enriched for "difficult-to-interpret" cases, and are not representative of most variants. Nevertheless, our re-interpretations of these variants showed high reproducibility:

- 39 (92.5%) matched the previous interpretation exactly
- 2 matched within the same pathogenicity group:
  - 1 P to LP change with inconsistent use of "Clinical Findings"
  - 1 B to LB change with omission of Misc. Benign evidence
- 1 changed (VUS to LP) with addition of Misc. Pathogenic evidence

Sherloc Interpretations

A

PALB2 c.3136G>A (p.Trp1038*)		Sherloc Score B	Sherloc Score P*
Population	EV0015: Absent from the population DBs (Not in ExAC)		~~1~~
Variant Type	EV0016: Truncating mutation, presumed loss of protein		5
Clinical Findings	EV0050: Moderate segregation (in 2 families)		2.5
Experimental	EV0027: Splicing defect: weak evidence (cDNA analysis)		1
Computational	Predicted splice site loss: : Majority say >25% reduction		~~0.5~~
	Invitae	**Pathogenic**	8.5
	Consensus	(Invitae only)	

* Strikethrough indicates weaker evidence not counted per Defining Principle #3 at left.

B

MYH7 c.1624+4A>T (Intronic)		Sherloc Score B	Sherloc Score P
Population	EV0101: Allele count within pathogenic range (1 in ExAC)		0.5
Variant Type	Intronic, within the consensus donor SS; No EV		
Clinical Findings	EV0079: 4 unrelated individuals with expected disease		3
Experimental	None		
Computational	EV0030: Predicted SS loss, majority say >25% reduction		0.5
	Invitae	**Likely Pathogenic**	4
	Consensus	**Likely Pathogenic**	

C

MSH6 c.2057G>A (p.Gly686Asp)		Sherloc Score B	Sherloc Score P
Population	EV0015: Absent from the population DBs (Not in ExAC)		1
Variant Type	Missense: no EV score		
Clinical Findings	EV0051: Weak segregation (3 individuals, from 2 families)		1
Experimental	None; No microsatellite instability detected in tumor		
Computational	EV0016: Multifactorial likelihood model (Likely Pathogenic)		1
	Invitae	**Uncertain Significance**	3
	Consensus	**Likely pathogenic**	
	Re-evaluation	**Uncertain Significance**	3

Conclusions

With the 2015 ACMG ISV recommendations for variant interpretation as our guide, we developed Sherloc, a score-based classification system with detailed evidence criteria, inherent logic for handling interdependent evidence, and comprehensive notes outlining caveats, various use cases, and evidence considerations for each criteria. This system has been implemented in our clinical testing workflow and refined over the past 15 months. Importantly, Sherloc interpretations are highly consistent with those submitted to CinVar. The methods used in this study to evaluate the concordance of Sherloc may be widely applicable to any clinical molecular laboratory looking to evaluate the concordance and consistency of their own interpretations.