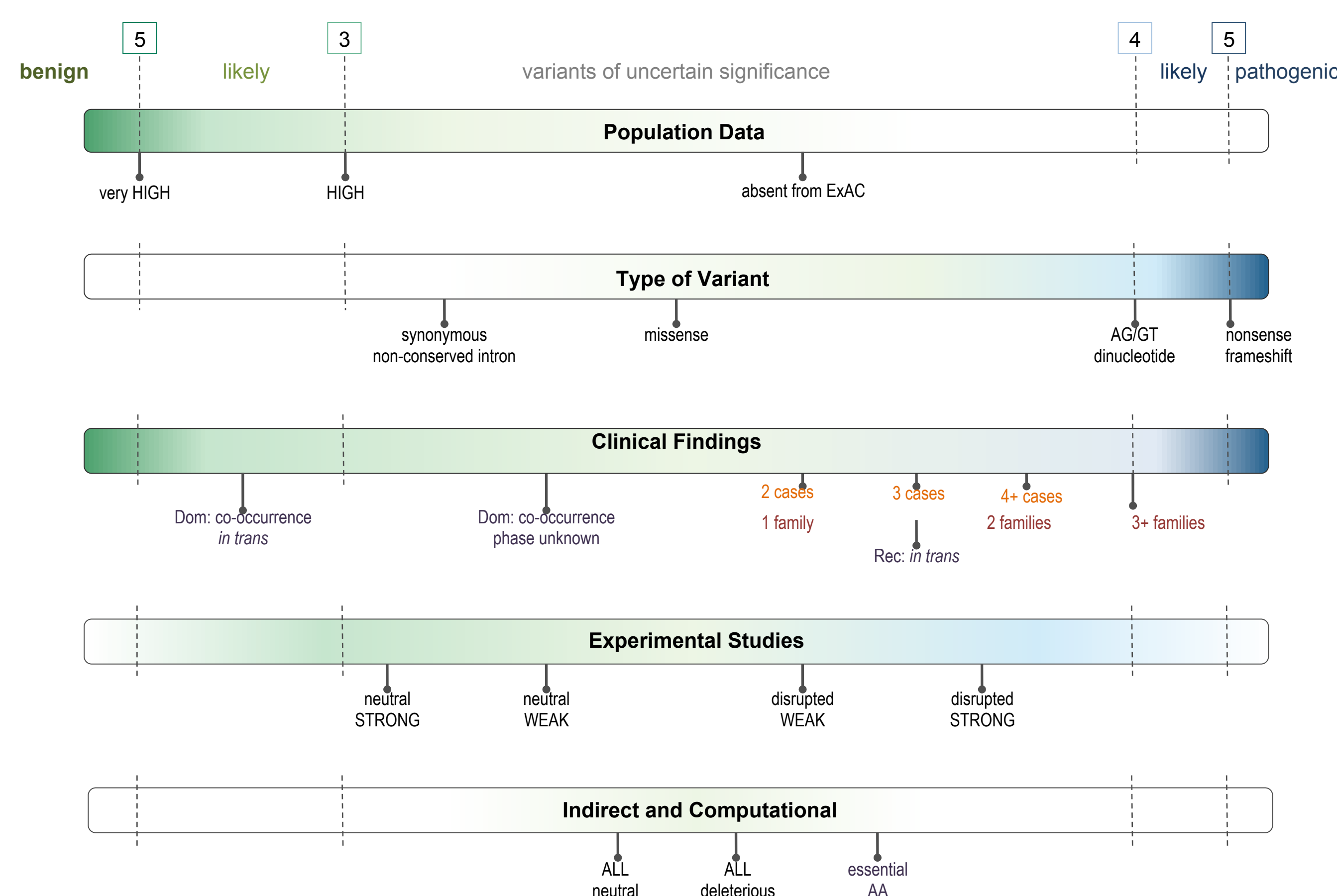


## Introduction

Access to large-panel genetic testing has steadily increased as improvements in technology have driven down the cost of generating genomic sequence data. This steady expansion of test menus has highlighted the need for high quality and cost-effective ways to provide a clear and thoughtful clinical report to clinicians. Communicating the aggregate evidence supporting a variant classification in a succinct, cohesive manner to clinicians and patients can be extremely challenging. To facilitate this process, we use a unique score-based evidence system for variant classification named Sherlock and created a framework for describing the relevant information in a clinical genomics report. We've categorized evidence into four main variant details (VD) sections: (1) a molecular description of the variant, (2) evidence that is directly applicable to the variant (e.g., population data, clinical, and functional data), (3) indirect and predictive evidence (e.g., computational analysis), and (4) a final summary. Informatics solutions have been developed by our team to support the automation of these VD sections, and metrics are routinely gathered to optimize text language and further increase writing efficiency. Importantly, these efforts reduce the time spent generating a clinical report, allow scientists to focus more time on variant interpretation, and remove a significant barrier to low-cost genetic testing.

## Sherloc

Sherloc is a hierarchical and integrative approach for aggregating and evaluating multiple lines of evidence into a consistent and confident variant classification. Based on the elements outlined in the 2015 ACMG/AMP ISV guidelines\*, we created highly detailed and specific evidence criteria (EVs) with point scores that reflect the relative strength of corresponding data. Higher scores equal stronger evidence and provide more weight to a definitive classification, either Pathogenic or Benign.



## What is the Variant Details System?

The variant details system was developed at Invitae to provide a structured framework for reporting the interpretation (pathogenic, variant of uncertain significance, etc.), of a given variant as well as describing all evidence used to classify the given variant. To support this structured framework, common text was developed to guide consistency in describing the evidence for variant classification across variants as well as variant interpreters.

Element	What is this?	Details
Element 1	Variant description	Factual description of the observed DNA change; no judgments
Element 2	Direct evidence	Overview: presence/absence in population databases and literature <ul style="list-style-type: none"> <li>Description of affected individuals in literature, including segregation data</li> <li>Presence in clinical databases</li> <li>Functional data</li> </ul>
Element 3	Indirect evidence	In silico predictions Important functional domains
Element 4	Summary	Summarize elements 2-3

## Results

We collected VD text usage metrics for a 12 month period with >21,000 accumulative data points. This data was used to evaluate bottlenecks in the system and identify areas of improvement in the description of variants and the evidence supporting their classification. We observed a decline in manual revisions of all sections. Indirect and predictive evidence was the least modified VD section, whereas population, clinical, and functional evidence was the most frequently modified section. An evaluation of the framework prompted a separation of population data, clinical evidence, and functional evidence into three independent descriptions, which further streamlined the reporting process. In addition, non-dynamic variables (i.e., gene name) were selected for automation to reduce manual data entry.

### Element 1 refinements after usage metrics analysis

(underlined variables are auto-generated with informatics support)

Text	"This sequence change creates a premature translational stop signal at codon <u>CODON NUMBER</u> ( <u>PROTEIN NOMENCLATURE</u> ) of the <u>GENE</u> gene. It is expected to result in an absent or disrupted protein product."
Text with input	This sequence change creates a premature translational stop signal at codon <b>143</b> ( <b>p.Glu143*</b> ) of the <b>BRCA2</b> gene. It has been shown to result in an absent protein product.

### Element 2 refinement after usage metric analysis

Element	Bullets	Details
Element 2	2A	ExAC population data, including the highest subpopulation frequency
	2B	Affected individuals from the literature (include PMIDs), segregation data, clinical databases if present (include ClinVar variation ID)
	2C	Functional data, if available (include PMIDs)

### Example of Variant Details Write-up (pre-analysis)

Element 1



#### KCNQ1, Exon 6, c.830C>T (p.Ser277Leu), heterozygous, PATHOGENIC

- This sequence change replaces serine with leucine at codon 277 of the KCNQ1 protein (p.Ser277Leu). The leucine residue is highly conserved and there is a large physicochemical difference between serine and leucine.
- This sequence change has been reported in the literature and is not currently found in any individuals from the population databases (rs199472730, no frequency). It has been shown to segregate with long QT syndrome in multiple families (PMID: 21895724, 21241880, 12442276) and has been identified in several unrelated individuals with isolated long QT syndrome (PMID: 19716085).
- Experimental studies have shown that this missense change leads to a non-functional KCNQ1 protein and acts in a dominant-negative manner to reduce the activity and reduces the surface localization of normal KCNQ1 protein (PMID: 21241800, 21895724).
- This sequence change is absent from population databases, has been shown to segregate with long QT syndrome in several families, and leads to a deleterious effect on protein activity. For these reasons, this sequence change has been classified as Pathogenic.

Element 2

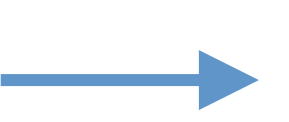


Element 4



### Example of Variant Details Write-up (post-analysis)

Element 1



Element 2



Population data  
Clinical information in literature  
Functional data

#### KCNQ1, Exon 6, c.830C>T (p.Ser277Leu), heterozygous, PATHOGENIC

- This sequence change replaces serine with leucine at codon 277 of the KCNQ1 protein (p.Ser277Leu). The leucine residue is highly conserved and there is a large physicochemical difference between serine and leucine.
- This variant is not present in population databases (rs199472730, no frequency).
- This variant has been shown to segregate with long QT syndrome in multiple families (PMID: 21895724, 21241880, 12442276) and has been identified in several unrelated individuals with isolated long QT syndrome (PMID: 19716085). ClinVar contains an entry for this variant (Variation ID: 53116).
- Experimental studies have shown that this missense change leads to a non-functional KCNQ1 protein and acts in a dominant-negative manner to reduce the activity and reduces the surface localization of normal KCNQ1 protein (PMID: 21241800, 21895724).
- This sequence change is absent from population databases, has been shown to segregate with long QT syndrome in several families, and leads to a deleterious effect on protein activity. For these reasons, this sequence change has been classified as Pathogenic.

## Informatics

Informatics incorporation does not result in the complete automation of a clinical report. Utilizing informatics in the report is a crucial aspect to improve:



Within this current framework, further development of Element 1 for more complex variant types (i.e., frameshift, splicing, gross deletions, etc.) may reduce manual entry of this information. In addition, specific portions of the variant details involve summarizing information found in databases (public and private). Element 2 is specifically broken down into three subcategories to streamline incorporation of informatics processing where data input is needed. Complete automation of Element 2 is not always possible as scientific review of the literature is required to provide a comprehensive and accurate interpretation of the variant.

## Conclusions

Our analysis provides evidence that improving the efficiency of reporting variant information is a compelling need in genomics. Expansion of test menus, introduction of new databases into the variant interpretation process, and changes to the VD workflow all create influx in user customization. Furthermore, variant evidence will continue to require updates as we learn more about various genes and as new literature becomes available. When components of the variant details workflow are incorporated into informatics planning, a decrease in the frequency of modified text across various areas can be appreciated. This optimizes the reporting process of each variant as the analysis and description of data can then be reproducible across variant interpreters.

## References and Acknowledgements

- \*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. (PMID: 25741868)
- Clinical Genomics Group, Invitae
- Genetics Development Group, Invitae