**Introduction**

Recently updated guidelines from the American College of Medical Genetics and Genomics (ACMG) provide a detailed framework for using evidence-based criteria for the interpretation of sequence variants identified through clinical genetic testing. Phenotype information is routinely used to determine the significance of rare variants identified in the course of genetic testing. Determinations are made by either the laboratory performing the testing or the physician collecting laboratory results and other evidence to make a final diagnosis. The use of such information is not yet standardized, and ACMG guidelines generally state that when tested genes have high clinical sensitivity and relatively little benign variation, phenotype information should be used as supporting evidence if it is highly specific for a disorder. This information may influence the prior probability that a rare variant found in a gene associated with a specific disorder has clinical significance. However, there is rational concern that this approach introduces the risk of circular reasoning and type I errors—i.e., wrongly concluding that a rare variant identified in a gene must be pathogenic because it was found in an individual with symptoms compatible with the disorder associated with that gene. This risk increases as the specificity of clinical findings in a disorder decreases. We investigated the use of phenotype ontologies to identify sets of phenotypic features that together establish high specificity to evaluate whether such scores could be useful in minimizing the risk of type I errors while still incorporating phenotypic information into variant assessment.

**High phenotype Annotation Sufficiency scores (AS) have the highest average rate of positive molecular findings**

To investigate the utility of phenotype Annotation Sufficiency (AS) and Phenomizer scores in the process of variant interpretation, we first tested the hypothesis that disorders with high AS scores are highly specific and have high prior probabilities of being diagnosed correctly. We examined the distribution of AS scores across all disorders in the Monarch database and created three groups according to AS score: high (≥0.79), mid-range (0.64 < AS < 0.79), and low (≤0.64).

**Deep dive into high-AS conditions**

We next examined phenotypic information provided by referring clinicians for individuals with disorders that have high AS scores. Our goal was to test the validity of Phenomizer scores for predicting a positive molecular diagnosis. However, for the majority of cases, the reported clinical information was insufficient, thus reducing the specificity of the diagnosis and obscuring the relevance of observed rare variants. We use CHARGE syndrome as an example below.

**Using phenotype to support variant claims**

The ACMG standards and guidelines* for the interpretation of sequence variants (Richards, et al., 2015) recommend that a patient’s phenotype can be considered supporting evidence if the following criteria are met:

- **1.** The clinical sensitivity of testing should be high.
- **2.** The gene should not be subject to substantial benign variation.
- **3.** The family history should be consistent with the mode of inheritance.
- **4.** The patient should have a well-defined syndrome with little overlap with other clinical presentations.

In previous studies, we have successfully incorporated evidence that takes into account clinical sensitivity, variant frequency and the mode of inheritance into our variant classification schema known as Sherloc (see ACMG Poster #592; Murillo, et al., 2017). However, accurately quantifying phenotypic specificity has proven to be much more difficult and, as a result, is highly susceptible to individual bias and type I errors.

The Human Phenotype Ontology (HPO) and the Monarch Initiative have established useful mechanisms for linking phenotypes with cognate genes, and together they provide a basis for integrating phenotype information into variant interpretation. Specifically, the Monarch Initiative has developed “annotation sufficiency” (AS) scores to provide a measure of the depth of available phenotypic features for a given disorder, whereas the HPO-driven Phenomizer tool uses available phenotype information to produce confidence scores on the likelihood that the information represents a specific annotated genetic disorder.

**Using phenotype to support variant claims**

In previous studies, we have successfully incorporated evidence that takes into account clinical sensitivity, variant frequency and the mode of inheritance into our variant classification schema known as Sherloc (see ACMG Poster #592; Murillo, et al., 2017). However, accurately quantifying phenotypic specificity has proven to be much more difficult and, as a result, is highly susceptible to individual bias and type I errors.

The Human Phenotype Ontology (HPO) and the Monarch Initiative have established useful mechanisms for linking phenotypes with cognate genes, and together they provide a basis for integrating phenotype information into variant interpretation. Specifically, the Monarch Initiative has developed “annotation sufficiency” (AS) scores to provide a measure of the depth of available phenotypic features for a given disorder, whereas the HPO-driven Phenomizer tool uses available phenotype information to produce confidence scores on the likelihood that the information represents a specific annotated genetic disorder.

**Conclusions**

Our results show that the average diagnostic yield was highest for genes associated with conditions with high AS scores, suggesting that the AS scores can be used to identify genes associated with highly specific phenotypes. The next step is to determine whether Phenomizer scores can be used to accurately predict which patients have a positive molecular diagnosis based on their reported phenotypes.

Our study highlights the potential value of specific clinical information from referring clinicians for gene panels and exome sequencing. A critical issue that we hope to address in future evaluations is the weight (if any) that should be granted to phenotype information within a variant interpretation process that systematically considers several lines of evidence.