Importance of phenotype in exome variant analysis

Dhanya Lakshmi N and Shubha R Phadke

Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow

Email: shubharaophadke@gmail.com

Mutation detection is the gold standard for the diagnosis of monogenic disorders. The advent of next generation sequencing strategies has revolutionized diagnostics in genetics. The whole genome and whole exome can now be sequenced with ease and at an affordable cost. For the disorders caused by large sized genes or with etiological heterogeneity, this is a great boon. Enormous data is generated which needs expertise for correct interpretation. The challenge is to correctly identify the causative sequence variation from the thousands of variations identified in each individual.

There are various methods which are employed for filtering the variants obtained and these include a series of filters designed to remove the low quality and common variants and non-pathogenic variants which are defined as the variants in non-coding, non-splice site regions and synonymous or missense mutations. Subsequently the best candidate gene correlating with the disease is chosen. But this strategy of identification of a deleterious variant purely based on the sequence variant pathogenicity will not help in correct identification of the disease-causing mutant gene in very rare diseases and diseases with unknown etiology. This is why the idea of using phenotypic data for prioritizing the variants has emerged. In this article we review the various phenotype-based tools used to analyze data of genetic variants obtained by next generation sequencing based technology and we wish to stress upon the importance of correct delineation and description of phenotype as the first step in genomic analysis.

Human Phenotype Ontology (HPO)

The phenome is defined as the set of all phenotypic traits of an organism. The variation in phenotype needs to be described in a systematic human and computer interpretable form. Standard measures for capturing phenotypic abnormalities are needed.[1] With this aim, the Human Phenotype Ontology (HPO) was initiated in 2007, with over 8000 terms, representing human phenotypic abnormalities and annotating all clinical entries in the Online Mendelian Inheritance in Man (OMIM) database with the terms of HPO.[2] It has 10,088 classes and 13,326 subclass relationships between these classes.[3] The annotated terms in HPO follow the true path rule by which a disease directly annotated to a term implicitly implies annotation to all the previous terms.

The HPO has three independent sub ontologies which include the mode of inheritance, the onset and clinical course and phenotypic abnormalities. Each of the HPO classes has a unique identifier, a label and a list of synonyms. More than 65% of the classes have a textual definition. HPO classes have cross references to other resources like Disease Ontology, Unified Medical Language System, Medical Subject Headings and International Classification of Diseases 10th revision.

The Human Phenotype Ontology includes a wide range of phenotypic abnormalities including morphological abnormalities, abnormal processes and abnormal investigations. (Table1)[3]

Tools using HPO

The important tools used in differential diagnosis and exome analysis which uses HPO are Phenomizer and Exomiser. Other clinical databases and analysis tools like PhenoTips, DECIPHER and Cartagenia also use HPO.

- Phenomizer: Phenomizer is an online tool that is freely available at http://compbio.charite.de/phenomizer/.[3] The Phenomizer allows searching for various phenotypes and prior-

Type of abnormality	Examples from HPO
Morphological abnormality	2,3 finger syndactyly (HP:0001233)
Abnormal process(organ)	Hyperacusis (HP:0010780)
Abnormal process (cellular)	Abnormal glucose homeostasis (HP:0011014)
Abnormal laboratory finding	Ketonuria (HP:0002919)
Abnormal imaging	Aplasia or hypoplasia of cerebellar vermis (HP:0006817)

Table 1 Phenotypic abnormalities listed in the Human Phenotype Ontology. (Adapted from Kohler et al.[3])

itizing exome variants. The search data can be downloaded in PDF format. In a study by Kohler et al. in 2009, the performance of Phenomizer was evaluated by developing a testing scenario based on "simulated patients" presenting with features of one of the 44 known genetic syndromes.[4] It was shown that ontological approaches performed better than diagnostic algorithms based on exact matching of terms in a phenotypic feature vector.

- PHIVE (Phenotypic Interpretation of Variants in Exomes)[5]: This software allows prioritization of variants by integrating the phenotypic similarity between human diseases and genetically altered mouse models. There are phenotyped mouse mutant orthologs for 4836 protein coding genes. By using this method, a correct gene hit was observed in 83% cases. This tool is freely available at `http://www.sanger.ac.uk/resources/databases/exomiser`.

- Phevor: (Phenotype Driven Variant Ontological Re ranking Tool)[6]: This software is freely available online and allows uploading the variants in vcf format and entering the phenotypic description using terms from HPO. (`http://weatherby.genetics.utah.edu/cgibin/Phevor/PhevorWeb.html`). Phevor works by combining the output of various variant prioritization tools along with the information contained in various Ontologies including the Human Phenotype Ontology, the Mammalian Phenotype Ontology, the Disease Ontology and Gene Ontology. The phenotypic information is propagated across and between various ontologies and this helps in accurately reprioritizing variants identified by the variant prioritization tools. Phevor is not a substitution for any of the available variant prioritization tools, rather it improves the performance of every tool. Singleton et al. in 2014 showed that post prioritization by using Phevor improved the variant prioritization in autosomal dominant as well as recessive diseases.

To conclude, phenotypic data plays an important role in translational bioinformatics. Phenotype should be systematically captured and used for exome variant analysis and the various tools available should be used appropriately to identify disease causing variants.

References

1. Biesecker LG. *Nat Genet* 2004; 36: 323-4.
2. Robinson PN, et al. *Am J Hum Genet* 2008; 83: 610-5.
3. Kohler S, et al. *Nucleic Acids Res* 2014; 42 (Database issue): D966-74.
4. Kohler S, et al. *Am J Hum Genet* 2009; 85: 457-64.
5. Robinson PN, et al. *Genome Res* 2014; 24: 340-8.
6. Singleton MV, et al. *Am J Hum Genet* 2014; 94: 599-610.