Applied & Translational Genomics 3 (2014) 60–67

Contents lists available at ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Medical genomics: The intricate path from genetic variant identification to clinical interpretation

B. Quintáns [a,b], A. Ordóñez-Ugalde [a,c], P. Cacheiro [a,c], A. Carracedo [a,b,c], M.J. Sobrido [a,b,*]

[a] Fundación Pública Galega de Medicina Xenómica and Instituto de Investigación Sanitaria, SERGAS, Santiago de Compostela, Spain
[b] Centro para Investigación Biomédica en red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Spain
[c] Universidade de Santiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 2 June 2014
Accepted 2 June 2014

Keywords:
Genetic variant
Interpretation challenge
Pathogenicity assessment
Clinical significance
Informatics pipeline
Family co-segregation

ABSTRACT

The field of medical genomics involves translating high throughput genetic methods to the clinic, in order to improve diagnostic efficiency and treatment decision making. Technical questions related to sample enrichment, sequencing methodologies and variant identification and calling algorithms, still need careful investigation in order to validate the analytical step of next generation sequencing techniques for clinical applications. However, the main foreseeable challenge will be interpreting the clinical significance of the variants observed in a given patient, as well as their significance for family members and for other patients.

Every step in the variant interpretation process has limitations and difficulties, and its quote of contribution to false positive and false negative results. There is no single piece of evidence enough on its own to make firm conclusions on the pathogenicity and disease causality of a given variant.

A plethora of automated analysis software tools is being developed that will enhance efficiency and accuracy. However a risk of misinterpretation could derive from biased biorepository content, facilitated by annotation of variant functional consequences using previous datasets stored in the same or linked repositories. In order to improve variant interpretation and avoid an exponential accumulation of confounding noise in the medical literature, the use of terms in a standard way should be sought and requested when reporting genetic variants and their consequences. Generally, stepwise and linear interpretation processes are likely to overrate some pieces of evidence while underscoring others. Algorithms are needed that allow a multidimensional, parallel analysis of diverse lines of evidence to be carried out by expert teams for specific genes, cellular pathways or disorders.

Contents

1. Introduction

Next generation sequencing (NGS) technologies are rapidly becoming a routine tool in the diagnostic workup of patients with diverse

* Corresponding author at: Neurogenetics Group, Fundación Pública Galega de Medicina Xenómica, Hospital Clínico de Santiago, Travesía da Choupana s/n, 15706 Santiago de Compostela, Spain.
E-mail address: ssobrido@telefonica.net (M.J. Sobrido).

http://dx.doi.org/10.1016/j.atg.2014.06.001

conditions, including tumor profiling. Medical genomics refers to the ability to simultaneously query the diagnostically relevant gene set of a given person for clinical decisions. Sequencing of the complete set of protein coding exons of an individual — whole exome sequencing (WES) — has enhanced the identification of genetic defect of rare diseases (Wan et al., 2012; Keller et al., 2013). These technologies can also be applied to decipher more common syndromes (Cirulli and Goldstein, 2010; Kiezun et al., 2012). Decision-making in oncology can now be based on the singular molecular signature of the tumor with implications in pathology and response to treatment or individual prognosis (Normanno et al., 2013). Another approach to the diagnosis of genetically heterogeneous disorders is the simultaneous sequence of a panel of genes associated with a given syndrome. NGS also harbors potential to delineate an individual's pharmacogenetic profile (Patrinos et al., 2013). The use of high throughput molecular analysis for clinical decision making is often referred to as personalized medicine or personal genomics, although warnings have also been raised about myths and inflated expectations that may come along with these somewhat blurry terms (Salari et al., 2012).

How far are we still from being able to interpret all genetic variations accurately in a clinical context? Many challenges lie ahead before NGS can be integrated as part of routine medical care. The process to know which one among the thousands of genetic variants harbored within an individual's genome is clinically relevant generally involving a number of steps summarized in Fig. 1. In the following sections we review some of the challenges and limitations encountered along this path, as well as potential sources of errors that must be taken into account for an adequate clinical interpretation of genetic variants.

2. Need for accurate use of terms on genetic variations and their consequences

A first source of difficulty comes from the imprecise use of vocabulary referred to genetic variations and their consequences. The terms polymorphism and mutation do not bear implications on their functional consequences, however they are often used with that meaning. A polymorphism is a genetic variant present in ≥1% of the population, whereas a mutation is any change in the DNA sequence compared to the previous state or wild type. Neither concepts imply whether they are or are not disease-causing. Just because a polymorphism is not so rare, it does not necessarily mean that it is benign (not associated with a disorder) or neutral (without functional consequences). Because of the potential for misinterpretation of polymorphism and mutation, the term genetic variant is currently favored, as defined by the presence of a particular allele — at a nucleotide position, gene or locus — that is not the most commonly encountered allele in the general population. Thus, the term genetic variant does not imply any a priori assumption on the frequency of the variant allele or its potential effect on the health of the individual carrying it. Also, terms such as neutral, benign, functional, pathogenic, deleterious, damaging, disease-associated and causal, when referring to a genetic variant, are often used in ill-defined manner throughout the medical literature. For instance, pathogenic is often equaled to disease-causing, which is not necessarily always the case. While functionality, deleteriousness, pathogenicity and disease causality may be strongly related terms, they are not interchangeable. As for the term phenotype, it must be specified whether the authors mean abnormalities detectable at a cellular/organ level, to biochemical alterations that can be measured, or to abnormal clinical traits that can be observed in an individual, animal model or cellular construct. A phenotype can be made up of several endophenotypes that may provide useful clinical measures (Mann et al., 2009).

Another issue is the system level at which the consequence of a genetic variant is being described. For example, the variant may be deleterious at a cellular level (causes a loss of function in a given cellular process), but not necessarily deleterious for the organ or individual. When discussing the potential effects of a given genetic variant on disease, there is a tendency to classify the variant in a simple three or five-tiered scheme (pathogenic, likely pathogenic, unlikely pathogenic, non pathogenic, unknown). This scheme, however, ignores the complexity of biological processes that can imply other types of relationships between a variant and a clinical manifestation (predisposing, triggering, modifying, protective, etc.), as well as digenic or polygenic disorders. We call for using terminology – and requesting its use in scientific publications – with more precision when describing the consequences of genetic variants, such as done in the recent paper by MacArthur et al. (2014). While a consensus is developed by the genetics community on the definition of these terms and how they should be used, it would be a good practice that curators of genetic databases define their intended meaning.

3. Variant identification and annotation

The first step towards genetic variant interpretation is the ability to correctly determine the presence of, and subsequently annotate, the alleles at each position of the target sequence. Obviously, variants that have not been identified and annotated will not be subject to further

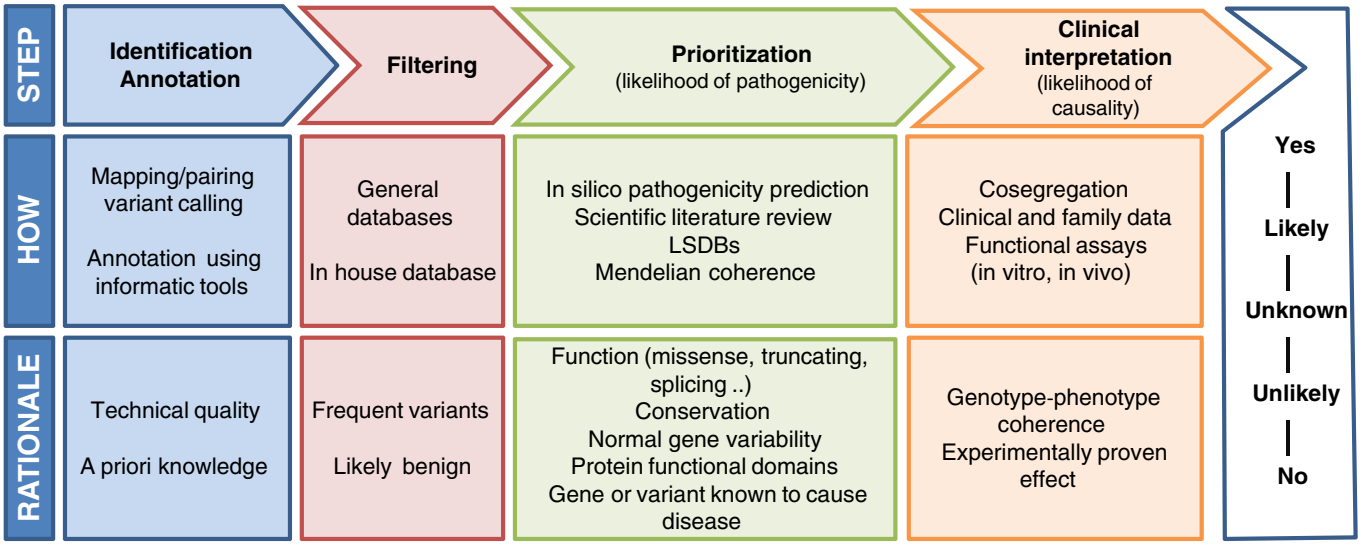

Fig. 1. Stepwise evidence pipeline for clinical interpretation genetic variants. After identification and automatic annotation, likely benign variants are filtered out and the remaining variants are prioritized. The weight of different lines of evidence leads to final clinical interpretation.