

Genomic Era: the Challenges Ahead

Editorial

The advent of genomic techniques like microarray and massively parallel sequencing has revolutionized the diagnosis of genetic disorders. Massively parallel sequencing or next generation sequencing (NGS) which can sequence many genes or the exome and even the whole genome in one go, has been a great boon for single gene disorders. It is very useful for phenotypes which can be caused by many genes and where there is no genotype-phenotype correlation, as is the case usually. NGS based testing for a panel of genes or exome sequencing has become a first tier testing for many monogenic disorders in the clinical settings. The large sizes of some genes, genetic heterogeneity and lack of a definitive clinical diagnosis are no longer problems. NGS -based testing can detect the causative mutation(s) in 30 to 50% of cases in various clinical scenarios and thus is a diagnostic test with great clinical validity. This is possible due to massively parallel sequencing. This virtue also has a flip side as it generates massive data of genetic variations in all the sequenced genes or in case of exome sequencing, in the entire exome. Each individual personal genome sequence still reveals 200,000–500,000 SNVs that have not been observed in other publicly available personal genomes, many of which may be unique to that individual's family or clan. Hence, fishing out the causative sequence variation from the huge data generated by NGS is a big task. The article about identification of Mendelian disease genes using NGS in this issue provides an overview of steps and strategies used in identification of the causative genetic mutation(s) from the massive data of genes / exome sequenced. Many of the 'disease causing mutations' identified are novel and their pathogenic nature needs to be supported by *in silico* analysis and family data about segregation of the mutation with disease-affected family members.

The identification of the causative mutation(s) for more and more patients with monogenic disorders is exciting for clinicians. In addition to confirmation of the clinical diagnosis, it is of great help for the families in genetic counselling,

pre-symptomatic diagnosis and prenatal diagnosis. However, the pathogenic nature of many 'mutations' continues to be uncertain and the label of 'likely pathogenic' is preferred for the novel variants without functional validation. It is thought that as more and more exomes get sequenced, the data about disease-causing mutations as well as variants present in 'normal individuals' i.e. polymorphic variants will expand. This will help in decision-making about the pathogenic nature about each nucleotide variation easier. Thus, it is hoped that as more and more of the genome gets annotated, the interpretation of genome variants will become easier.

Now, already the data of thousands and thousands of genomes are available in various laboratories in addition to databases like the 1000 Genomes, ExAC (Exome Aggregation Consortium), etc. The analysis of data of the genomes of probably 'normal' individuals in these databases shows that there are 20 to hundred or more sequence variations which appear to be damaging to the function of the protein in each individual. These include stop codons and frameshifts leading to stop codons. Many of them are even present in a homozygous state. This raises the question as to why these individuals do not manifest the disease. Some of these individuals with normal phenotypes and damaging sequence variations may express the disease phenotype which may not be mentioned in the data or might have mild manifestations and hence might be missed clinically. Late onset of the disease-symptoms may also be one of the causes of no apparent disease manifestations at the time of reporting, in mutation carriers. But incomplete penetrance as a cause of no manifestations in some mutation carriers for autosomal dominant disorders has been known to clinicians and geneticists for long. The non-penetrance is an all-or-none phenomenon and different from variable expressivity. This can be identified by skipped generation or by screening of extended family members by mutation testing. Data from exome and genome sequencing now show that non-penetrance may be more common

than perceived. Non-penetrance in autosomal recessive diseases has also been reported.

The causes of non-penetrance can be modifier genes, epigenetic effects, variable expressions from maternal and paternal alleles, etc. The modifier genes for some diseases like cystic fibrosis, beta thalassemia and Hirschsprung disease have been identified. Digenic diseases like Bardet-Biedl syndrome, modifier genes, microdeletion syndromes and polygenic disorders have blurred the artificial boundaries between monogenic disorders, genomic/ chromosomal disorders and multifactorial disorders. Same is true for recessive and dominant disorders. No gene functions in isolation and therefore, in reality there is no actual 'single gene' disorder. All the so called pathogenic mutations for monogenic disorders act on the background of sequence variations and levels of functions of

other genes and hence the phenotype effect of the same mutation in different members of a family or in different families may be different and difficult to predict. This implies that the task of prediction of the pathogenicity of a sequence variation will continue to be a Herculean and very complex task.

Genomic techniques which have ushered in the era of genomic medicine have brought in new and exciting diagnostic tools, but at the same time they have also brought in new challenges which are different and much more complex than the traditional simplistic concepts of genetics.



Dr. Shubha R Phadke
1st October, 2015