The challenges in analyzing NGS data for molecular diagnosis of primary immunodeficiency diseases

Wanling Yang, Yu Lung Lau

Recent advances in next-generation sequencing (NGS) technology provide a cost-effective approach to large-scale resequencing of human genome for medical diagnosis. In recent years, whole exome sequencing (WES), the targeted sequencing of protein-coding regions, has become a powerful and widely-used tool for dissecting the genetic basis of Mendelian diseases. We have performed whole exome sequencing on nearly a hundred patients with primary immunodeficiency disease (PID) and made molecular diagnosis for more than half of these patients. NGS also showed superb power by detecting the causal mutations in cases that were missed by Sanger sequencing technology. In addition, the new sequencing platform allows us to discover novel PID genes and in certain cases, the molecular diagnoses prompted us to revisit the clinical phenotypes. Thus, NGS technology has become an integral part of clinical diagnosis of Mendelian diseases, particularly PID.

However, many challenges remain. Nearly half of the patients cannot reach definitive molecular diagnosis by WES. There are likely a number of different reasons for the failure in finding the causal mutations. Certain regions of the genome are not covered well, especially for the regions with repeats and paralogous sequences. The short reads of NGS platform and the discontiguous nature of coverage of the genome make it difficult to detect structural mutations such as copy number variations, deletions and duplications, balanced translocations and inversions. Many of the causal genes may not be realized even when the mutations are detected and for known causal genes, mutations in non-coding regions, such as in the regulatory regions are poorly characterized and as a result, may not be realized as disease causal. Extreme genetic heterogeneity for PID and many other Mendelian diseases makes it difficult to draw references from different families.

Thus, improving coverage of the genome and improved analysis of NGS data, broader collaboration in studies of rare Mendelian diseases, building up of population genetic databases with population-specific genetic data from large number of samples, and functional characterization of the disease genes and regulatory elements are the needed steps to improve molecular diagnosis rate. In the meantime, targeted approach combining whole exome sequencing and sequencing of targeted whole genes might be a practical way of improving NGS success rate at the time being.