Genomic Sequencing: Overcoming Challenges to a Bright Future

Whole genome sequencing (WGS) and whole exome sequencing (WES), which sequences only the protein-coding regions of the genome, have already begun to transform clinical medicine. They are being used to home in on the causes of rare and undiagnosed genetic diseases, determine appropriate cancer treatments for a given tumor, and match drugs and doses to an individual’s genomic makeup. But as WGS takes on greater relevance in the clinic, it is increasingly important to consider the benefits and challenges of this technology.

Currently, the technology used for genome sequencing requires scientists to fragment the DNA into thousands of small pieces—“short reads”—that are then sequenced in parallel. After aligning the fragments to a human reference sequence, algorithms determine the patient’s consensus sequence. Next, scientists compare the patient’s DNA to the human reference sequence using a variety of computational tools that vary widely in their speed, strengths and limitations. This “variant calling” provides a list of the 3 to 3.5 million positions where individuals differ from the reference, with about 100,000 of these variants being very rare or novel.

Various aspects of each step in the process generate downstream consequences. First, short fragments can be misaligned when they exactly match more than one genomic region. Second, the ability to identify genetic variants is highly dependent on the depth of coverage, or the number of sequence reads that line up at each position in the genome. Compared with WES, WGS is generally expected to provide improved coverage of certain genomic regions, such as introns and other noncoding regions that are associated with disease risk and drug response. However, in a recent study published in JAMA,1 we found that while WGS coverage is fairly high, there is still incomplete coverage of some important inherited disease genes. Finally, while variant calling algorithms often reliably identify single nucleotide variants, we and others have found that they are less consistent when it comes to identifying insertions, deletions, and larger variations (i.e., copy number variants or structural variants). This is a notable limitation, as these types of variants are often particularly important in genetic diseases.

Determining which variants matter for disease risk is also nontrivial. In the JAMA paper, we used automated variant annotation to help prioritize variants most likely to be impactful. We found that 50 to 100 variants per person (fewer in the undiagnosed diseases context) typically merit manual review to determine their implications for disease. Manually evaluating these candidate variants takes an average of 50 minutes of curation time. One challenge is that available information is often conflicting or limited. For instance, while specific variants may be present in variant databases, several studies have found that these databases contain high error rates, with up to 25 percent of variants incorrectly categorized as disease causing when in fact they may be common benign variants.2

Each of these technical, computational and interpretation challenges is currently being addressed. Advances in sequencing technology, such as long read sequencing, should allow identification of larger types of variations while also reducing errors in alignment and assembly. Incomplete coverage of specific genomic regions can be targeted with orthogonal approaches. And improved curated variant databases will greatly assist with variant assessment and the interpretation bottleneck in clinical WGS.

WGS thus has a very bright future. Clinical WES has already demonstrated a diagnostic yield of approximately 30 percent—a sensitivity higher than many routinely used genetic tests.3 As sequencing becomes more accessible and reliable, knowledge of disease-gene relationships expand, and bioinformatics algorithms improve, our ability to interpret WGS in any context will rapidly advance.

NOTE:
The information presented represents the author’s own views and does not necessarily represent the views of Stanford Hospital and Clinics, Lucile Packard Children’s Hospital and/or Stanford University or its affiliates.

ACKNOWLEDGMENTS:
The author would like to thank Dr. Euan Ashley and Rachel Goldfeder for their helpful commentary on this editorial.

REFERENCES:
