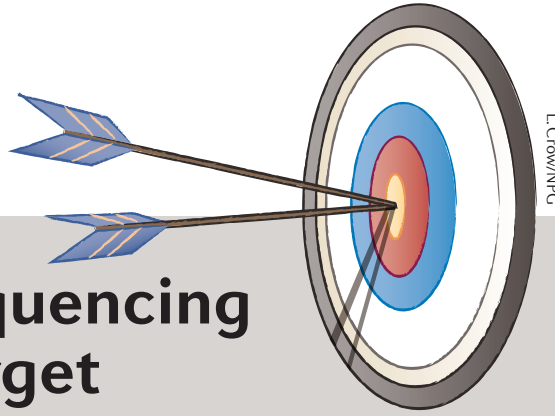


GENETIC TESTING

Clinical sequencing right on target



L. Crow/NPG

As genome sequencing moves from the research laboratory to the clinic, the accuracy of variant calls used in diagnoses is increasingly important. In a new study published in *Genome Medicine*, Goldfeder *et al.* reveal that many disease-associated genes and variants lie within areas of the genome that are prone to sequencing errors.

The team set out to assess the accuracy of variant calls generated using representative whole-genome sequencing (WGS) and whole-exome sequencing (WES) pipelines. To this end, they sequenced a standard reference genome that had previously been highly characterized through extensive resequencing by the Genome in a Bottle (GIAB) Consortium as a benchmark of human genome sequencing.

Comparison of WES and WGS variant calls with consensus calls from the reference assembly revealed the presence of false-positive and false-negative results in both datasets. For

WES data, false-negative variant calls were mainly attributable to inadequate sequencing depth, whereas those generated by the WGS pipeline were caused by over-stringent filtering, demonstrating the importance of fine tuning filters for data analysis. The team identified more than 39,000 loci in which at least one sequencing pipeline incorrectly called a variant, and almost 7,500 of these variants could be found in publicly available variant databases. One variant call identified as a false positive using the reference assembly was a recognized pathogenic frameshift in the *BRCA2* gene, which is implicated in hereditary cancer. This finding highlights the importance of well-characterized reference genomes and confirmatory testing in clinical sequencing.

Focusing their attention on clinically significant genes, the group calculated the proportion of exon bases within high-confidence areas of the reference genome

for 3,300 disease-associated genes from the ClinVar and OMIM databases. For almost 18% of these genes, less than half of the exonic DNA lay within high-confidence regions of the genome. Next, the team examined a set of 56 'medically actionable' genes, variants in which are recommended for clinical reporting by the American College of Medical Genetics and Genomics (ACMG). Almost 18% of the exonic sequences from these genes lay outside of high confidence areas of the genome, revealing a risk of false variant calls that could have medical consequences for patients.

Notably, analysis of the reference assembly revealed that high-confidence regions were enriched for unique and non-repetitive sequences, which can be explained by the exclusion of repetitive regions during high-confidence variant calling. The team conclude that the continued development of reference materials to incorporate more complex areas of the genome is essential to improve our understanding of the predictive and technical characteristics of clinical sequencing.

Denise Waldron

ORIGINAL ARTICLE Goldfeder, R. L *et al.* Medical implications of technical accuracy in genome sequencing. *Genome Med.* <http://www.dx.doi.org/10.1186/s13073-016-0269-0> (2016)