

## **Use of Pangenome reference improves variant calling in clinical genome sequencing**

Erica D Smith, Ermanno Florio, Julian Stone-Farhat, Matthew Schultz, Ali Alhafidh, David Brohawn, Marcus Mahar

### **Introduction:**

Assembly of genome short read sequences is most effective and efficient when aligned to a reference genome. The choice of which reference genome to use, however, can affect which variants can be detected and their resulting sequence quality, called reference bias. Iterations on these linear references improved quality and coverage of previously inaccessible regions. Previous publications have reported, however, hundreds of megabases of sequence in global populations that are not represented in linear reference genomes and are therefore missed in traditional genotyping. Using a Pangenome reference improves on linear reference genomes by allowing alignment of more contigs and elucidating previously uncharacterized alternative alleles. This approach has been shown to improve Structural Variant (SV) detection and genotyping accuracy in difficult-to-sequence regions in research settings.

### **Methods:**

To determine the value added from graph-based genotyping in clinical genome sequencing, we used DRAGEN to compare sensitivity, precision, and F1 scores between linear hg38 vs hg38 with Pangenome. We first benchmarked performance using Genome in a Bottle samples and cell lines from the expanded 1000 Genomes Project. We then evaluated performance on clinical specimens, comparing linear hg38 vs hg38 with Pangenome against cross-platform benchmarked long read genome sequencing data.

### **Results:**

Using pangenome enhanced sensitivity for insertions between 36-50bp and single nucleotide variant (SNV) detection in both benchmarking studies and in clinical specimens from people of all genetic ancestries. Performance improvements due to pangenome were more pronounced in downsampling experiments, suggesting that it will be particularly beneficial for specimens with low coverage and areas of poor mappability. Notably, the improved sensitivity was particularly apparent for heterozygous SNVs in easy-to-sequence regions of clinical specimens. Using a pangenome reference is therefore expected to be highly impactful in clinical rare disease diagnostics, as heterozygous SNVs are the most frequently reported variant type in exome.

## Conclusions:

These findings show that using pangenome reference can improve sensitivity and precision for multiple variant types in both people from underrepresented ancestries and in people with European ancestry. Broadly, this suggests that ongoing refinements to the standard linear reference genomes (eg hg19 to hg38) can be further enhanced by utilizing pangenome reference genomes.

## Ref:

Nyaga et al Front Genet. 2025 Sep 19;16:1679660. doi: 10.3389/fgene.2025.1679660

Liao et al Nature. 2023 May;617(7960):312-324. doi: 10.1038/s41586-023-05896-x