INVERSION BREAKPOINTS RESOLVED USING LONG-READ GENOME SEQUENCING

The group of Anna Lindstrand focused in recent years on the study of chromosomal rearrangements, aiming to resolve derivative chromosomes at the breakpoint level using short-read genome sequencing (srGS) and, increasingly, long-read genome sequencing (lrGS).

In a paper published in Genome Res. (1), they report on twelve inversion carriers: nine analyzed using long-read genome sequencing (lrGS) and three cases where short-read genome sequencing (srGS) data were reanalyzed. By aligning both lrGS and srGS data to multiple reference genomes, nine inversions were resolved (9/12, 75%). Notably, in four cases, one inversion breakpoint was located in a genomic region missing from at least one of the human reference genomes (GRCh37, GRCh38, or T2T-CHM13), requiring a reference-agnostic analysis that included T2T-CHM13. In some instances, de novo assembly was also necessary, as long-read mapping alone was insufficient.

The clinical relevance of these findings is underscored by a case where an inversion disrupted *EHMT1*, resulting in a diagnosis of Kleefstra syndrome 1. The discovery that four inversions could only be mapped using specific reference genomes led ~~us~~ them to investigate the presence and population frequencies of differential reference regions (DRRs) between T2T-CHM13, GRCh37, GRCh38, and the genomes of chimpanzees and bonobos. This investigation uncovered hundreds of megabases of DRRs.

Overall, the study highlights the critical role of reference genomes and the enhanced resolution that lrGS provides in the detection of structural variants.

1. https://genome.cshlp.org/content/early/2024/10/30/gr.279346.124.long