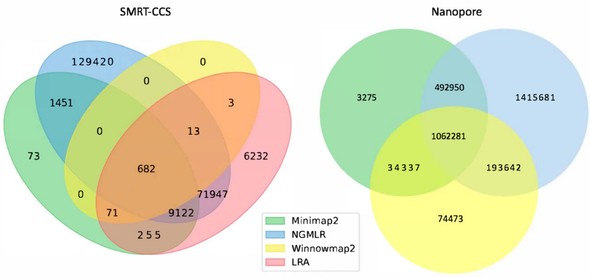
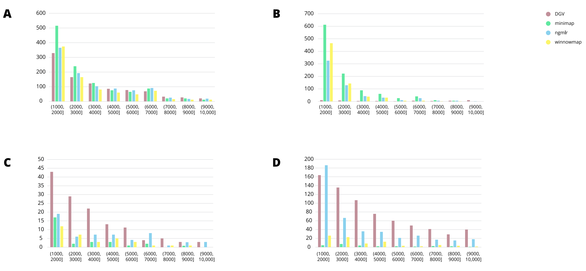
Benchmarking long-read genome sequence alignment tools for human genomics applications



Aligning long-read sequence data from humans to a reference is an informatic task that has become more important as access to the technology has increased. However, which reads to include and exclude in an alignment to resolve human variation remains a question. We find that using multiple alignment tools, and thus different combinations of reads, allows for the greatest number of first-pass variants to be discovered.



Regardless of the sequencing platform, and computational resources, it should be best practice to align an LRS human genome with three alignment tools. Minimap2, winnowmap2, and NGMLR will provide a strong foundation to gain better insight into the architecture of a genome of interest, but there are circumstances where use of LRA for SMRT data may make sense in lieu of NGMLR. When computational resources are limited, minimap2 is a strong choice, and when time is a limiting factor, winnowmap2 is the best choice.