Bivartect : accurate and memory-saving breakpoint detection by direct read comparison

We developed Bivartect (bit-based variant detection), a simple yet accurate computational approach to detecting genomic variants based on direct comparison of sequence reads, which skips initial mapping. To reduce memory use and attain a speed-up, Bivartect converts all sequences into bit strings, and keeps only a small part of the suffixes of the reads in the memory space during identification of breakpoints. Moreover, we adopted a strategy where part of the suffixes of the normal and mutated reads with a common prefix are sorted to detect potential breakpoints. Bivartect can detect not only single nucleotide variants but also insertions/deletions, inversions and their complexes. Bivartect achieves high predictive performance with an elaborate memory-saving mechanism, which allows Bivartect to run on a computer with a single node for analyzing small omics data. Tests with simulated benchmark and real genome-editing data indicate that Bivartect was comparable to state-of-the-art variant callers in positive predictive value for detection of single nucleotide variants, even though it yielded a substantially small number of candidates. These results suggest that Bivartect, a reference-free approach, will contribute to the identification of germline mutations as well as off-target sites introduced during genome editing with high accuracy.

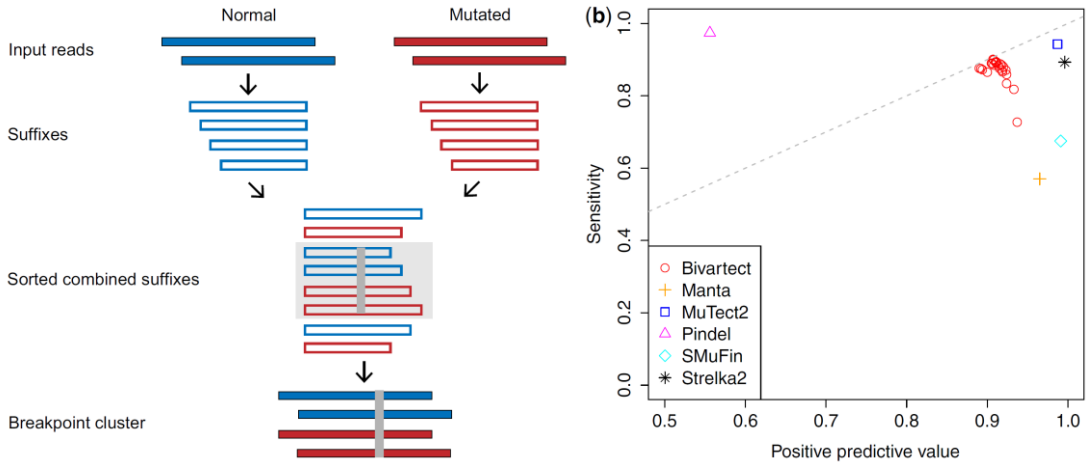


Fig. K.Shimmura et al, Bioinformatics, 2020