[BIOINFO 2023]

Germline indel calling performance is highly associated with indel property and parameter optimization

Mooyoung Kim[1], Dongchan Yang[1] and Inkyung Jung[1,*]

[1] *Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea*

**Corresponding author: ijung@kaist.ac.kr*

Insertions and deletions (indels) are a major source of genomic variations. Indels have clinical implications in various human diseases and also individual-specific genome function. However, unlike single nucleotide polymorphisms (SNPs), the accurate identification of indels remains challenging. Despite efforts for systematic evaluation of indel calling performances using various tools, the effect of indel properties or library status on detection accuracy is not well characterized, yet. To gain a deeper understanding of how these conditions influence the indel calls and to identify the most optimal variant detection method, we evaluated performance of four widely used indel variant callers with a focus on detection accuracy and reproducibility. For these, in this study, whole genome sequencing from NA12878 platinum genomes data was used for detection accuracy and identical twins data for biological replicates are produced by two different sequencing platform to investigate the reproducibility of each caller. Overall, both Strelka2 and Mutect2 showed great performance, but each caller has specific conditions that outperform other callers, such as improved performance of VarDict and VarScan2 in homozygous genotype and the high sensitivity of VarDict. Moreover, parameter optimization significantly improves the performance of all indel callers especially in VarDict and VarScan2. Our systematic evaluation of various sequencing and factors on the performance of indel calling provides a best practice to accurately detect indels from patient samples at various experimental conditions.