A robust benchmark for evaluating and improving mosaic variant calling strategies

Yoo-Jin Ha, Jisoo Kim, Seungseok Kang, Junhan Kim, Se-Young Jo, and Sangwoo Kim*

Abstract

The rapid advances in sequencing and analysis technologies have enabled the accurate detection of diverse forms of genomic variants, including germline, somatic, and mosaic mutations. However, unlike for the former two mutations, the best practices for mosaic variant calling still remain chaotic due to the technical and conceptual difficulties faced in evaluation. Here, we present our benchmark of nine feasible strategies for mosaic variant detection based on a systematically designed reference standard that mimics mosaic samples, with 390,153 control positive and 35,208,888 negative single-nucleotide variants and insertion–deletion mutations. We identified the condition-dependent strengths and weaknesses of the current strategies, instead of a single winner, regarding variant allele frequencies, variant sharing, and the usage of control samples. Moreover, feature-level investigation directs the way for immediate to prolonged improvements in mosaic variant calling. Our results will guide researchers in selecting suitable calling algorithms and suggest future strategies for developers.