Improving variant accuracy with Copy number variant pipeline for target sequencing

George de Vasconcelos Carvalho Neto[1], Wilder Barbosa Galvao[1], Marcel Caraciolo[1], Rodrigo Bertollo[1], Joao Bosco Oliveira[1],

1 Genomika

Abstract

Studies comparing human genomes have been shown that more base pairs are altered as a result of structural variants (SVs), including copy number variants (CNVs), than as result of point mutations. Structural variants were first defined as insertions, deletions and inversions greater than 1 kb size. However, with the high-throughput sequencing becoming a routine for genome analysis, the spectrum size of SVs and CNVs have been extended to events >50 bp in length. Due to the cost and complexity of analyzing whole genome sequence data, target sequencing (TS) has become the major approach for clinical diagnostic purposes. Also, TS allows for the detection of CNVs in addition to single-nucleotide variants (SNVs) and small insertions/deletions (INDELs). More recently, novel tools have been developed to improve CNVs identification from targeted panel sequencing data. With efforts conducted to develop an internal protocol for CNVs detection in TS data, and increase possible genetic case elucidation, the aim of this study was to analyze the performance of state-of-the-art CNV detection methods on targeted next-generation sequencing (NGS) before implementation at our laboratory. The chosen softwares to the analyses were ExomeDepth, panelcn.MOPS, CoNVaDING and VisCap. Based on samples from patients that sequenced BRCA1 gene and had variant confirmed by Multiple ligation probe assay (MLPA), which represents the gold standard for molecular analysis of diseases caused by CNV, is possible to evaluate accuracy and sensitivity for each tool. Finally, in the end of this evaluation we expect to maximize our variant detection accuracy using the best algorithm tested or the combination of the best callers for CNV on target sequencing data.

Funding: Genomika Diagnósticos