A comparative study of copy number variation detection methods for next-generation sequencing technologies

Yu-Ping Wang*[1,2] and Junbo Duan[1]
[1]Tulane University
[2]Shanghai University for Science and Technology

Copy number variation(CNV) has played an important role in the studies of susceptibility or resistance to complex diseases. Traditional methods such as fluorescence in situ hybridization (FISH) and array comparative genomic hybridization (aCGH) suffer from low resolution of genomic regions. Following the emergence of next generation sequencing (NGS) technologies, CNV detection methods based on the short read data have been developed recently. However, the performances of these methods have not been fully studied. To help investigators choose their suitable CNV detection methods for their specific tasks, comparative studies are needed.

We compare five publicly available CNV detection methods: CNV-seq, FREEC, readDepth, CNVnator and SegSeq. They are evaluated both on simulated data and real data with different experiment settings. In the simulated data processing, receiver operating characteristic (ROC) curve and box plot are employed to compare their performances in terms of breakpoint detection and copy number estimation. In the real data processing, Venn diagram and F-score are employed to show the quality of detected CNVs. The computational demands are also studied. The results of our comparative studies show the strength and weakness of each CNV detection method under a set of conditions, which is of great help for biological investigators to make informed decisions when choosing a suitable detection method for their specific needs.