SVDetect - FREEC: two complementary tools for the detection of genomic structural variants from deep sequencing data

Bruno Zeitouni[1,2,*], Valentina Boeva[1,2], Andrei Zinovyev[1,2], Kevin Bleakley[3], Isabelle Janoueix-Lerosey[1,4], Sophie Loeillet[1,5], Patricia Legoix-Ne[1], Alain Nicolas[1,5], Olivier Delattre[1,4], Emmanuel Barillot[1,2]

[1] Institut Curie, 26 rue d'Ulm, Paris, F-75248 France

[2] Inserm U900, 26 rue d'Ulm, Paris, F-75248 France ; Mines ParisTech, Fontainebleau, F-77300 France

[3] INRIA Saclay Ile-de-France, 4 rue Jacques Monod, Orsay, F-91893 France

[4] Inserm U830, 26 rue d'Ulm, Paris, F-75248 France

[5] CNRS, UMR3244, Paris, F-75248 France

*corresponding author: bruno.zeitouni@curie.fr

Keywords Cancer, Structural variants, Deep sequencing, Paired-ends, Copy-number profiles

Abstract

The detection of structural variants (SVs) in the human genome plays an important role in the understanding of many genetic diseases, including cancer. With the arrival of new high-throughput sequencing technologies, our current power to detect SVs has significantly improved. Here, we report two complementary methods for the prediction of structural variants from read mapping data provided by current short read aligners.

The first method detects clusters of paired-end/mate-pair reads anomalously mapped to the reference genome and uses all the characteristics of reads inside the clusters (orientation, order and clone insert size) to identify the SV type. This method, implemented in a program called SVDetect [1], allows identifying a large spectrum of rearrangements including large insertions-deletions, duplications, inversions and balanced/unbalanced intra/inter-chromosomal translocations.

The second method uses a different approach based on the detection of copy-number alterations from read depth-of-coverage data. Implemented in a program called FREEC (control-FREE Copy number caller) [2], it normalizes raw copy number profiles either using the profile from a control dataset or by the local GC-content if the control sample is not available. FREEC applies a LASSO-based segmentation procedure to the normalized profile and predicts genomic regions of gain and loss.

The use in parallel of the two SV detection strategies improves the confidence of calling true positive rearrangements and facilitates their interpretation at the genome scale. Both tools are compatible with the SAM alignment format as input file and provide output files for a graphical visualization of predicted structural variants.

Acknowledgements

This work was supported by The Ligue Nationale contre le Cancer.

References

[1] Zeitouni B *et al*. (2010) SVDetect - a bioinformatic tool to identify genomic structural variations from paired-end next-generation sequencing data, Bioinformatics, 26: 1895-1896

[2] Boeva V *et al*. (2011) Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization, Bioinformatics, 27(2):268-9.