MobiCNV : a simple and rapid method for detecting Copy Number Variants in Illumina gene panel, clinical exome and exome data

Henri Pégeot[1], Charles Van Goethem[2], Gema Garcia-Garcia[1], Reda Zenagui[1], Delphine Lacourt[1], Olivier Ardouin[1], Mireille Claustres[1], Michel Koenig[1], Mireille Cossée[1], Anne-Françoise Roux[1], and David Baux*[1]

[1]Laboratoire de génétique des maladies rares, EA7402, Université de Montpellier, Laboratoire de génétique moléculaire, CHU de Montpellier, Montpellier France (LGMR, CHU Montpellier) – CHU Montpellier, Université de Montpellier – France

[2]Laboratoire de biologie des tumeurs solides, CHU de Montpellier, Université de Montpellier, Montpellier France – CHU Montpellier, Université de Montpellier – France

Résumé

The identification of small DNA events (substitutions, small insertions/deletions) has been tremendously improved during the last 3 years but it is still challenging to effectively and easily identify larger events such as deletions/duplications of more than one exon. Sophisticated algorithms have been developed based on the depth of coverage of the regions of interest (ROIs), however, if useful, they are often complex to operate and rely on large pre-computed data sets. We have been using a method for several years in our laboratory based on normalized depth of coverage comparison between samples and regions of a single Illumina run. We present here MobiCNV (https://github.com/mobidic/MobiCNV), a simple and rapid python script implementing this algorithm that can be run with only basic bioinformatics knowledge on UNIX or Microsoft Windows platforms. The script computes as input csv or tsv files summarizing the depth of coverage for each ROI obtained either directly from Illumina regular pipelines Local Run Manager® or MiSeq Reporter® or from a slightly modified samtools bedcov command. The more samples the run contains (at least 4, ideally between 8 and 20), the more sensitive the method is. MobiCNV performs more efectively with runs having an average sequencing depth above 100X and is therefore well suited for gene panel analyses in clinical context or for exomes with sufficient coverage. This algorithm also relies on the uniformity of the sequencing experiment and consequently requires good quality data. The method is particularly effective for multi-exon events. The output is a single Excel spreadsheet which contains several worksheets. The most useful one is the summary sheet, which presents all the regions containing a predicted event surrounded by the neighbouring regions for comparison. Moreover, MobiCNV includes a gender prediction module, based on the analysis of reads mapping on the X chromosome which ultimately produces a table summarizing the gender predictions of each sample. It is worth noting that sexual chromosomes and autosomes are treated separately by the algorithm. MobiCNV can also focus on a list of genes of interest provided as a simple text file (e.g. for large panels including multiple pathologies). MobiCNV is a simple script which efficiently analyses CNVs from gene panel and exome data, and that can easily be implemented as part of any analysis pipeline.

*Intervenant

Mots-Clés: NGS, CNV, Copy Number Variants, Illumina sequencing