A suite of R packages for the analysis of DNA copy number microarray experiments

Philippe Hupé[1,2,*]

1. U900 Institut Curie, INSERM, Mines ParisTech
2. UMR144 Institut Curie, CNRS

* Contact author: philippe.hupe@curie.fr

Keywords: DNA copy number, microarray, biostatistics algorithms, oncology

Microarray technology is a powerful tool very helpful in oncology in order to better understand the molecular mechanisms involved in tumoral progression. A ccommon charecteristic of tumours is the presence of chromosome alterations and especially a change of their DNA copy number. There are microarrays which allow the quantification of DNA copy number. The raw data obtained from the microarray technology need appropriate statistical processing so that they can be biologically and clinically meaningful. Thus, we developed statistical methods in order to normalise and extract the biological information from microarrays devoted to the study of DNA copy number in tumours. Our methods are implemented within three R packages which are part of the Bioconductor project:

MANOR - Neuvial et al. (2006) This package implements a normalisation method devoted to the spatial normalisation of DNA copy number data. Briefly, the method consists of a spatial smoothing of the data followed by a segmentation which identifies aberrant spatial areas on the chip.

GLAD - Hupé et al. (2004) This package allows the detection of breakpoints in the DNA copy number molecular profiles (this step is called *segmentation*) and the assignment of a status (either loss, normal, gain or amplification) to each region identified (this step is called *calling*). The calling step provides valuable information for downstream analyses. The development of such an algorithm also avoid the tedious task of a manual expertise which is subject to error, non-reproducible and time-consuming (and even untractable for high-density chips).

ITALICS - Rigail et al. (2008) This package proposed a normalisation method devoted to the analysis of Affymetrix® Genome-Wide Human SNP Array. Besides normalisation, the proposed method has the originality to perform the identification of the DNA copy number alterations using the GLAD algorithm. The algorithm alternatively identifies the DNA copy number alterations and normalises the data. Those two alternative steps are iterated to improve the signal-to-noise ratio of the data at each iteration. The normalisation step takes into account the information of the genome alterations to better estimate the sources of variability to correct during the normalisation step.

The packages we have developped have already been widely used within the scientific community. Moreover, Institut Curie has developed client/server application named CAPweb which integrates the three previous packages (Liva et al., 2006). CAPweb is a user-friendly tool enabling biologists to analyse DNA copy number experiments from raw data to visualisation and biological interpretation. With CAPweb it is possible to manage the data, to normalise the DNA copy number experiments data with MANOR, to detect breakpoints with GLAD, to analyse Affymetrix data with ITALICS, to visualise and analyse the genomic profiles with VAMP (La Rosa et al., 2006).

References

Hupé et al. (2004). Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics.

La Rosa et al. (2006). VAMP : Visualization and Analysis of CGH array, transcriptome and other Molecular Profiles. Bioinformatics.

Liva et al. (2006). CAPweb: a bioinformatics CGH array Analysis Platform. Nucleic Acids Research.

Neuvial et al. (2006). Spatial normalization of array-CGH data. BMC Bioinformatics.

Rigaill et al. (2008). ITALICS: an algorithm for normalization and DNA copy number calling for Affymetrix SNP arrays. Bioinformatics.