Cancer research – R package to analyze genomic regulation and tumor pathways based on array data from single nucleotide polymorphism (SNP) and comparative genomic hybridization (CGH) experiments

E. KORSCHING, W. NADLER[1], H. BÜRGER
Institute of Pathology, University of Münster (korschi@uni-muenster.de)
[1]John von Neumann Institute for Computing, Research Centre Jülich

Aims: Cancer research is focused on a better understanding of tumor progression pathways. Progression means that there are successive development stages of a certain tumor type which often can be discriminated by means of molecular biology methods or classically by morphologic classification of an ultra thin tissue section. This knowledge on types and grades in turn helps to understand and identify early events which makes it easier to intercept the progression on a level where the therapeutic action is feasible. In this context it is important to decipher marker molecules for diagnostic and also therapeutic purpose. Marker means that some molecular factors play a more specific role in a certain tumor event than other involved molecules.
One of the approaches to get a better insight in tumor development is to analyze genomic events by means of the CGH / matrix CGH and more recent by SNP analysis.
The SNP analysis on the whole human genome with a very dense coverage gives the possibility to go beyond the limits of resolution of the classical CGH and also the matrix CGH. We are trying to explore the new details arising with this high resolution and established a set of algorithms supporting this analysis approach.

Implementation: The set of algorithms was originally developed in S because the platform SPlus has an excellent visual data browser very useful in big and complex projects and also in routine work. But a lot of problems concerning the closed source platform their evolvement away from science and also some technical limitations convinced us to switch to R with our activities.
The functionality of the R code cover the following tasks: a) show the relation between classical CGH and SNP experiments b) associate SNP and gene expression data to analyze regulatory motives and c) visualize results in an appropriate full genome or chromosome mode. The implementation is actually based on the CSV format of the exported Affymetrix data, but can be easily adapted to other data structures.
The introduction of the 10K SNP microarray by Affymetrix has brought an average resolution of 210 kb up to 6 kb with the 500K SNP chip set. The developed software tools allow on one hand a rapid, global overview of all unbalanced chromosomal alterations within a tumor by the use of a smoothing procedure. On the other hand our tools allow a much more detailed view into the fine structure of chromosomal alterations giving a much better insight into the complex picture of chromosomal alterations.

Conclusion: The developed analysis tools are a stable platform to promote new models in genomic deregulation occurring in cancer events.

Reference: *Improvements in the analysis strategy make single nucleotide polymorphism (SNP) analysis a powerful tool in the detection and characterization of amplified chromosomal regions in human tumors. Korsching E. et al., Pathobiology 2006, in press.*