Methods for integration of genome-wide miRNA and mRNA paired expression data sets

Stephan Gade, Stephan Artmann, Klaus Jung and Tim Beissbarth

German Cancer Research Center, Heidelberg &
Dept. Medical Statistics, University Medical Center Göttingen
Germany

Keywords: data fusion, miRNA expression, gene expression, classification

Motivation: Genome-wide expression measurements, e.g. via microarray platforms, are a widespread means to research the molecular features that are crucial in the development of diseases such as cancer and that could determine the fate of a patient, for example by being predictive for response to clinical treatment. Most common are measurements of mRNA expression that allow an approximation of a steady-state snapshopt of the gene-activity within a cell or a tissue sample, e.g. a tumor sample or needle biopsy. Recently micro-RNAs (miRNAs) have become a focus of research. miRNAs are small non-coding RNA molecules which can regulate mRNA-expression post-transcriptionally based on specific sequence complementarity. In recent studies it becomes more and more common that both types of measurements are performed on the same patient samples. Methods to combine these two data types utilizing their special biological features are still rare.

Methods: We have devised and tested several methods for the combination of miRNA and mRNA data. The targetet mRNAs to each miRNA are predicted based on their specific sequence and collected in several known databases. The information from the target databases is used in conjunction with the miRNA and mRNA expression profiles. When searching for differentially expressed miRNAs, we utilize this information by combining statistical tests on the miRNA data as on the mRNA targets. We present a method to combine the information available on miRNAs and their target sets' expression. miRNA-wise p-values from established component-wise testing and gene set p-values from different self-contained and competitive gene set tests are combined using methodology often used in meta-analysis. Besides testing for differentially expressed genes, another problem often faced in clinical research is the prediction of clinical outcome. Here we present a method to fuse mRNA and miRNA data in a classification algorithm utilizing the target predictions and the correlation structure.

Results: We have tested our approaches in simulation studies as well as on several example data-sets.

Availability: All methods are implemented in R-packages that can be obtained by the authors. The R-packages will be made publicly available in the near future.