Significance and Reliability of Hybridization Signals Derived from Expression Profiling on High-Density cDNA Arrays

Hilmar Lapp[1], Marion Weissmann[1], Judith Boer[2], Gudrun Werner[1]
[1]*Novartis Forschungsinstitut Vienna, Genetics Dept.,*
Hilmar.Lapp@pharma.novartis.com
[2]*German Cancer Research Center Heidelberg, Dept. of Molecular Genome Analysis,*
J.Boer@dkfz-heidelberg.de

Motivation: By hybridizing labeled complex probes derived from total Poly(A)$^+$-RNA of tissues, cell-lines, or cultured cells to high-density cDNA arrays the expression of a large number of clones can be profiled and analyzed in parallel (e.g., see [1], and references therein). We hybridize radioactive labeled probes to arrays gridded onto nylon membranes, such that each membrane carries 16,100 up to 27,600 clones in duplicate. Hybridization intensities obtained from the resulting spot array images frequently proved to be not directly comparable between different hybridizations and even not within the same image due to various sources of bias. The detection limit below of which measured intensities rather represent noise than true expression levels is unknown. The hybridization signal of a particular clone with a particular probe is of only limited reproducibility. However, quantitative analysis to investigate expression patterns and differences requires comparable hybridization signals with a known significance of being different from background, as well as the significance of observed differential expression.

Results: We developed methods and algorithms to address these issues mentioned above, i.e., to remove bias, to derive a significance of being different from background, and to assess the reproducibility of hybridization signals. In particular, we developed two methods that successfully remove the prevailing bias, which is caused by inhomogeneous background brightness, by employing local (neighborhood-based) normalization without distorting the intensity distributions. The methods compute average local background spots and average local spots. Regarding the detection limit, we devised an algorithm that can resolve the distribution of signals observed for a particular hybridization into the sum of two populations representing background spots and positive spots. The algorithm assigns a probability to each spot for belonging to the population of positives, which can also be used to score hybridization signals independently of particular image properties. In order to assess the reproducibility of hybridization signals, we developed a program that fits theoretically expected to observed distributions of ratios of duplicates and repetitions, and, based on this, derives confidence intervals, which can be used to determine the significance of observed fold change.

In summary, we established a set of tools for the analysis of high-density cDNA arrays on nylon membranes. The algorithms yield quantitative, accurate, and comparable expression data together with significance measures, and run without user-interaction. All data analysis functions were written in R. Code is available on request by e-mail to Hilmar.Lapp@pharma.novartis.com.

References:

1. Pietu G, et. al. (1999). The genexpress IMAGE knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. Genome Res 9:195-209