

Expression Profiling on cDNA Arrays: A Robust Method For Resolving Hybridisation Intensities Into Background and Positives

Hilmar Lapp, Marion Weissmann, Gudrun Werner

Novartis Research Institute Vienna, Computational Biology and Chemistry

Hilmar.Lapp@pharma.novartis.com

Motivation: As the first step in analysing hybridisations of labelled cDNA samples to high-density arrays carrying cDNA clone libraries [1], image analysis assigns hybridisation intensities to every clone on the array. In any downstream analysis, such as comparisons of multiple samples, differential expression screening, comparison of clones by their expression patterns etc., the intensity values are assumed to quantitatively reflect the abundance level of a transcript in the sample. Since a high proportion of genes is expressed only at a low level in a given tissue at any point in time [2], the majority of clones will be assigned only low intensities in the case of genome-wide target libraries. However low intensity does not necessarily correspond to either low abundance or absence of the respective transcript, due to various sources of experimental noise and the inherent detection limit of the system. Obviously, one would like to exclude signals representing noise from further downstream analysis. To this end, we devised a method that robustly resolves the population of observed intensities into two populations, namely background and positives, and provides a significance score for each intensity being different from background.

Results: The method we developed is based on the model that the observed distribution of intensities is a mixture of two populations, which are referred to as background (noise, unspecific binding, and intensities below the detection limit) and ‘true’ positives. The model is fitted along the lines of an EM algorithm by iterating over alternating steps of distribution parameter estimation and classification. Robustness is achieved by employing iterated robust parameter estimation. The resulting estimates of the two populations are superior to minimising an error function in terms of robustness, and can be used to compute a significance score for a spot being different from background. This score is independent of individual properties of an image, like average background. We demonstrate the method, its performance and applicability to different kinds of cDNA arrays (arrays on membranes, microarrays) and hybridisations (complex samples, DNA mass detection probes). In addition, we compare the results to other methods for estimating the background fraction.

References:

- [1] Piétu G., et al., (1996), *Genome Res.* 6:492-503; DeRisi, et al. (1996), *Nature Genet.* 14:457-460; Lennon G.G. and Lehrach H. (1991), *Trends Genet.* 7:314-317
- [2] Zhang L., et al. (1997), *Science* 276:1268-1272