Poster L-1

A calibration method for estimating absolute expression levels from microarray data

Authors:
Kristof Engelen (*Department of Microbial and Molecular Systems, KULeuven*)
Hui Zhao (*Department of Electrical Engineering, KULeuven*)
Bart De Moor (*Department of Electrical Engineering, KULeuven*)
Kathleen Marchal (*Department of Microbial and Molecular Systems, KULeuven*)

Short Abstract: We describe an approach for estimating absolute expression levels from two-color, spotted microarray data. The procedure is based on a physically motivated calibration model of which the parameters and error distributions are estimated from external control spikes.

Long Abstract:

A calibration method for estimating absolute expression levels from microarray data

Kristof Engelen, Hui Zhao, Bart De Moor and Kathleen Marchal

Normalization of microarray measurements is the first step in a microarray analysis flow. It aims at removing consistent sources of variations to make measurements mutually comparable. Reliable normalization is essential since the results of all subsequent analyses (such as e.g. clustering) might largely be influenced by the normalization procedure. For normalization of cDNA different methods have been described. Although some approaches inherently work with absolute intensities (e.g. ANOVA[1,2]), in general, preprocessing, of cDNA microarrays largely depends on the calculation of the log-ratios of the measured intensities. A common normalization step consists of the linearization of the Cy3 vs. Cy5 intensities (e.g. loess[3]). It assumes the distribution of gene expression is balanced and shows little change between the biological samples tested (Global Normalization Assumption). Global mRNA changes that result in an uneven distribution of expression changes however, have been shown to occur more frequently than what is currently believed[4,5], and could have a significant impact on the interpretation of data normalized according to the Global Normalization Assumption.

Therefore, in this study we describe a different way of normalizing cDNA microarray data. In contrast to previous approaches, our methodology is based on a physically motivated model, consisting of two major components. We explicitly model the hybridization of mRNA transcripts to their corresponding cDNA probes and the relation between the measured fluorescence and the amount of hybridized, labeled mRNA. The parameters of this model and the incorporated error distributions are estimated from external control spikes: mRNA transcripts that are added to the hybridization solution in known concentrations. Using a publicly available data set, we show that our procedure, due to the inherent nonlinearity of the model, is capable of adequately linearizing the data, without making any assumptions on the distribution of gene expression (as opposed to the Global Normalization Assumption). More importantly, since our model links mRNA concentration to measured intensity, we are

able to estimate the absolute concentrations of mRNA transcripts in the hybridization solution with fair accuracy.

1. Kerr MK, Martin M, en Churchill GA 2000. Analysis of variance for gene expression microarray data. J. Comput. Biol. 7: 819-837.
2. Jin W, Riley RM, Wolfinger RD, White KP, Passador Gurgel G en Gibson G 2001. The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. Nat Genet 29: 389-395.
3. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J en Speed TP. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30: e15.
4. van de Peppel J, Kemmeren P, van Bakel H, Radonjic M, van Leenen D, Holstege FC. 2003 Monitoring global messenger RNA changes in externally controlled microarray experiments. EMBO Rep. 4(4):387-393.
5. van Bakel H, Holstege FC. 2004 In control: systematic assessment of microarray performance. EMBO Rep. 5(10):964-969.