Modelling of Distortion of Microarray Time-Series Experiments Caused by Loss of Cell Synchrony

Tommi Aho, Harri Lähdesmäki, Heikki Huttunen,
Marja-Leena Linne, Jari Niemi, Ronald Pearson and Olli Yli-Harja

In this paper we consider the problem of degradation of cell synchrony during microarray time-series experiments. We explore the phenomenon using real data and show its significance to the quality of microarray data. We formalize it within the framework of signal processing.

The homogeneity of the sample cell population is important in many microarray experiments. This is because the measurement result is an average over the population which, in the case of microarray experiments, consists of millions of cells. Various synchronization methods to this problem have been proposed c.f. [Futcher99], which e.g. stop all cells to a certain cell cycle phase. However, when the cells are released again, they are moving along the cell cycle at different rates and hence the synchrony is lost during time. Gradually the synchrony vanishes and the phase distribution of the cell population widens. This results in the observed averaging of the measured gene expression values.

We have explored the yeast time-series data provided by Spellman et al. [Spellman98]. The averaging of the data can be observed by studying the cell cycle -dependent genes that are assumed to have similar expression patterns within each cell cycle. We explore the behaviour of those genes by visualising their gene activity profiles using different multidimensional scaling methods. The results resemble spiral shaped trajectories, which indicates that the synchronization of the sample cell population degrades over time.

In order to model the loss of synchrony, we estimate the spread of the phase distribution of the cell population. The spreading is modelled using convolution, which describes the observed transformation between consecutive cell cycles. With simulated data this approach works successfully, for example the expression profiles taken from the second cell cycle can be predicted based on the first cell cycle. With the real data [Spellman98] the prediction is more difficult, but certain stationary impulse responses can be extracted out of expression patterns of cell cycle dependent genes in the first and second cell cycle. The small number of samples per cell cycle and the noisy data make the estimation of the impulse response harder. The performance of the method will improve when sampling will be more frequent and measurements more reliable. However, the problem of cell synchronization probably remains due to biological reasons.

The synchronization problem can be formalized with the help of signal processing concepts and methods. We expect this approach to be useful in improving the quality of microarray time-series data.

References:

[Futcher99] B. Futcher. Cell Cycle Synchronization. Methods in Cell Science 1999, Vol. 21, pp. 79-86.

[Spellman98] P. Spellman, G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein, B. Futcher. Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Molecular Biology of the Cell 1998, Vol. 9, pp. 3273-3297.