

FARMS: a probabilistic latent variable model for summarizing Affymetrix array data at probe level

Djork-Arné Clevert¹ and Sepp Hochreiter¹

¹ *Institute of Bioinformatics, Johannes Kepler Universität Linz 4040 Linz, Austria*

Motivation:

High-density oligonucleotide microarrays, and in particular Affymetrix GeneChip arrays, are successfully applied in many areas of biomedical research. However, the large number of genes, which have high variation but are irrelevant for the experimental question, leads to many false positives in extracting relevant genes. The high variation which originates from measurement noise can be reduced by identifying the probes of a probe set which vary synchronously across the arrays. The cause for synchronous changes is the gene-specific mRNA concentration in the cell detected by those probes.

This concentration is the hidden factor in our factor analysis model, which automatically models the probe-specific measurement error and is optimized by Bayesian maximum a posteriori estimation. In contrast to previous methods our new summarization method called "Factor Analysis for Robust Microarray Summarization" (FARMS) supplies model-based signal intensity values. We have evaluated FARMS in terms of sensitivity and specificity on all public available spike-in data sets and at the Affycomp competition where it outperformed its competitors.

Results:

We think that this paper is particularly suitable for the presentation in the highlight track, because FARMS automatically reduces microarray data sets by 95% by filtering out signal containing, informative genes. Common gene selection methods are only applicable to prediction or classification tasks and do not work on probe level. Therefore, FARMS can be very important for biologists and medical researchers, which have to face the problem of false discovery rates. This problem is considerably reduced because genes are filtered out without looking at conditions; therefore higher significance values are obtained in the post-processing.