A New Probabilistic Framework for cDNA Microarray Data Analysis

Babatunde A. Ogunnaike*, Claudio A. Gelmi and Jeremy S. Edwards
University of Delaware, Department of Chemical Engineering

1. Data representation and transformation

- Let $I_t$ (treated cell) and $I_r$ (reference cell) be the fluorescence signal intensities measured from each spot "i" on a microarray.
- Transform the data from a "cartesian" representation to a "polar" form, i.e.: $I = (R, \theta)$

\[
R = \sqrt{I_t^2 + I_r^2} \quad \text{(intensity magnitude)}
\]

\[
\theta = \frac{\tan^{-1}(I_t/I_r)}{\text{(fractional intensity)}}
\]

2. Drawing inferences

(i) Bias detection and correction

Theoretically, if there is NO BIAS the mean of $I_t(x)$ should be $\mu_t = 0.5$; remove via correction for $X_i$:

$$X_i = 2I_i/\left(1 + \left|I_i\right|/\mu_t\right)$$

Bias is indicated if $|X_i - 0.5| > 0.25$.

(ii) Probability of expression status

Using the probability distribution functions it is possible to calculate the probability of expression status (down, up or not regulated) for every gene.

$$P(g_i \in D) = \phi_{fl}(x) + \phi_{h0}(x) + \phi_{h+}(x)$$

3. Application results

- Table 1 shows the identification of relevant genes using classical fold-change criteria. These results are compared before and after bias correction using our probabilistic framework.
- Table 2 shows the number of candidates genes identified per group (low, medium and high R). All of them with a very high probability of expression status.

4. Future work

- Extend to GeneChip® technology.
- Compare different microarray technologies; develop appropriate metrics for assessing the "quality".
- Incorporate degree of confidence in the inference regarding the probability of expression status.

(*) : ogunnaike@che.udel.edu