

genArise

genArise is an easy to use tool for dual color microarray data implemented in the R language. genArise is a package that contains specific functions to perform an analysis of cDNA microarrays data to detect genes that are significantly differentially expressed under different growth conditions.

Before this analysis, genArise carry out a number of transformations on the data to eliminate low-quality measurements and to adjust the measured intensities to facilitate comparisons.

Background correction

Given the data set, a background correction can be performed. This is optional, but recommended. The default background correction action is to subtract the Cy3 background intensity from the Cy3 foreground intensity and the Cy5 background intensity from the Cy5 foreground intensity for each spot in the microarray.

R and I values

After this, the log-differential expression ratio and a measure of the overall brightness for each spot are computed. The log-ratio is represented by the $R = \log_2(\text{Cy5}) - \log_2(\text{Cy3})$, and the overall intensity by the $I = \log_{10}(\text{Cy5}) + \log_{10}(\text{Cy3})$. genArise use base-2 logarithms for R so the ratio is units of 2-fold change.

Lowess normalization

genArise can carry out the normalization within the array. The imbalance in the Cy5 and Cy3 intensities can vary widely across the spots within the array, and this variation can be due to overall spot intensity, location on the array, slide origin, and possibly other variables.

Due to spatial and intensity dependent biases in numerous experiments, genArise use a local weighted regression (loess) analysis. This normalization method can remove such intensity-dependent effects in the $\log_2(\text{ratio})$ values.

Lowess, can be applied either globally (over all the microarray data set) or locally (over each subgrid). We recommend to apply the local normalization, because it has the advantage that it can help correct for systematic spatial variation in the array, for example, inconsistencies among the print tips.

Filter by intensity

With the purpose of eliminate those points where the intensity is too close to their corresponding background. Also, each Cy3 value in the spot must be greater than two standard deviations from Cy3 background values, the same criterion for Cy5. In other way, the element is eliminated.

Replicates analysis

Replicates analysis is a replica averaging function. This function merges the replicated spots by presenting the Cy3 intensities and Cy5 intensities as a single average value of the replica.

Three methods have been implemented in genArise to obtain this average value. The first of them is the arithmetic mean of both values, the second is the geometric mean and finally the extreme values (if the log-ratio of the element is positive and the replicated is positive too, genArise keep the maximum of both of them, if the log-ratio of the element is negative and the replicated is negative, keep the minimum, if both of them have different sign, the element is eliminated).

Selecting Differentially Expressed Genes

The goal of genArise is to identify which of the genes show good evidence of being differentially expressed. This function identifies differentially expressed genes by calculating an intensity-dependent Z-score. We use a sliding window algorithm to calculate the mean and standard deviation within a window surrounding each data point, and define a Z-score where Z measures the number of standard deviations a data point is from the mean.

$$z_i = (R_i - \text{mean}(R)) / \text{sd}(R)$$

where z_i is the z-score for each element, R_i is the log-ratio for each element, and $\text{sd}(R)$ is the standard deviation of the log-ratio.

With this criterion, the elements with a $|z\text{-score}| > 2$ standard deviations would be the significantly differentially expressed genes.