



MicroRNA Profiling at Ocean Ridge Biosciences miRBase Version 16 Multi-Species Microarrays

MicroRNA Microarrays

RNA samples were sent to Ocean Ridge Biosciences (ORB, Palm Beach Gardens, FL) for analysis using custom multi-species microarrays containing XXXX probes covering XXXX human/mouse/rat/rhesus mature microRNAs present in miRBase version 16. The sensitivity of the microarray is such that it could detect as low as 20 amoles of synthetic microRNA being hybridized along with each sample. The microarrays were produced by Microarrays Inc. (Huntsville, Alabama), and consisted of epoxide glass substrates that had been spotted in triplicate with each probe.

Sample Processing

Quality control of the total RNA samples was assessed using UV spectrophotometry and agarose gel electrophoresis. The samples were DNase digested and low-molecular weight (LMW) RNA was isolated by ultrafiltration through YM-100 columns (Millipore) and subsequent purification using the RNeasy MinElute Clean-Up Kit (Qiagen). The LMW RNA samples were 3'-end labeled with Oyster-550 fluorescent dye using the Flash Taq RNA labeling Kit (Genisphere). Labeled LMW RNA samples were hybridized to the MicroRNA microarrays according to conditions recommended in the Flash Taq RNA labeling Kit manual. The microarrays were scanned on an Axon Genepix 4000B scanner, and data was extracted from images using GenePix V4.1 software.

Data Pre-processing

Spot intensities were obtained for the XXXX features on each microarray by subtracting the median local background from the median local foreground for each spot. Detection Thresholds for each array were determined by calculating the 10% trim mean intensity of the negative controls spots and adding 5X the standard deviation of the background (non-spot area). The spot intensities and the Threshold were transformed by taking the log (base 2) of each value. The normalization factor (N) for each microarray was determined by obtaining the 20% trim mean of the spot intensities for the human/mouse/rat/rhesus probes above threshold in all samples. The log₂-transformed spot intensities for all XXXX features were normalized, by subtracting N from each spot intensity, and scaled by adding the grand mean of N across all microarrays. The mean probe intensities for each of the XXXX human/mouse/rat/rhesus probes on each of the X arrays were then determined by averaging the triplicate spot intensities. Spots flagged as poor quality during data extraction were omitted prior to averaging. The XXX human/mouse/rat/rhesus non-control log₂-transformed, normalized, and averaged probe intensities were filtered to obtain a list of XXXX human/mouse/rat/rhesus microRNA probes showing probe intensity above T in all samples from at least one treatment group.

Microarray Quality Control

Each array contains probes targeting 11 different synthetic miRNAs, each of which is added at a mass of 200 amoles to each RNA sample prior to labeling and hybridization. Sensitivity of the microarray hybridization was confirmed by detection of hybridization signal for all 11 spikes well above the detection threshold. The array also contains a set of specificity control probes complementary to three different miRNAs. Each specificity control includes a perfect match,

single mismatch, double mismatch, and shuffled version of the probe. Specificity of the hybridization was confirmed by detection of hybridization signal on the microarray for the perfect match probes and not the double mismatch and shuffled version of the probes. Reproducibility of the arrays was determined by monitoring the hybridization intensity for the triplicate human/mouse/rat/rhesus spots on each array. The sensitivity, specificity, and reproducibility data for the arrays were compiled into a Quality Control report.

Differential Expression Analysis

For statistical analysis, samples were binned in X groups (X,X,X,etc...). The log2-transformed and normalized spot intensities for the XXX detectable probes were examined for differences between the groups by 1-way ANOVA using National Institute of Ageing (NIA) Array Analysis software (1). This ANOVA was conducted using the Bayesian Error Model and 10 degrees of freedom. A total of 15 probes showed significant differences with $P < 0.01$ and $FDR < 0.1$. The statistical significance was determined using the False Discovery Rate (FDR) method which was proposed by Benjamini and Hochberg (2). It is the proportion of false positives among all probes with P values lower or equal to the P value of the probes that we consider significant. It can also be viewed as an equivalent of a P-value in experiments with multiple hypotheses testing. FDR is an intermediate method between the P-value and Bonferroni correction (multiplying P-value by the total number of probes). The equation is:

$$FDR_r = \min_{i \geq r} \left[\frac{p_i N}{i} \right]$$

where r is the rank of a probe ordered by increasing P values, p_i is the P value for probe with rank i, and N is the total number of probes tested. FDR value increases monotonously with increasing P value.

Principal Component Analysis was also performed on the XXX detectable probes using the module built in to the NIA software.

Hierarchical Clustering Of MicroRNA Array Data

Data for the XXX detectable microRNA probes were clustered using Cluster 3.0 software (3). Three rounds of gene median centering and gene median normalization were used to preprocess the data. Hierarchical clustering was conducted using Centered Correlation as the similarity metric and Average Linkage as the clustering method. Intensity scale shown is arbitrary.

References

- (1) Sharov, A.A., Dudekula, D.B., Ko, M.S.H. (2005) Principal component and significance analysis of microarrays with NIA Array Analysis tool. *Bioinformatics*. 21(10): 2548-9.
- (2) Benjamini, Y. & Hochberg, Y., (1995). *J Roy Stat Soc B* 57: 289-300
- (3) De Hoon, M. J. L., Imoto, S., Nolan, J. Nolan, and Miyano, S. (2004) Open Source Clustering Software. *Bioinformatics*, 20 (9): 1453-1454.