

A Novel Method of Studying the Disease Regulatory Activities of MicroRNAs (Supplementary details)

Sanghamitra Bandyopadhyay and Malay Bhattacharyya
Machine Intelligence Unit, Indian Statistical Institute,
203 B. T. Road, Kolkata - 700108, India
E-mail: {sanghami, malay_r}@isical.ac.in

July 23, 2009

Introduction

Microarray analysis of microRNAs (miRNAs) is a high throughput method for studying miRNA expression in cultured cells or tissues. Some of the topical studies strongly suggest that the disorders in the normal activities of miRNAs might cause many diseases. Generally, such studies concern patient-specific expression profiles for the purposes like pruning, clustering or classification. This paper describes a novel relative co-expression measure to take in account the deviation in microarray expression profiles between diseased and general people. The dataset used in this analysis is the resultant of microarray experimentation using human (*Homo Sapience*) miRNAs. We describe the materials and methods employed in the study in the sections ahead.

1 Materials

A single microarray dataset has been analyzed in this computational study. This dataset consists of expression values of schizophrenia patient-specific miRNAs [Perkins *et al.*, 2007]. This dataset contains the expression values of postmortem human brain tissues, obtained from the Harvard Brain Tissue Resource Center, consisting of frozen blocks (300-500 mg/block) from the postmortem prefrontal cortex (Brodmann area nine from 15 individuals with schizophrenia and 21 unaffected comparison subjects). The subjects considered in this study were free of neurodegenerative pathology and the tissues were groupmatched for gender, age, postmortem interval, and hemisphere.

The dataset was prepared by synthesizing oligonucleotide probes in duplicate for 264 miRNAs antisense to the mature sequence in a microarray analysis. The mature sequences were verified from the reported results accumulated in the Sanger miRNA registry. All arrays were prepared from the same batch using the conventional hybridization process. The microarray analysis started with the extraction of data from the GPR files and the data points having foreground values lower than 1.5 times of the local background were eliminated. Following this elimination, the probes from which greater than 40% of the data points were removed, were discarded out of the analysis. This pre-processing reduced the dataset comprising a total of 239 miRNAs. After the background subtraction all the data were log-transformed and the missing values were estimated using the well-known k-NN [Troyanskaya *et al.*, 2001]. The data were normalized using rank invariant normalization for comparison across the samples. The per-sample mean of the two rank invariant

normalized probes was used for the analysis. Univariate calculations of differential expression were estimated using two-class (unpaired test) Statistical Analysis of Microarrays through 500 permutations with an FDR of 5%.

Thus, the final dataset includes the expression values received over 36 experiments for 239 miRNAs. The statistical information pertaining the microarray data is - minimum expression value = 6.03, maximum expression value = 15.88, average expression value = 7.27, standard deviation of the expression values over the entire dataset = 1.38.

2 Methods

The methods employed in the study are described in detail in the following subsections.

2.1 Computation of p -value

Suppose, an event is found to occur n times out of a total N observations and given the evidence that the event originally occurs e times out of E total cases. Then, the p -value of the observation is computed, assuming a hypergeometric distribution, as

$$p\text{-value} = \sum_{i=n}^N \frac{\binom{e}{i} \binom{E-e}{N-i}}{\binom{E}{N}}. \quad (1)$$

3 Color Figures

The color version of Fig. 2 appearing in the main paper is shown in Fig. S1.

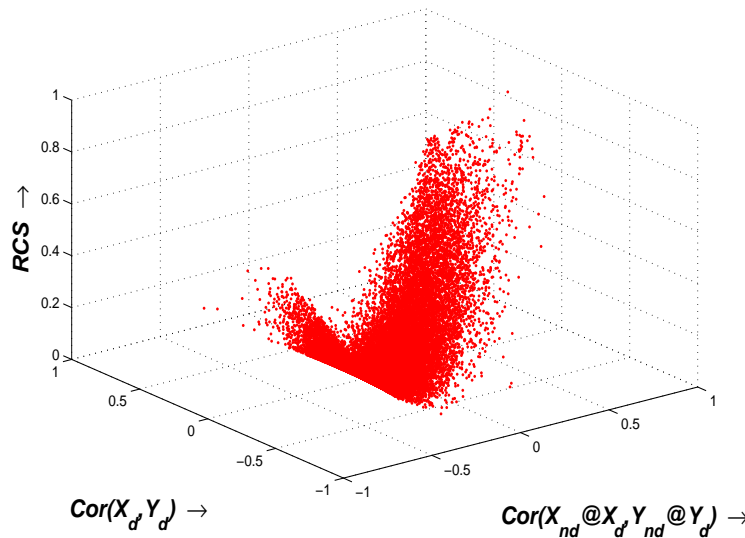


Figure S1. Original RCS values computed from the schizophrenia dataset.

The color version of Fig. 3 appearing in the main paper is shown in Fig. S2.

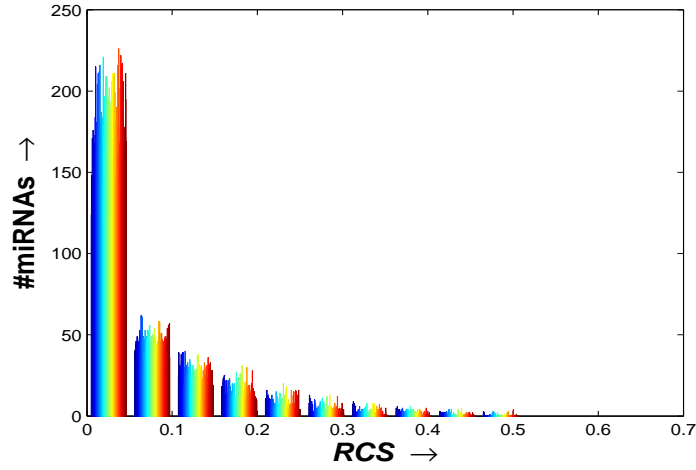


Figure S2. Histogram of #miRNAs over RCS values.

The color version of Fig. 4 appearing in the main paper is shown in Fig. S3.

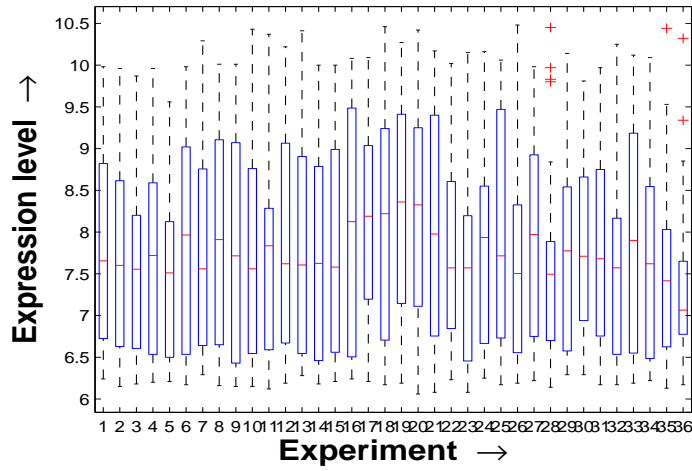


Figure S3. QDP of the largest DAN identified from the Schizophrenia dataset.

The color version of Fig. 5 appearing in the main paper is shown in Fig. S4.

References

[Perkins *et al.*, 2007] Perkins, D.O. *et al.* (2007) microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder, *Genome Biol.*, **8**, R27.

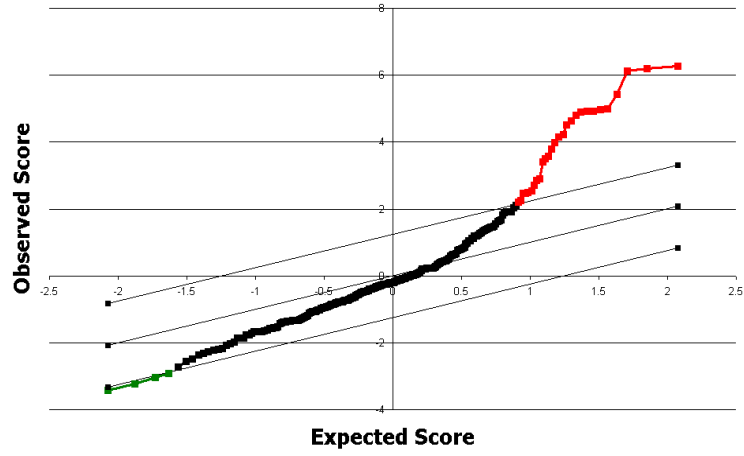


Figure S4. The SAM plotsheet obtained by the significance analysis of the schizophrenia microarray dataset.

[Troyanskaya *et al.*, 2001] Troyanskaya, O. *et al.* (2001) Missing value estimation methods for DNA microarrays, *Bioinformatics*, **17**, 520-525.