

A Novel Method of Studying the Disease Regulatory Activities of MicroRNAs

Sanghamitra Bandyopadhyay and Malay Bhattacharyya*

Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata – 700108, India

Abstract: MicroRNAs (miRNAs) are small, non-coding RNAs that participate in the post-transcriptional regulation of messenger RNAs (mRNAs) by degrading or inhibiting translation. 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 compute deviation in microarray expression profiles of diseased people over a set of people. This measure is used by an unsupervised algorithm of complexity $O(n^3 \log n)$, where n denotes the number of miRNAs, to locate the group of miRNAs responsible for the specific disease. The results taken over the expression data of schizophrenic patients show efficiency in locating brain-enriched miRNAs, which have earlier established support to be associated with schizophrenia neuropathology.

Keywords: Schizophrenia, miRNA, microarray, p -value.

INTRODUCTION

Present researches in genomics considerably focus on a class of small (~22nt), non-coding RNAs, known as microRNAs (miRNAs), which regulate mRNA expression by degradation or inhibition of translation at the post-transcriptional level [1]. Study of miRNAs in the current decade has been enriched in several directions like exploring the functionality of miRNAs, identification of miRNA encoding genes, tissue-specificity analysis [2], reconstruction of regulatory networks [3], disease regulation [4], etc. Among these, the study of miRNA regulated diseases has received major attention due to its pharmaceutical significance. Certainly, this is a demanding study in state-of-the-art research in computational molecular biology. MiRNAs are responsible, directly or indirectly, for a large number of diseases as they can dysregulate post-transcriptional gene expression [5]. Earlier, major of these studies have been conducted with biological experiments [6-8]. However, the motivation of the current work is to identify the disease regulatory miRNAs by computational analysis. We present and study an unsupervised algorithm to locate the group of miRNAs responsible for the schizophrenia disease.

Microarray profiling is a high-throughput technology that enables to measure the expression profiles of multiple genes in parallel. Microarray analysis is a commonplace study employed for genes studies [9-11]. Similar high-throughput analysis is currently also performed for miRNAs [12]. Biologically, microarray profiling of miRNAs is done for therapeutic, diagnostic, and prognostic applications. Since it is highly anticipated that an miRNA might regulate a disease due to disorders in their normal function, a large deviation of

expression of miRNAs is naturally expected between normal people and affected people. With this insight, this paper proposes a relative score to measure the variation of co-expression between miRNA pairs. This probabilistic score computes the correlation between two miRNAs X and Y over the universe of discourse (as observed in all the subjects including the affected patients and normal people), with respect to the background correlation between X and Y in patients. This helps to identify the highly co-expressed miRNA pairs in general people, which become poorly co-expressed while considering the patients only, probably because of their abnormal functions causing those diseases. These miRNAs are expressed abnormally in diseased cells.

The framework of the method proposed in this paper is based on fuzzy complete graphs (FCGs) [13]. The above mentioned relative co-expression score, assigned to each pair of miRNAs, can be interpreted as the fuzzy relations in a fuzzy co-expression network. These co-expression networks are apparently similar to FCGs. A subset of vertices of an FCG is generally denoted as an N-vertexlet. We have used a novel measure to find out association of miRNAs within an N-vertexlet. By identifying such dense association within the N-vertexlets, noted as DANs, a statistically significant module of miRNAs has been recognized. An $O(n^3 \log n)$ algorithm is proposed in the paper that could efficiently locate the largest DAN from an FCG, given the density of it. The algorithm employs a heuristic that makes the mining method efficient for scale-free graphs.

In this work, we exclusively study the regulatory activities of miRNAs over schizophrenia, for which there are evidences of miRNAs regulating brain development, dendritic spine morphology, and neurite outgrowth that relates to schizophrenia neuropathology [4]. Promising outcome of the methodology is established by comparing the results with the ground truths existing in the literature. It also gives rise to some novel findings.

*Address correspondence to these authors at the Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata - 700108, India; E-mail: sanghami, malay_r@isical.ac.in

RELATED WORKS

During the past decade a large number of complex diseases have been suggested to be regulated by miRNAs [5]. Schizophrenia is one of the most studied diseases in miRNA research. It has been extensively studied that miRNAs play an important role in regulating brain development early in life. It also mediates synaptic plasticity in a matured organism [4]. Earlier, it was hypothesized that schizophrenia might be associated with altered expression or function of miRNAs [14]. Later, this hypothesis has been well established in a study by investigating the expression of human miRNAs in the prefrontal cortex [4]. They have computed the fold change in the expression of human miRNAs in affected people taking the background of unaffected people. From then on a large number of works has been done in this specific direction of research. Earlier researches have studied the activities of single miRNAs like hsa-miR-130b [15], hsa-miR-181b [6] to find their responsibility in the regulation of schizophrenia. In [6], dysregulation of miRNA 181b has been observed in the temporal cortex in schizophrenic tissues. Again, some of the works have focused on the expression of multiple miRNAs in parallel for studying their specificity, fold-change or their dysregulation [4, 7] to be a cause of schizophrenic disorder.

These earlier studies strongly suggest the involvement of miRNAs in regulating schizophrenia. But, most of these analyzes are based on biological studies. Unfortunately, extensive studies to identify significant disease regulatory modules of miRNAs from a co-expression network by unsupervised learning on expression profiles have still not been carried out. The literature lacks from the inclusion of unsupervised machine learning methods in these directions.

FOUNDATION OF THE PROBLEM

An $m \times n$ microarray data is considered throughout the analysis as a combination of m (corresponding to miRNAs) n -dimensional (corresponding to expression values profiled from different subjects) expression row vectors. The subjects considered in this study are human (*Homo sapiens*) and are either diseased or non-diseased. Thus an expression vector X can be divided into two portions X_{nd} (non-diseased) and X_d (diseased) such that $X_{nd} \cap X_d = \emptyset$. Suppose, an append operator ($@$) is defined between two real expression row vectors X_1 and X_2 resulting in $X_1 @ X_2$ such that,

$$X_1 @ X_2 = [X_1 X_2]. \tag{1}$$

Notably, the append operation (by applying $@$) retains the following properties,

1. $X_1 @ X_2 \neq X_2 @ X_1$, unless $X_1 = X_2$,
2. $|X_1 @ X_2| = |X_1| + |X_2|$.

With this, the relative co-expression score (RCS) is now defined between two expression vectors X and Y as follows,

$$RCS(X, Y) = \frac{Cor^2(X_{nd} @ X_d, Y_{nd} @ Y_d)}{\varepsilon + Cor^2(X_d @ Y_d)}, \text{ for } X \neq Y$$

$$= 0, \text{ otherwise.} \tag{2}$$

Here, the constant term $\varepsilon > 0$ is incorporated to map the range of RCS from $[0, \infty]$ to $[0, \frac{1}{\varepsilon}]$ and $Cor(X, Y)$ denotes the Pearson correlation coefficient between X and Y . Pearson correlation coefficient between two expression vectors X and Y is computed as follows.

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$$= \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^N (x_i - \bar{x})^2][\sum_{i=1}^N (y_i - \bar{y})^2]}} \tag{3}$$

The complexity of computing the RCS score for each miRNA pair is $O(n^2)$, n being the size of the expression vectors. We expect higher values of RCS to be probabilistically an indication of high co-expressibility of the miRNA pair in general people (including the patients and non-patients), but low in patients. We include patients in general people to avoid any bias arising from the effect of any other disease. Moreover, in other view, we are incorporating noise in the correlation measure by including patients within the complete set of subjects. Significantly, in spite of this the results are promising. It reflects the co-expression (otherwise higher in normal case) loss in patients that might be due to functional disorder that is the probable origin of the disease.

We prefer to choose $\varepsilon = 1$ for this study, targeting the range $RCS \in [0, 1]$. The purpose is to use the RCS scores to form a co-expression network or equivalently an FCG [13]. Some theoretical details of this framework are described hereunder.

Definition 3.1 (Fuzzy Complete Graph) A fuzzy complete graph (FCG), $\tilde{G} = (V, \tilde{E}, \Omega)$, is defined as a graph in which V denotes the set of vertices, \tilde{E} denotes the set of fuzzy relations (v_i, v_j) ($v_i \neq v_j, \forall v_i, v_j \in V$) and Ω is a fuzzy membership function defined over the set \tilde{E} such that $\Omega: \tilde{E} \rightarrow (0, 1]$.

Definition 3.2 (Association Density of a vertex) Given an FCG, $\tilde{G} = (V, \tilde{E}, \Omega)$, the association density, μ_{v_i/V_{let}^N} , of a vertex v_i of \tilde{G} is defined, with respect to a set of vertices V_{let}^N ($v_i \notin V_{let}^N$), as the ratio of the sum of the fuzzy edge memberships between v_i and each of the vertices belonging to V_{let}^N , and N . Thus, the association density of a vertex v_i with respect to V_{let}^N is computed as,

$$\mu_{v_i/V_{let}^N} = \frac{\sum_{v_j \in V_{let}^N} \Omega_{v_i v_j}}{|V_{let}^N|}. \tag{4}$$

In Eqn. (4), $\Omega_{v_i v_j}$ denotes the fuzzy membership value of the vertex pair (v_i, v_j) . This density definition computes the

degree of participation of a single vertex within a set of vertices. Putting the constraint of a lower bound to this density factor for every vertex within an N-vertexlet, a DAN is defined as follows.

Definition 3.3 (DAN) Given an FCG $\tilde{G} = (V, \tilde{E}, \Omega)$, an N-vertexlet $V_{let}^N \subseteq V$, is defined to be a DAN, with respect to an Association Density threshold δ , that satisfies,

$$\min_{\forall v_i \in V_{let}^N} (\mu_{v_i/V_{let}^N - \{v_i\}}) \geq \delta \quad (5)$$

Now, we describe the proposed methodology in the subsequent section.

METHODOLOGY

Our target is first to construct an FCG computing the RCS value between each miRNA pair, and then, to locate the largest DAN within the FCG. The RCS values originally map to the fuzzy relations (similarity values between the vertex pairs) in the FCG. The proposed algorithm uses a multilevel linked list, the *NList*, which stores the ordered list of neighbors for each vertex. In Algorithm 1, we present the formal steps to construct an *NList* from an FCG. In the beginning of the algorithm (steps 1-3), a list of vertices, in the descending order of fuzzy edge weights from the vertex v_i , is prepared for each v_i in the FCG. Then (steps 4-11), the vertices are taken one by one from the beginning at each level of *NList* until they satisfy the Association Density threshold δ . Finally, the residual portion is removed out of the *NList* (step 7). Here, $NList(v_m, n)$ denotes the n^{th} dense neighbor of vertex v_m , i.e., the n^{th} member in this list of vertices in the order of descending RCS score.

Algorithm 1 A method to construct the *NList*

Input: An FCG $\tilde{G} = (V, \tilde{E}, \Omega)$ and an Association Density threshold δ .

Output: The multilevel linked list *NList* corresponding to \tilde{G} .

Algorithmic Steps:

- 1: for each vertex $v_i \in V$ do
- 2: Set $NList(v_i, n) \leftarrow v_i$, where $v_i \in V - \{v_i\}$, such that $\Omega_{v_i, NList(v_i, j)} \geq \Omega_{v_i, NList(v_i, k)}$, if $j < k$ and $NList(v_i, j) \neq NList(v_i, k)$, if $j \neq k, \forall n \in \{1, 2, \dots, |V| - 1\}$

- 3: end for
- 4: for each vertex $v_i \in V$ do
- 5: for $k=1$ to $|V| - 1$ do
- 6: if $\sum_{j=1}^k \Omega_{v_i, NList(v_i, j)} < \delta k$ then
- 7: $NList(v_i, n) \leftarrow \phi, \forall n \in \{k, k + 1, \dots, |V| - 1\}$
- 8: Proceed to next v_i breaking out of the inner for loop
- 9: end if
- 10: end for
- 11: end for

The actual unsupervised method that can locate the largest DAN in an FCG with the help of an *NList* is given in Algorithm 2. Here, for each of the vertices taken as an initial seed vertex once, a largest DAN, V_{let}^N , is produced by expansion and finally the largest one of them is selected as the final largest DAN. This heuristic expansion algorithm starts (step 3) with a seed vertex and expands it by including (step 6) the first vertex in the *Temp* list. Initially, the *Temp* list is taken (step 4) from the *NList* entry of the seed vertex. Then, both the *NList* and V_{let}^N are iteratively constructed by weighted aggregation (steps 11-12) using the *NList* until all the vertices are exhaustively taken (step 5) or the density constraint is violated (step 7).

An example of the process of combining two ordered lists by weighted aggregation is shown in Fig. (1). Here, the first list gets a weight of 2 while the second one gets 1 depending on their current sizes. The location numbers of the vertices in both the lists are multiplied with the corresponding weights and these two are sum up to finally prepare an ordered list to construct the new *Temp* list.

The average time complexity of this algorithm is $O(n^3 \log n)$, n being the order of the FCG. Although this complexity is of a higher order of n , but still this is reasonable considering the low dimension of miRNA expression data.

Algorithm 2 A method to find out the largest DAN from an FCG

Input: An FCG $\tilde{G} = (V, \tilde{E}, \Omega)$ and an Association Density threshold δ .

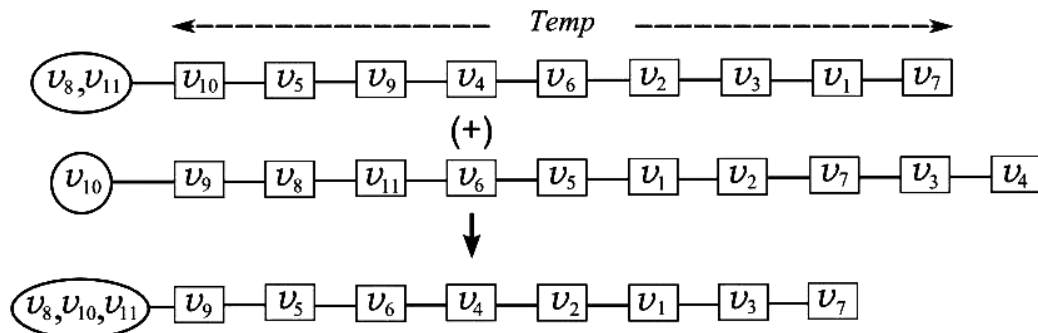


Fig. (1). Combining two neighboring lists ordered by density.

Output: The largest DAN $V_{let}^{N_{max}}$ in \tilde{G} with respect to δ .

Data Structure: The multilevel linked list $NList$ corresponding to \tilde{G} .

Algorithmic Steps:

- 1: $V_{let}^{N_{max}} \leftarrow \phi$
- 2: **for** each vertex $v_i \in V$ **do**
- 3: $V_{let}^N \leftarrow \{v_i\}$
- 4: $Temp(n) = NList(v_i, n), \forall n \in \{1, 2, \dots, |V| - 1\}$
- 5: **while** $|V_{let}^N| < |V|$ **do**
- 6: $V_{let}^N \leftarrow V_{let}^N \cup \{Temp(1)\}$
- 7: **if** $\mu_{V_{let}^N} < \delta$ **then**
- 8: $V_{let}^N \leftarrow V_{let}^N - \{Temp(1)\}$
- 9: Exit from the **for** loop
- 10: **else**
- 11: $Order(v_i) = Index(Temp(n), v_i) / (|V_{let}^N| - 1) + Index(NList(Temp(1), n), v_i),$
 $\forall v_i \in V$
- 12: $Temp(n) \leftarrow v_i$, where $v_i \in V - V_{let}^N$, such that $Order$
 $(Temp(i)) \leq Order(Temp(j))$, for each $i < j$
- 13: **end if**
- 14: **end while**
- 15: **if** $|V_{let}^N| > |V_{let}^{N_{max}}|$ **then**
- 16: $V_{let}^{N_{max}} \leftarrow V_{let}^N$
- 17: **end if**
- 18: **end for**

EMPIRICAL STUDY

The original dataset used for experimental analysis contained expression profiles of 264 human miRNAs. These are collected from postmortem prefrontal cortex tissue in 21 non-schizophrenic and 15 schizophrenia affected people [4]. A total of 239 miRNAs was considered over these 36 cases which does not contain any missing values. Statistically, the microarray data have minimum expression value = 6.03, maximum expression value = 15.88, average expression value = 7.27 and standard deviation of the expression over the entire dataset = 1.38.

The RCS values are computed for each miRNA pair and these are plotted with respect to the values of (X_d, Y_d) and $Cor(X_{nd}@X_d, Y_{nd}@Y_d)$ in Fig. (2). The range of the RCS values obtained was found to be within $[0, 0.511]$. On preparing an integrated histogram plot (shown in Fig. 3) over the column vectors of the RCS scores obtained for the miRNA pairs, we received a long tail. This is due to the property of power law degree distribution found in the FCG.

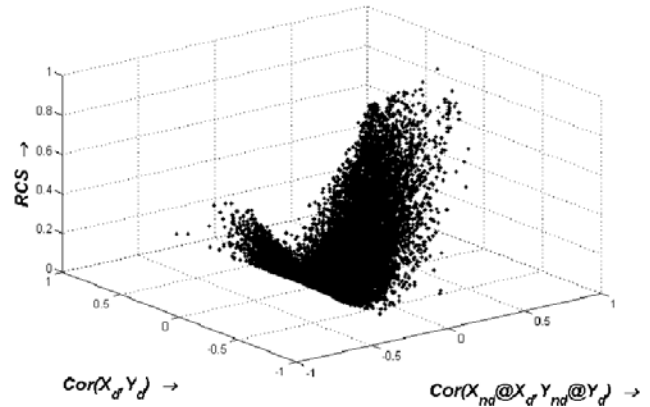


Fig. (2). Original RCS values computed from the schizophrenia dataset.

It highlights the scale-freeness of the network constructed from the Schizophrenia dataset.

For analyzing the biological significance of the RCS computation, the topmost five (~0.02% of the total) miRNA pairs, which comprises eight separate miRNAs (hsa-miR-7, hsa-miR-9, hsa-miR-27a, hsa-miR-27b, hsa-miR-137, hsa-miR-218, hsa-miR-368 and hsa-miR-376a), those having higher RCS scores are selected from the complete set of miRNAs. These are listed in Table 1. In most of the cases, the higher correlation values obtained in the combination of non-diseased and diseased subjects (column 3) get reduced on considering only the diseased subjects (column 4). It indicates a possible functional disorder that might have decreased the co-expression between the miRNA pairs within a diseased cell. Only, in the case of miRNA pairs hsa-miR-27b and hsa-miR-27a yielding $RCS = 0.504$, this deviation is relatively small. However, a recent work has shown the involvement of hsa-miR-27a in regulating schizophrenia [7]. The unsupervised finding employed here is thus capable to figure out significant schizophrenia associated miRNAs. On performing a comparison of these miRNAs vis-a-vis schizophrenia, we found a match of 5 out of total 8 miRNAs which gives a hypergeometric p-value of 0.002 (with a background

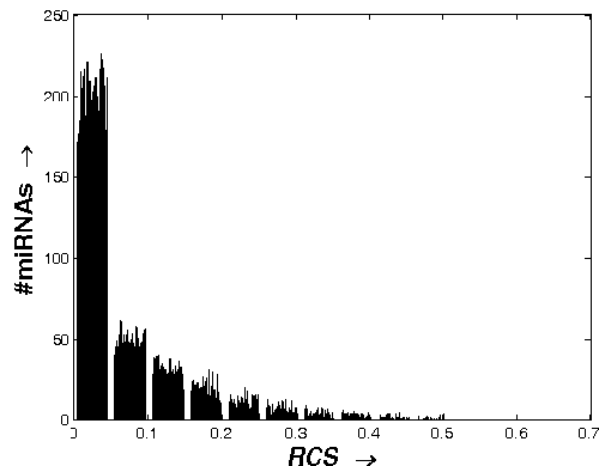


Fig. (3). Histogram of #miRNAs over RCS values.

Table 1. Most Significant miRNA Pairs Found

miRNA1 ($m1$) miRNA2 ($m2$)	hsa-miR-7 hsa-miR-218	hsa-miR-27b hsa-miR-27a	hsa-miR-7 hsa-miR-9	hsa-miR-376a hsa-miR-218	hsa-miR-137 hsa-miR-368
$Cor(m1_{nd}, m2_{nd})$	0.870	0.978	0.898	0.935	0.855
$Cor(m1_d, m2_d)$	0.461	0.912	0.564	0.173	0.257
$Cor(m1_{nd}@m1_d, m2_{nd}@m2_d)$	0.787	0.961	0.807	0.711	0.723
$RCS(m1, m2)$	0.511	0.504	0.494	0.491	0.490

truth of 30 detected miRNAs out of 239) depending on two established results [4, 7].

Now, we apply Algorithm 1 to construct the $NList$ and consecutively employing this $NList$ in Algorithm 2 to mine the largest DAN from the RCS matrix, which can be thought of as an FCG by altering $RCS(X, Y)$ values from 0 to 1. Keeping the Association Density threshold as high as 0.15 (lower than top ~9.6% of the total RCS values), we have identified a largest DAN comprising of 32 miRNAs, with higher RCS values, out of the total 239 (~13%) miRNAs. To show the cohesiveness within the DAN identified from the Schizophrenia dataset the quartile deviation plot (QDP) [13] has been prepared. It is shown in Fig. (4). QDP is a tool for visualizing the compactness of expression values within a cluster. It becomes evident from Fig. (4) that the deviation in expression values within this set of miRNAs is very small and therefore they are very compact in nature.

We have validated the significance of the result by computing the probability of occurrence of the event. Seeing that the randomization model suffers from the loss of accuracy, the chance of receiving this result has been measured by computing the p-value over a hypergeometric distribution. The results on the biological validation of the largest DAN on the basis of schizophrenia literature are given in Table 2. Each of the matching miRNAs found in literature are marked with a tick over the rows of the table.

The p -value is a well-known measure to compute the probability of occurrence of an event. We compute the p -value of the result obtained by the above study. Suppose, an event is observed n times out of total N observations and

given the evidence that the event originally occurs e times out of E total cases. Then, the p-value is computed assuming a hypergeometric distribution as given follows.

$$p\text{-value} = \sum_{i=n}^N \frac{\binom{e}{i} \binom{E-e}{N-i}}{\binom{E}{N}} \quad (6)$$

The hypergeometric p -value obtained was as low as 2.9×10^{-7} . This is due to the finding of a high number of resembling miRNAs (11 in total) depending on an established result selecting 16 miRNAs as a regulating module of schizophrenia from a total of 264 miRNAs [4]. Another previous study identified 16 miRNAs out of 101 accumulated by literature survey [7] of which we received a match of 4 miRNAs out of the 32 found in the largest DAN. Another miRNA (hsamiR-130b), selected by a genetic investigation coupled with gene expression analysis of miRNA in an independent study [15], is also identified within the largest DAN. Thus, the largest DAN produces a biologically well-validated group regulating the specific disease studied.

In the experimental analysis, the paper uses two different schemes for selecting miRNA modules regulating schizophrenia. The first one selects a group of 8 miRNAs (hsa-miR-7, hsa-miR-9, hsa-miR-27a, hsa-miR-27b, hsa-miR-137, hsa-miR-218, hsa-miR-368, hsa-miR-376a) by ordering the RCS values without any heuristics. Interestingly, some of these miRNAs are well known to be enriched in brain. A very early study identified hsa-miR-9 to be active in brain [2]. Because of its enrichment in the brain, schizophrenia might appear in case of functional disorder. The other scheme of selecting miRNA module regulating schizophrenia employs computational identification of the largest DAN. This produces comparatively better modules as a statistical cut-off participation of each of the miRNAs is applied in the module. This could be biologically good as they can overlook noisy data. These modules can also be helpful in studying the complex regulatory network that might be formed due to disorders, thus, creating diseases.

Earlier researches have studied the activities of single miRNAs like hsa-miR-130b [15], hsa-miR-181b [6] or expression of multiple miRNAs [4, 7] for studying their specificity, fold-change or their dysregulation. These studies indirectly suggest the involvement of miRNAs in regulating diseases. The present work highlights an unsupervised method that establishes that tissue-specific expression profiles are useful in the study of disease regulating module of miRNAs.

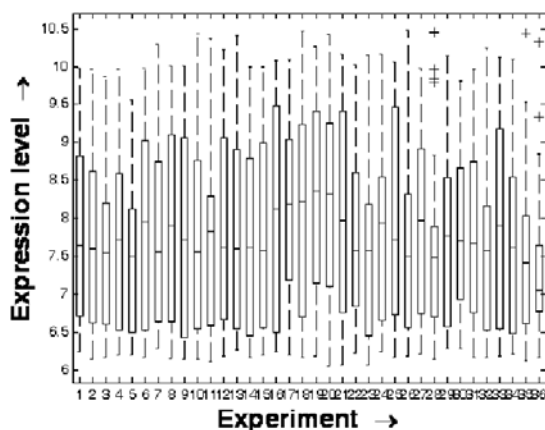


Fig. (4). QDP of the largest DAN identified from the Schizophrenia dataset.

Table 2. Biological Validation of the Set of miRNAs (Located as the Largest DAN Present in the FCG Constructed from the Rat CNS Dataset) to be Associated with Schizophrenia Based on Published Knowledge

MiRNAs in the largest DAN	Perkins <i>et al.</i> [10]	Burmistrova <i>et al.</i> [6]	Hansen <i>et al.</i> [8]
hsa-miR-30a*			
hsa-miR-29a	√		
hsa-miR-29c	√		
hsa-miR-30b	√		
hsa-miR-26b	√		
hsa-miR-195	√		
hsa-miR-92	√		
hsa-miR-16			
hsa-miR-126			
hsa-miR-9*			
hsa-miR-22			
hsa-miR-7	√		√
hsa-miR-29b	√		
hsa-miR-135			
hsa-miR-143			
hsa-miR-218			
hsa-miR-184			√
hsa-miR-30c	√		
hsa-miR-345			
hsa-miR-30c			
hsa-miR-30d	√		
hsa-miR-9			√
hsa-miR-130b		√	
hsa-miR-520e			
hsa-miR-190			
hsa-miR-106a			
hsa-miR-20b	√		
hsa-miR-182*			
hsa-miR-182			√
hsa-miR-200b			
hsa-miR-181a			
hsa-miR-137			

The miRNAs (hsa-miR-30a* and the others in the list) found by this study otherwise unexplored earlier is to be experimented biologically for judging their evidence in regulating schizophrenia. We have compared our approach with a widely used method for selecting significant genes from a microarray analysis. This method is commonly abbreviated as SAM [16]. The result obtained by using the SAM on the same dataset is shown in Fig. (5).

To keep the analysis comparable, we set the tuning parameter $\delta = 1.24$ for SAM such that 32 miRNAs are selected as significant. We chose the quantitative response type for

the study and the results are obtained after 100 permutations. We found a set of 32 significant miRNAs with median number of false positives = 0, false discovery rate = 0%, tail strength = 51.7% and se = 24%. Amongst the miRNAs selected by SAM 28 are found to be positive and 4 are negative.

We verified the significance of the miRNAs selected by SAM as compared to the miRNAs identified by the proposed method. The comparison is made based on the set of miRNAs having established support to be associated with schizophrenia neuropathology. For two cases (shown under

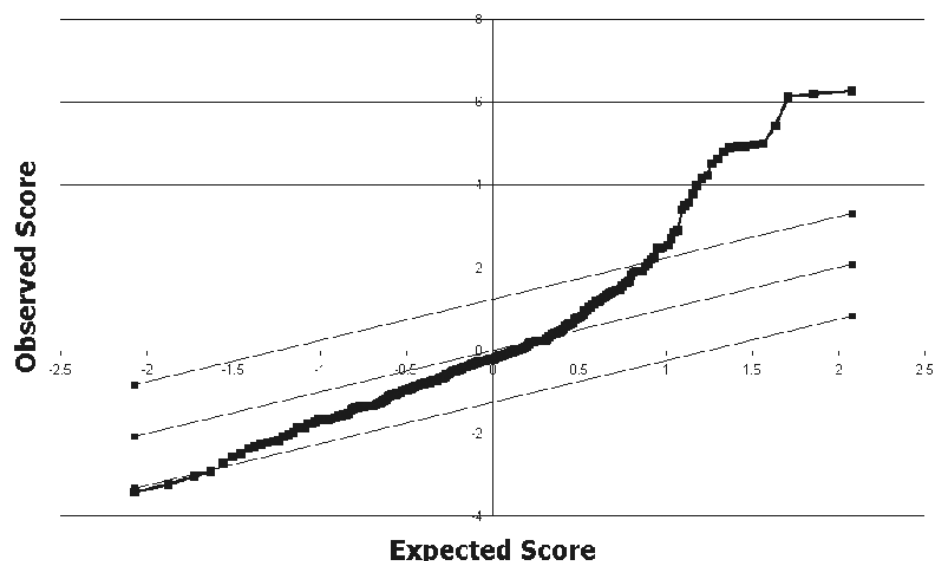


Fig. (5). The SAM plotsheet obtained by the significance analysis of the schizophrenia microarray dataset.

Table 3. Comparative Study with the Well-Known Gene Selection Method SAM

Method	#miRNAs Selected	Perkins <i>et al.</i> [10]	Burmistrova <i>et al.</i> [6]	Hansen <i>et al.</i> [8]
SAM [13]	32	11	1	1
Proposed	32	11	1	4

column 3 and 4 of Table 3), we found the results are comparatively same in finding significant miRNAs related with schizophrenia. However, by a comparison based on the result from a recent work [7] (shown under column 5 of Table 3), only one miRNA (hsa-miR-184) is found to be selected by SAM whereas the proposed one found four miRNAs that have established biological support. Thus, the unsupervised method is very effective in locating the miRNAs regulating schizophrenia.

CONCLUDING REMARKS

The current paper exploits patient-specific expression profiles of miRNAs [17] for the analysis of diseases. It carries out an unsupervised study on the altering activities of miRNAs in regulating Schizophrenia disease. Introducing a novel relative co-expression measure, the proposed methodology establishes a strong possibility of identifying the distinction in co-expressibility/co-repressibility of miRNAs between patients and normal people. This helps in the probabilistic selection of the module of miRNAs responsible for the specific disease. The result obtained through the study gives evidence, both statistically and biologically, of the novelty of the analysis. Similar empirical studies are in progress for studying expression profiles over various cancer patients.

SUPPLEMENTARY MATERIALS

Supplementary details along with the color figures are available at the webpage of the corresponding author: http://www.isical.ac.in/~malay_r/Supplementary.html

ACKNOWLEDGEMENT

The first author gratefully acknowledges the financial support from the grant no.- DST/SJF/ET-02/2006-07 under the Swarnajayanti Fellowship scheme of the Department of Science and Technology, Government of India.

REFERENCES

- [1] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **2004**; 116(2): 281-97.
- [2] Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **2002**; 12: 735-9.
- [3] Shalgi R, Lieber D, Oren M, Pilpel Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol* **2007**; 3: e131.
- [4] Perkins DO, Jeffries CD, Jarskog LF, *et al.* microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. *Genome Biol* **2007**; 8(2): R27.
- [5] Brown D, Conrad R, Devroe E, *et al.* Methods and compositions involving microRNA. *US Patent 20070161004* **2007**; A1.
- [6] Beveridge NJ, Tooney PA, Carroll AP, *et al.* Dysregulation of miRNA 181b in the temporal cortex in schizophrenia. *Hum Mol Gen* **2008**; 17(8): 1156-68.
- [7] Hansen T, Olsen L, Lindow M, *et al.* Brain expressed microRNAs implicated in schizophrenia etiology. *PLoS ONE* **2007**; 2(9): e873.
- [8] Wang W, Rajeev BW, Stromberg AJ, *et al.* The expression of microRNA mir-107 decreases early in alzheimers disease and may accelerate disease progression through regulation of β -site amyloid precursor protein-cleaving enzyme 1. *J Neurosci* **2008**; 28(5): 1213-23.
- [9] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **1998**; 95: 14863-8.
- [10] Bandyopadhyay S, Mukhopadhyay A, Maulik U. An Improved Algorithm for Clustering Gene Expression Data. *Bioinformatics* **2007**; 23(21): 2859-65.

- [11] Bhattacharyya M, Bandyopadhyay S. Integration of co-expression networks for gene clustering. Proceedings of the 7th International Conference on Advances in Pattern Recognition. Kolkata, India **2009**; pp. 355-8.
- [12] Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **2005**; 11: 241-7.
- [13] Bandyopadhyay S, Bhattacharyya M. Analyzing miRNA co-expression networks to explore TF-miRNA regulation. *BMC Bioinformatics* **2009**; 10: 163.
- [14] Perkins DO, Jeffries CD, Sullivan P. Expanding the 'central dogma': the regulatory role of nonprotein coding genes and implications for the genetic liability to schizophrenia. *Mol Psychiatry* **2005**; 10: 69-78.
- [15] Burmistrova O, Goltsov AY, Abramova LI, Kaleda VG, Orlova VA, Rogaev EI. MicroRNA in schizophrenia: Genetic and expression analysis of mir-130b (22q11). *Biochemistry* **2007**; 72(5): 578-82.
- [16] Tusher V, Tibshirani R, Chu G. Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. *Proc Natl Acad Sci USA* **2001**; 98: 5116-21.
- [17] Wang S-T, Li C, Liu L. miRNA microarray technology in miRNA Profiling. *Curr Bioinformatics* **2009**; 4(2): 141-8.

Received: January 21, 2009

Revised: April 29, 2009

Accepted: June 18, 2009