Context-Specific miRNA Regulation Network Predicts Cancer Prognosis

Xionghui Zhou[1], Juan Liu[1,§], Changning Liu[3], Simon Rayner[4], Fengji Liang[2], Jingfang Ju[5], Yinghui Li[2], Shanguang Chen[2], Jianghui Xiong[2,1,§]

1 School of Computer Science, Wuhan University, Wuhan, P.R. China

2 Bioinformatics Group and Data Coordination Center, State Key Lab of Space Medicine Fundamentals and Application (SMFA), China Astronaut Research and Training Center (ACC), Beijing, P.R. China

3 Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, P.R. China

4 Bioinformatics Group, State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, P.R. China

5 Department of Pathology, Stony Brook University Medical Center, New York, USA

§ Correspondence should be addressed to J.X. (laserxiong@gmail.com) or J.L. (liujuan@whu.edu.cn).

Abstract—**MicroRNAs can regulate hundreds of target genes and play a pivotal role in a broad range of biological process. However, relatively little is known about how these highly connected miRNAs-target networks are remodelled in the context of various diseases. Here we examine the dynamic alteration of context-specific miRNA regulation to determine whether modified microRNAs regulation on specific biological processes is a useful information source for predicting cancer prognosis. A new concept, Context-specific miRNA activity (CoMi activity) is introduced to describe the statistical difference between the expression level of a miRNA's target genes and non-targets genes within a given gene set (context).**

The microarray gene expression profile of brain tumors from 356 patients (The Cancer Genome Atlas dataset) was converted into a CoMi activity pattern, and showed significant positive correlation with the corresponding miRNA expression pattern. In a breast cancer cohort, the differential CoMi activity between good prognosis (longer survival) vs. bad prognosis patients forms a scale-free network, which highlighted a group of important cancer-related microRNAs and GO terms, e.g. hsa-miR-34a and 'cell adhesion'. Then two breast cancer cohorts were used in outcome prediction in an independent test. Using a popular T-test feature selection method and a support vector machine (SVM) classifier with 10-fold cross-validation, the CoMi activity feature achieves an area under curve (AUC) of 0.7155, better than the AUC value of 0.6339 for feature selection based on mRNA expression. In an independent test, CoMi feature selection achieved an AUC of 0.6874. Survival analysis also shows signatures defined by CoMi activity was predictive of survival and superior to mRNAs signatures.

In short, we have demonstrated the first interrogation of dynamic remodeling of context specific miRNAs regulation networks in cancer. The altered microRNAs regulation on specific contexts could be used to predict cancer prognosis and reveal hidden levels of cancer regulation mechanisms.

Keywords—context-specific microRNA activity; network biology; microRNA; microRNA Regulation Network; prognosis prediction; survival analysis

I. Introduction

MicroRNAs (miRNAs) are a class of ~22 nt endogenous small regulatory RNA molecules that regulate target mRNAs either via translation repression or mRNA degradation [1, 2]. The regulatory activities of miRNAs are involved in various biological processes such as development, proliferation, apoptosis, stress response, and cancer development, progression and metastasis [3-9]. In addition to this, many miRNAs such as let-7, miR-125, miR-17-92, miR-124, miR-155 and miR-223 have been reported to be related to cancer development or outcome [10]. Previous works show that infer miRNA activity by combining gene expression with miRNA target prediction is feasible [11-13]. Our group constructed a prognosis-related synergistic gene-gene interaction network as an efficient tool for pre-clinical drug prioritization. We found that the interaction of microRNA target gene sets with other gene modules is important to the robustness of cancer gene network [14], suggests that cancer prognosis might be modulated by miRNAs regulation network. However, few studies have examined the correlation between miRNAs with cancer outcome in the global context of microRNAs and their target gene network.

Recently, studies [15-18] have revealed that mRNA expression profiles could be used to effectively predict cancer prognosis. Dealing with breast cancer, [17] and [18] separately identified a 70 gene signature and 76 gene signature derived from mRNA expression profiles to classify cancer samples into a good-outcome group and a bad-outcome group with an accuracy between 60%-70% [19]. However, there are two main problems with these signatures. Firstly, when these signatures were applied to other data to perform an independent test, the performance declined significantly and secondly, there is little overlap between these signatures. Subsequently, alternative feature spaces such as protein-protein interaction networks, pathways and GO terms have been used to predict cancer outcome [19-23].In this study we proposed that context-specific miRNA activity networks might be a useful indicator of cancer prognosis. To test this hypothesis, a context-specific miRNA activity (CoMi activity) was introduced to describe the statistical difference between the expression level of miRNA target genes and non-target genes within specific

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

contexts. Here gene sets defined by Gene Ontology Terms (Gene Ontology Biological Process - GOBP, Gene Ontology Cell Context - GOCC) were used to define context. For a gene expression pattern of a breast cancer patient, we could generate a series of CoMi activities by combining information from microRNAs- target gene pairs. We then examined whether this new feature space (i.e. CoMi activity) could provide discriminating power for different prognostic groups of cancer patients.

II. Results and Discussions

A. *The pipeline used to identify Context specific miRNA activity*

To check microRNAs' regulation on a specific context, we proposed the following method to calculate CoMi activity (Fig. 1).

To calculate a miRNA's CoMi activity on a context (here we use GOBP as an example, unless stated otherwise, all subsequent context is based on GOBP), miRNA targets are predicted using a range of target prediction software (e.g. TargetScan). We then extract the intersection of the miRNA target set and the GO term gene set. Secondly, a two step filter method was applied to the miRNAs-GO term combination. Firstly, if the number of elements (probeSet of HG-U133A,see method) in the intersection set is smaller than 20, we discard the GOBP term because the intersection set is too small for T-test calculation. Secondly, the significance of the intersection set is estimated using an appropriate statistical method, here the hypergeometric distribution (see method) was chosen. Only the miRNA-GO term pairs that had a P-value smaller than 0.05 (We performed multiple testing adjustments and checked the P- value distributions, when P-value is smaller than 0.05, the FDR is not more than 0.005, data not shown) were passed to the next stage. Thirdly, for each biological sample profiled by microarray gene expression (Fig. 1b), we then constructed the expression vector for two sets: one for miRNAs target genes, another for non-target genes. Then the statistical difference between these two sets was calculated (Fig. 1c). Here the 2-sample T-test was employed and the T-score was used to represent the CoMi activity of a given miRNA on a specific GO term in each biological sample. According to traditional view of a miRNA's action on target gene abundance (down-regulation), if the expressions of target genes are significantly lower than those of non-target genes, then the CoMi activity is positive. If there is no significant difference between the expression distributions of the target genes vs. non-target genes, the CoMi activity is close to zero. In theory, the distribution of target genes expression levels might, on average, also be higher than those of non-target genes, we consider this might be due to the action of other factors.

Using the method outlined above, patients' CoMi activity patterns can be generated from their corresponding mRNA expression pattern. Although a miRNA expression profile could be interrogated by a simple miRNAs microarray or qRT-PCR, a clear advantage of the CoMi activity metric is that it reveals a comprehensive regulatory relationship between the miRNA and a specific context, which is often hidden in the gene expression profiles. As describe below, a CoMi activity network derived from a CoMi activity pattern might also be used to interrogate the synergistic or antagonistic action of multiple microRNAs on an individual biological process (GO term).

B. *Rationality, robustness and repeatability of the CoMi activity metric*

In order to validate the relevance of CoMi activity, we calculated the similarity between the CoMi activity pattern and the corresponding miRNA expression pattern (See methods). The brain cancer dataset which download from The Cancer Genome Atlas dataset (TCGA: http://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm) contains simultaneous measurement of mRNA expression and miRNA expression of 356 patients. We used this data to generate the CoMi activity patterns based on the mRNA expression profiles. When combined with various GO term, one miRNA have multiple CoMi activity values. We summarized them into one single value (see method). The Spearman correlation coefficient was calculated between the CoMi activity pattern and the miRNA expression pattern. From Fig. 2, it can be seen that all the CoMi activity patterns derived from GOBP (Gene Ontology biological process) have a significant (positive) correlation with their miRNA expression profiles (miRNA-GOCC shows similar results, data not shown). The correlation values are ranged from 0.25 to 0.45, and all of the P-values are smaller than 10e-4.

To further investigate the robustness of our CoMi estimation method, we recalculated the correlation using different microRNAs target prediction methods. All CoMi activity estimations based on these target predictions (TargetScan [2, 24, 25], miRanda [26, 27], ExprTarget [28], RNA22 [29], TargetScan ∪ RNA22), showed positive correlation with miRNA expression profiles (Fig. 2-Fig. 6). Furthermore, they were all significantly correlated with each other (data not shown). Thus our method is robust regardless of the miRNAs – target gene data source. Interestingly, the union set of TargetScan and RNA22 performed the best (Highest spearman correlation and lowest P-value, data not shown). Most miRNA target prediction algorithms (TargetScan, miRanda, PicTar) are based on the assumption that miRNAs bind to highly conserved seed matches in the 3'UTRs of targeted genes, but recent studies demonstrated that non-conserved sites can be as important as conserved sites [1, 2, 30]. Our results might suggests that a combination of these two strategies –conserved site analysis (TargetScan) and non-conserved site analysis (RNA22) might be complementary to each other and the union of these two miRNAs target gene sets represent a more comprehensive set of target predictions. In the following calculations, the union target gene set of TargetScan and RNA22 was the default choice.

To further validate the reproducibility of the calculated CoMi signature for cancer prognosis prediction, we checked the overlap of informative CoMis (miRNAs-GO term pairs) selected from two independent breast cancer data cohorts: Wang's data (GSE2034) [18] and GSE7390 [31]. The T-test was used to select the CoMi activity (miRNAs-GOBP) that was significantly differentially expressed in the good outcome group and bad outcome group respectively. We

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

calculated the overlaps of the features selected from two different breast cancer cohorts in different P-value grads, all of which are significant (table 3). Thus our method identified a consistent miRNAs regulation network involved in cancer prognosis related processes (e.g., recurrence/metastasis) from two independent breast cancer cohorts (data not shown).

C. *Context specific miRNA activity network (CoMiNet) could reveal hidden layer of mechanism related to cancer prognosis*

We used the 200 best features (For details, see table 4) selected by T-test in the bad-outcome group and good-outcome group in the Wang dataset (GSE2034) to construct CoMiNet (see method). In the network (Fig. 7a), the blue node represent miRNA, the yellow nodes represent GO terms (GOBP), red edges represent positive CoMi activity (positive t-score, higher miRNAs regulation activity in bad/good outcome groups), green edges represent negative CoMi activity (negative t-score, lower miRNAs regulation activity in bad/good outcome group), and the size of the nodes is proportional to their degrees.

We first investigated the topology of the CoMi network. Both the in-degree (GO term) and the out-degree (miRNA) fits a power law distribution, suggesting that the CoMi network is a typical scale free network (Fig. 7).

In the network, hsa-miR-34a (degree=5) are among the most connected (i.e., hub node) miRNAs. In additional to this, its family hsa-miR-34b (degree=2) are also in our CoMiNet. hsa-miR-34a is a well-known tumor suppressor, showing tumor suppressor activity in breast, pancreatic and colon cancers via the p53 network [32, 33],while hsa-miR-34b is significantly associated with the presence of breast cancer's metastases [33]. In the sub-network (Fig. 8b) of hsa-miR-34a and hsa-miR-34b, hsa-miR-34a is connected to a group of cancer-related GO terms: 'cell migration', 'cell death', 'negative regulation of apoptosis', 'post-embryonic development', while hsa-miR-34b connects to 'negative regulation of cell proliferation' and 'regulation of transcription from RNA polymerase II promoter'. In additional to this, hsa-miR-9 (degree=5) is also a important miRNA in our network, reported to be associated with vascular invasion and lymph node metastasis in breast cancer [34]; the sub-network of hsa-miR-9 (Fig. 8c) is direct to GO terms that play important roles in cancer metastasis, such as 'cell adhesion', 'cell-cell adhesion', 'response to organic cyclic substance', 'chemotaxis' and 'G-protein coupled receptor protein signaling pathway'. Other miRNAs, such as hsa-miR-124, a notable tumor suppressor in all cancers [33], is found to connect to 'Apoptosis', hsa-miR-221 and hsa-miR-222 are directed to "Oncogenic activity miRNA in breast cancer" [33], and hsa-miR-373 is a activator of metastasis of breast cancer [34].

The CoMi network also highlights many known cancer related miRNA families or clusters such as the let-7 family (hsa-let-7a, hsa-let-7b, hsa-let-7c, hsa-let-7d, hsa-let-7f, hsa-let-7g, and hsa-let-7i), which are tumor suppressors[33], In Fig. 8d, has-let-7a, has-let-7d and has-let-7i all negatively linked to 'immune response', which may indicate that the three let-7 family miRNAs synergistically act on 'immune response'. The miR-17-92 cluster (hsa-miR-18a, hsa-miR-19a, hsa-miR-19b) is another cluster in our CoMiNet , which shows oncogenic activity in various cancers include breast cancer [33]. Within the miR-520 cluster (hsa-miR-515-5p, hsa-miR-519d, hsa-miR-520a-3p, hsa-miR-520b, hsa-miR-520c-3p, hsa-miR-520d-3p, hsa-miR-520e, hsa-miR-520g, hsa-miR-526b), hsa-miR-520b is reported to regulate migration of breast cancer cells [35] and hsa-miR-520c is a activator of metastasis [34]. The hsa-miR-15 family (hsa-miR-15a, hsa-miR-15b) and hsa-181 family (hsa-miR-181a, hsa-miR-181b, hsa-miR-181d) are tumor suppressors [33].

The high degree GO term nodes in the network also show a relationship with cancer. 'cell adhesion' (degree=8) is a pivotal biological process involved in metastasis of breast cancer [36]. 'negative regulation of apoptosis' (degree=8) is a hub node in the network while apoptosis regulatory proteins have prognostic significance in breast cancer patients [37]. Finally, 'immune response' (degree=7) and 'response to drug' (degree=7) also play important roles in prognosis of cancer.

The combinatory miRNA-miRNA regulation on common target gene modules might reveal the synergistic mechanism of miRNAs regulation [38]. In our CoMiNet constructed from high ranking features, three tumor suppressor[33] miRNAs hsa-let-7f, hsa-181a and hsa-181d are negatively connected to 'anti-apoptosis', while hsa-miR-200a* (miR-200 is reported to enhances mouse breast cancer cell colonization to form distant metastases [39]) is positively directed to 'anti-apoptosis' (Fig. 8e). These finding may indicate that three tumor suppressor miRNAs and one onco-miRNA work synergistically on the same biological process 'anti-apoptosis' to influence the metastasis of breast cancer patients. In the bad-outcome group compared with good-outcome group, the decreasing activity of three tumor suppressor miRNAs (green edge) and the increase in one onco-miRNA's activity (red edge) increases in turn the metastasis risk of breast cancer patients. In another subnetwork, hsa-miR-125b and hsa-miR-520b are both negatively connected to 'DNA repair' (Fig. 8f). Hsa-miR-125b is reported to be down-regulated in breast cancer patients [40], and hsa-miR-520b is reported to regulate migration of breast cancer cells, additionally, polymorphisms in DNA repair genes have associations with cancer risk [41]. From these findings, we may argue that mir-520b and mir-125b co-regulate the biological process 'DNA repair' to influence the prognosis of breast cancer.

D. *CoMi activity patterns demonstrate better prognosis prediction performance in breast cancer than mRNA expression patterns*

We used CoMi activity patterns to predict distant-metastasis over 5 years on lymph node-negative primary breast cancers. 262 samples from the Wang dataset were used for training, and 164 samples from the GSE7390 dataset were used to perform an independent test. In both the training data set and the independent test data set, the CoMi activity pattern (GOBP) outperformed the mRNA expression profile, and the CoMi activity pattern achieved an AUC of 0.7155 and 0.6874 for the two datasets respectively, while the corresponding mRNA expression profiles achieved an

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

AUC of 0.6339 and 0.6647 respectively (Fig. 9). At the same time, the best accuracy of the CoMi feature is 69% on GSE2034,while the mRNA classifier is 65% (Fig.10,the sensitivity and specificity result in Wang data set of different feature spaces are shown in Fig. 11 and Fig. 12). From the ROC curve in Fig. 13, it is obvious that the CoMi activity pattern classifier can achieve better performance than the mRNA expression profile classifier. Similar results of the CoMi activity pattern for GOCC terms can be observed (data not shown).

Furthermore, we combined these two types of CoMi activity features (miRNA-GOBP, miRNA-GOCC). The combined features show marginally improved performance (AUC equals 0.7214 and 0.6954 respectively, data not shown).

We also use Cox proportional hazards regression to regress the value of each feature to the survival time on GSE2034 (22215 features in the mRNA expression pattern and 6314 features in the CoMi activity pattern of GOBP), we then calculated the FDR of the two kinds of features. In the CoMi activity pattern, there are 16 features for which the FDR is smaller than 0.1, while only 5 features are found in mRNA expression pattern (Table 5). After this, we used the log-rank test to select the best feature of both the CoMi activity pattern and the mRNA pattern to generate a survival curve; the 'best' feature (hsa-miR-154, response to organic cyclic substance) of the CoMi activity pattern (GOBP) divided the 286 patients into two group distinctly with HR of 2.79 ([95% CI 1.84 4.22], log rank P-value of 3.12E-07) (Fig. 14). Hsa-miR-154 has been reported to be negatively correlated with ER positivity in early breast tumors [42].The targets of hsa-miR-154 on GO term 'response to organic cyclic substance' was shown in Table 6.

The above results show that the CoMi activity feature space is more informative than the mRNAs feature space in various metrics (classification performance, log rank test, cox regression).

III. Conclusions

To our best knowledge, this is the first attempt to interrogate the global miRNAs regulation network by discover the regulation relationship between the miRNAs and contexts. We have demonstrated the proposed CoMi activity estimation method is a robust and repeatable method to discover the regulation mechanism of miRNA on specific biological processes. The CoMi activity pattern is a useful new feature space that can be used to predict breast cancer prognosis. Distinct from miRNA expression profile, the CoMi activity pattern can reveal regulation mechanisms that exist between miRNAs and gene modules, such as a miRNA regulates multiple gene modules and several miRNAs co-regulate one gene module. By uncovering the context specific miRNA activity pattern, the latent regulation within this highly rewired gene regulation network can be used to understand cancer prognosis and disease processes.

IV. Methods

A. Materials

mRNA expression profiles and miRNA expression profiles of brain cancer samples was download from TCGA (The Cancer Genome Atlas), mRNA microarray analysis was performed with Affymetrix U133A Genechips For miRNA expression level data, miRNA profiling from total RNA was performed using an Agilent 8 x 15K Human miRNA-specific microarray. We use level 3 (expression calls for miRNAs per sample, see TCGA data guidance, http://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm) data to calculate the similarity between the miRNA activity pattern and miRNA expression pattern. After eliminating the repeated samples and samples which only contain a miRNA expression profile or a mRNA expression profile, 356 unique samples with both mRNA expression profile and miRNA level 3 expression profiles remained.

Breast cancer samples were obtained from the GEO database (series entry GSE2034 and GSE7390), all of which were lymph node-negative primary breast cancers with clinic information (distant metastases days, distant metastases status etc). There are 286 samples in GSE2034 and 198 samples in GSE7390. Microarray analyses were performed on both of datasets using the Affymetrix U133a GeneChip. Expression values for each gene were calculated using Affymetrix GeneChip analysis software MAS 5.0 [18, 31]. In GSE2034, each data was log-transformed by base two to make the mRNA expression pattern fit a normal distribution.

B. miRNA target prediction tools

The following miRNA target predicting tools were used: TargetScan release 5.1 [2, 24, 25], RNA22 [29], miRanda Version 5.0 [26, 27] and ExprTarget [28] (Table2).

C. CoMi activity calculation

Two categories in the Gene Ontology were used in our analysis of gene modules: Biological Process, and Cellular Component (geneontology.org). The mapping probe sets ID (HG-U133A) to Gene Ontology was downloaded from http://www.biomart.org. All genes associated with one GO term were defined as one gene module and the module was named according to the name/title of GO terms.

To check whether the intersection set of set A (miRNAs target gene set) and B (a set of genes assigned to one GO term) is significant, the hypergeometric distribution (1) was calculated, and the final P-value p_2 is calculated by: $p_2=1-p$.

$$P = F(x/M,K,N) = \sum_{i=0}^{x} \frac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}} \qquad (1)$$

Where M is the size of the Universal set, x is the size of intersection set, K is the size of set A and N is the size of the set B. This equation was also used to check the significance of the intersection of the selected features between two breast cancer data set.

The two samples T-test (2) was used to calculate the difference between the expression vector of miRNAs target genes and the expression vector of non-target genes.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

$$t=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} \quad (2)$$

where $\overline{X}$ and $\overline{Y}$ are the sample means, s_x and s_y are the sample standard deviations, and n and m are the sample sizes.

D. Similarity calculation between miRNAs expression and CoMi activity

We calculated the Spearman correlation coefficient between the miRNA CoMi activity pattern and the miRNA expression pattern. In theory, there are two kinds of correlations (All miRNA expression values of a sample vs. all CoMi activity scores of a sample; a miRNA's expression value across all samples vs a miRNA's CoMi activity score across all samples), here the former correlation was used in our calculation. The brain cancer dataset from TCGA (see methods, dataset above) was used in our project. 356 mRNA expression profiles were converting into 356 CoMi activity patterns. For each miRNA, if combining with N GO terms, there will be N CoMi activities for this miRNA. Thus the multiple CoMi activities was merged into one single value for each miRNA as follows:

$$CoMi_{ALL}=\sum_{i=1}^{N}|CoMi_i| \quad (3)$$

Then the 356 mRNA expression patterns were converted into their corresponding miRNA CoMi activity patterns. Finally, the Spearman correlation coefficient was calculated between the 356 miRNA CoMi activity patterns and their corresponding miRNA expression pattern.

E. Cancer prognosis prediction

We classified breast cancer samples into high or low risk groups in the GSE2034 and GSE7390 datasets. The two groups (good-outcome group and bad-outcome group) were determined according to whether the samples developed distant-metastasis over 5 years (for details, see Table1). In GSE2034, 5 times 10-fold cross-validation was used, 9 out of 10 of the samples in GSE2034 were used to train and the remaining part was used for testing. At the same time, all the samples in GSE7390 were used to perform an independent test. A support vector machine algorithm implemented in the lib-svm software package [43] was used for classification. The area under curve (AUC) was used for evaluation.

F. Survival analysis

The univariate Cox scores and hazard ratio for each of the miRNAs-Gene Module and mRNA was calculated using the Matlab function coxphfit (Matlab version 2009b). The FDR of Cox proportional hazards regression analysis was calculated using Matlab function mafdr. The log-rank calculation and generation of the survival curve was performed using a Matlab package (http://www.mathworks.com/matlabcentral/fileexchange/22317).

G. Network topology analysis and visualization

We selected the best 200 features (selected by T-test) of GSE2034 to construct the miRNA-GOBP network. There are two kinds of nodes in the network, the miRNA and the GO term. The intersection of these two kinds of nodes is defined as the t-score of the T-test between the bad-outcome group and the good-outcome group. The network was visualized by Cytoscape 2.8.0 and the topology analysis was conducted by the Network Analyzer plugin for Cytoscape [44].

Acknowledgements

This work was supported by National Basic Research Program of China (973 Program, No2011CB711000), the National Natural Science Foundation of China to J.X. (30600759) and J.L. (60773010 and 60970063), the Ph.D. Programs Foundation of Ministry of Education of China (20090141110026) and the Fundamental Research Funds for the Central Universities (6081007) to J.L, grant from State Key Lab of Space Medicine Fundamentals and Application (SMFA) to J.X (SMFA09A07, SMFA10A03), and Advanced Space Medico-Engineering Research Project of China (01111014, 01111016). The study was carried out in SMFA.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

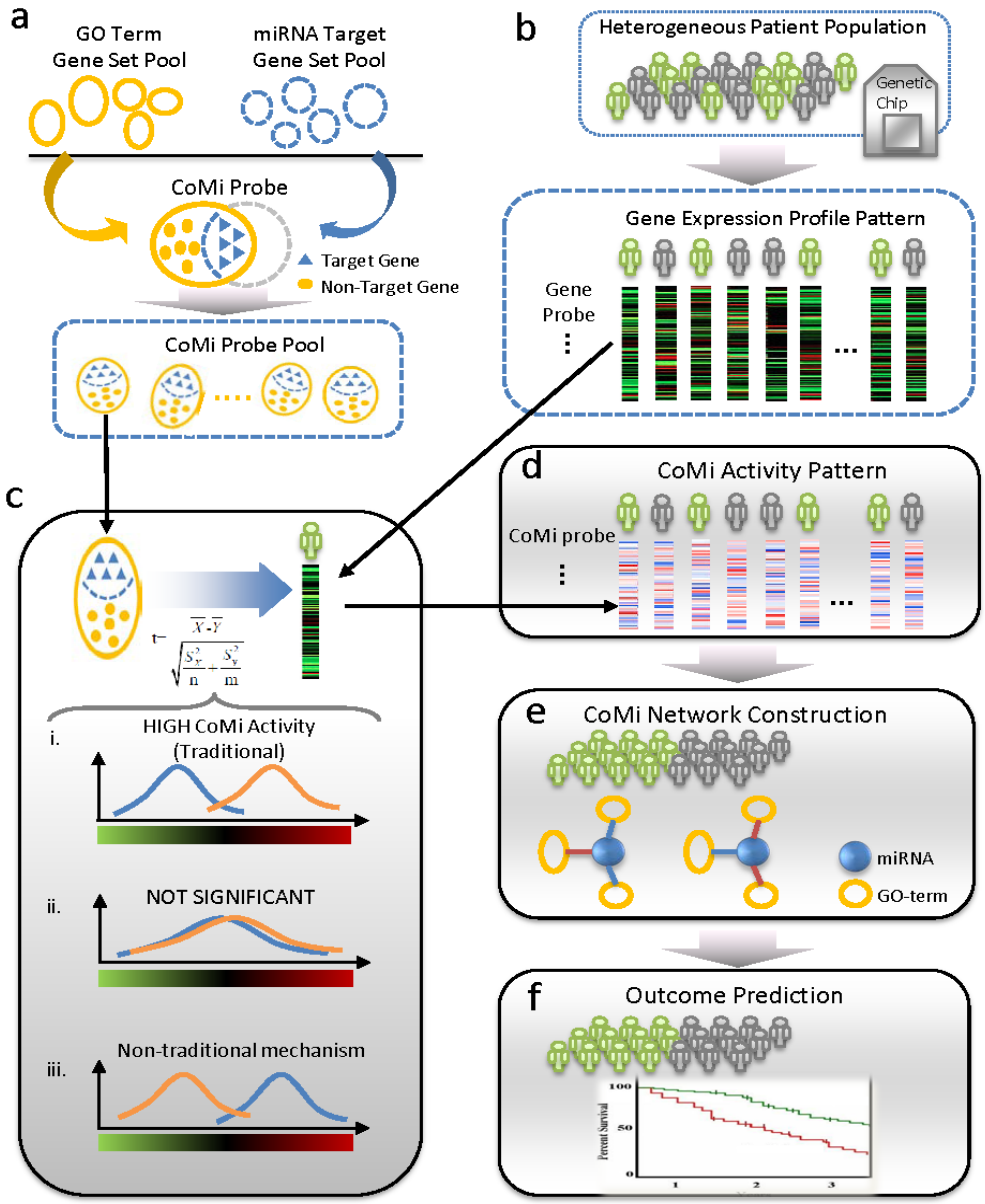

Figure 1. Pipeline of the CoMi activity estimation and application of the metric for outcome prediction; a. Identifying the significant Context specific miRNA probe(miRNA activity on a specific GO term); b. mRNA expression profiles of patient population; c. Based on CoMi probe and mRNA expression profiles, using T-test to calculate CoMi activity; d. The CoMi activity patterns; e. CoMi Network construction; f. Outcome prediction

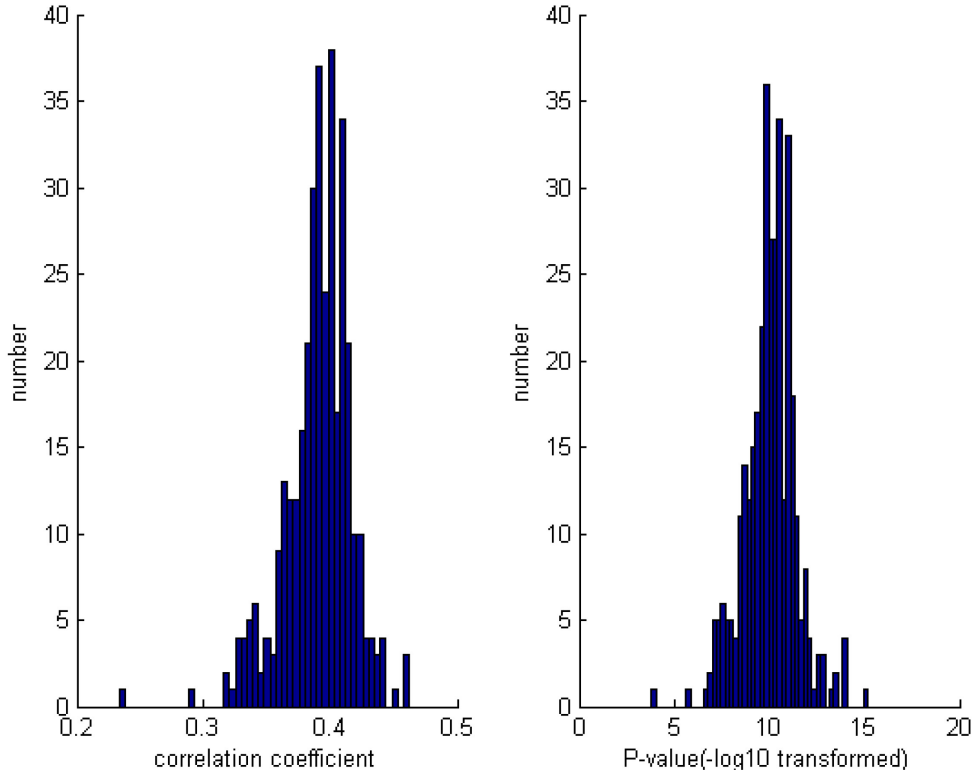

Figure 2. Distribution of Spearman correlation coefficients calculated for the correlation between the 356 miRNA CoMi activity pattern (GOBP) and miRNA expression profiles pairs (The miRNA target sets is predicted by targetscan U RNA22). a (left). Spearman correlation coefficient. b (right). P-value distribution. P-value was -log10 transformed.

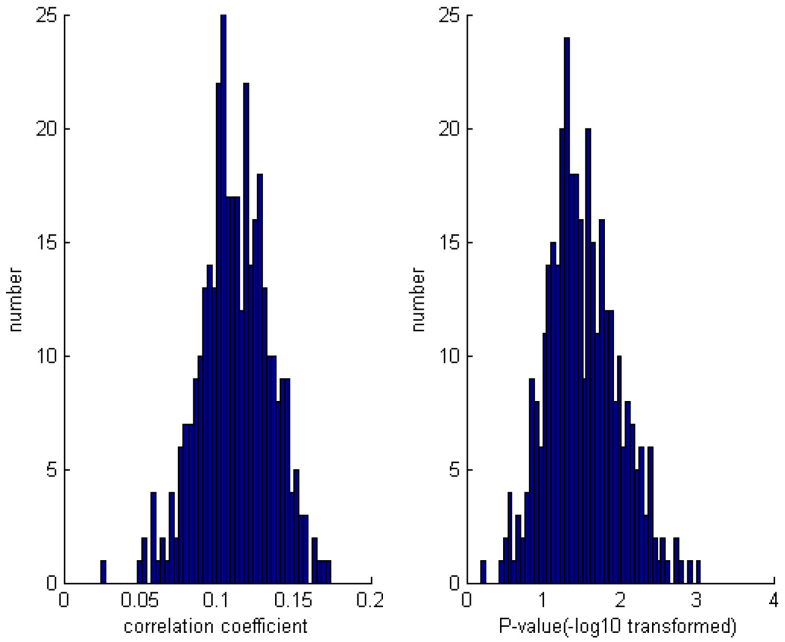

Figure 3. Distribution of Spearman correlation coefficients calculated for the correlation between the 356 miRNA CoMi activity pattern (GOBP) and miRNA expression profiles pairs (The miRNA target sets is predicted by Exprtarget). a (left). Spearman correlation coefficient. b (right). P-value distribution. P-value was -log10 transformed.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

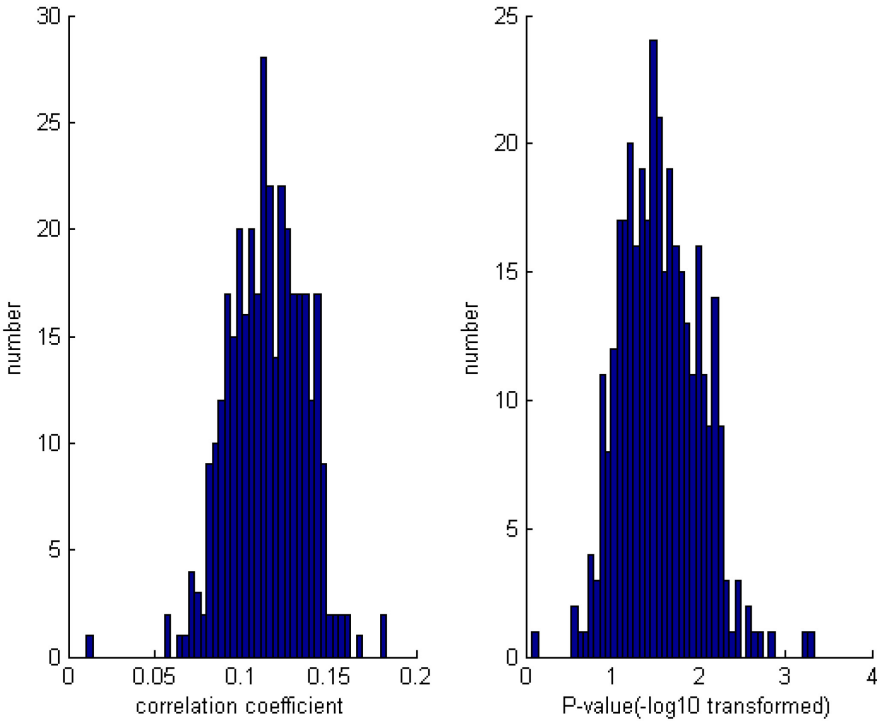

Figure 4. Distribution of Spearman correlation coefficients calculated for the correlation between the 356 miRNA CoMi activity pattern (GOBP) and miRNA expression profiles pairs (The miRNA target sets is predicted by miRanda). a (left). Spearman correlation coefficient. b (right). P-value distribution. P-value was -log10 transformed.

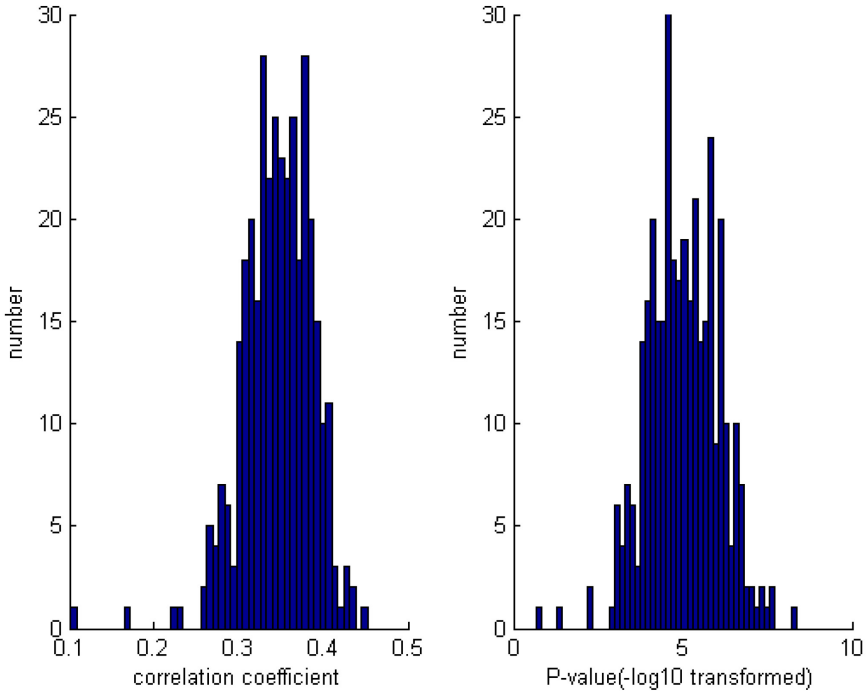

Figure 5. Distribution of Spearman correlation coefficients calculated for the correlation between the 356 miRNA CoMi activity pattern (GOBP) and miRNA expression profiles pairs (The miRNA target sets is predicted by targetscan). a (left). Spearman correlation coefficient. b (right). P-value distribution. P-value was -log10 transformed.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

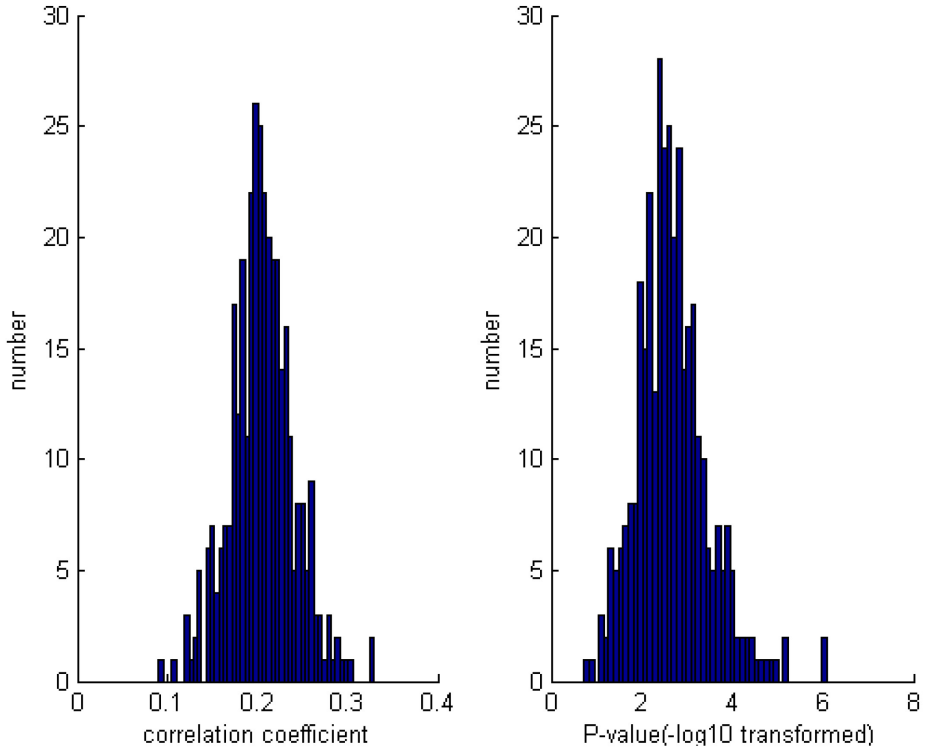

Figure 6. Distribution of Spearman correlation coefficients calculated for the correlation between the 356 miRNA CoMi activity pattern (GOBP) and miRNA expression profiles pairs (The miRNA target sets is predicted by RNA22). a (left). Spearman correlation coefficient. b (right). P-value distribution. P-value was -log10 transformed.

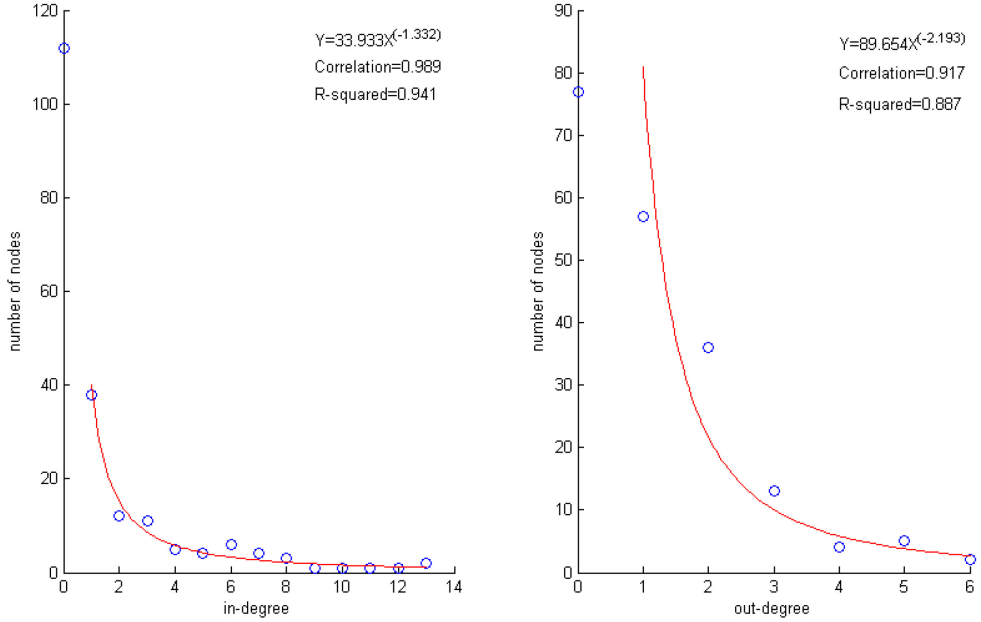

Figure 7. Topology analysis of estimated CoMi network. a (left). Power law fit to the in-degree distribution. b (right). Power law fit to the out-degree distribution.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

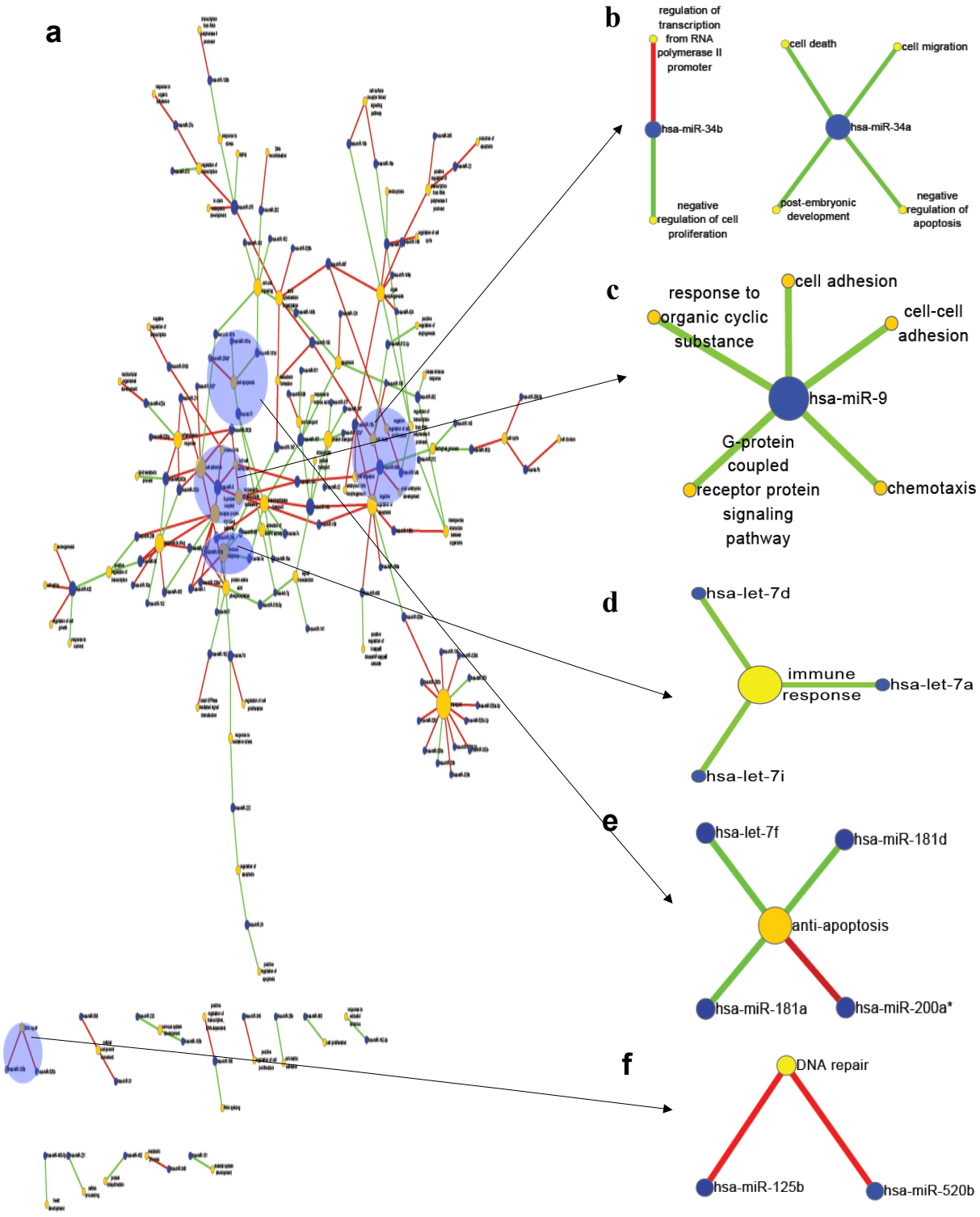

Figure 8 The CoMi network (CoMiNet): a. The CoMi activity network (CoMiNet) constructed on Wang's data; b. sub-network of miR-34a, miR-34b; c. sub-network of miR-9; d. synergistic function of let-7 family; e. miRNAs synergistic network on "anti-apoptosis"; f. miRNAs synergistic network on "DNA repair".

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

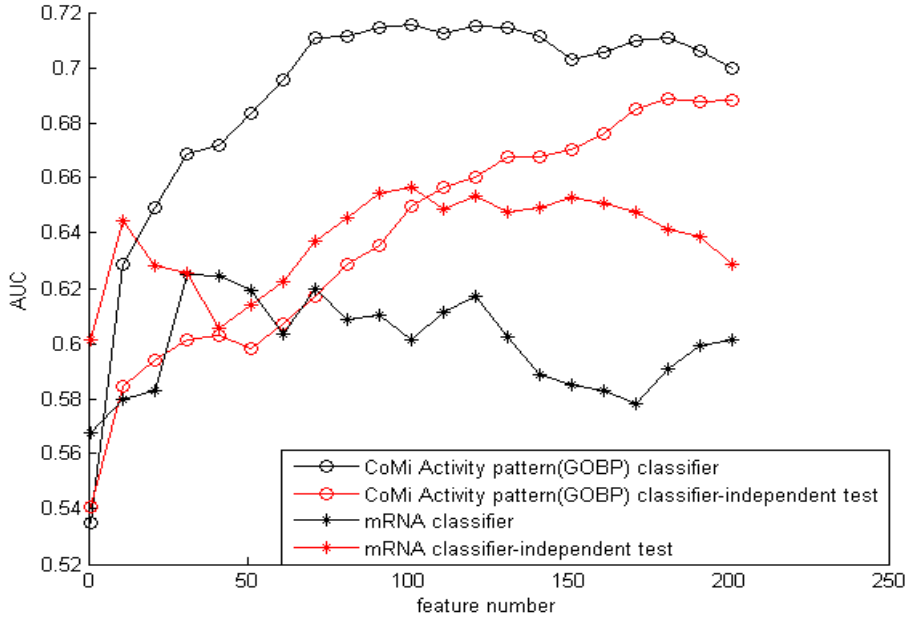

Figure 9. Relationship between the performance and the number of the selected features. The AUC of the classification as the selected feature from 1 to 201. The black line with small circles denotes the AUC of the CoMi activity pattern (GOBP) classifier on GSE2034, the red line with small circles denotes the AUC of the CoMi activity pattern classifier on GSE7390, the black line with stars denotes the AUC of the mRNA expression pattern classifier on GSE2034,and the red line with stars denotes the AUC of mRNA expression pattern classifier on GSE7390.

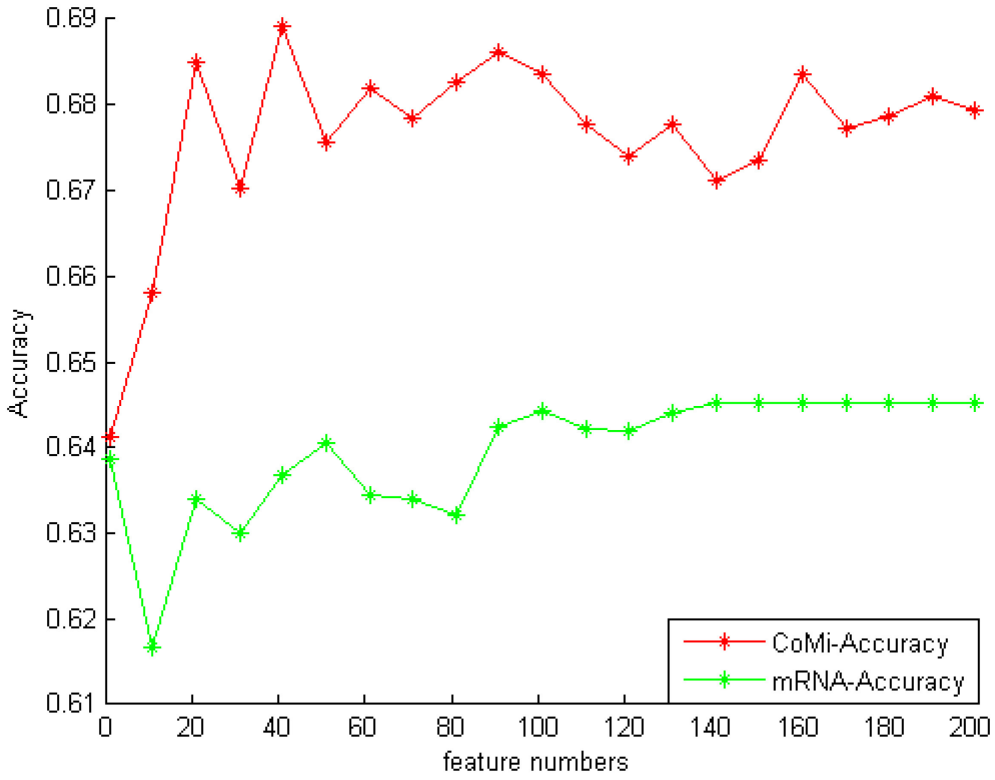

Figure 10. The accuracy of the CoMi(GOBP) classifier and mRNA classifier on GSE2034.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

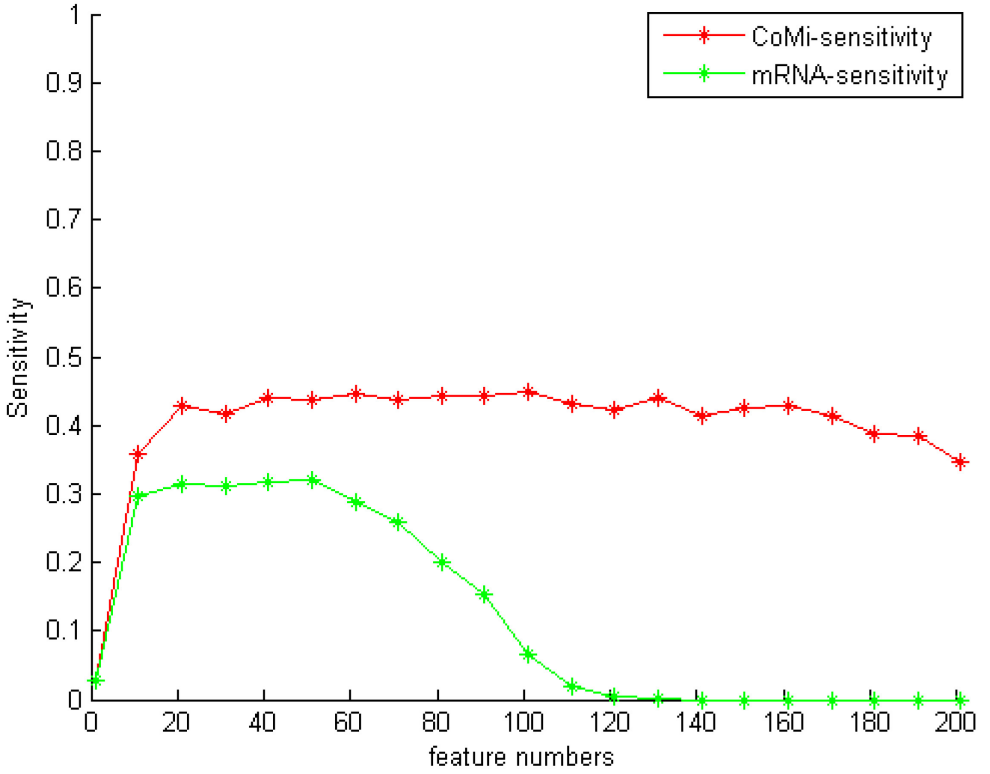

Figure 11. The sensitivity of the CoMi (GOBP) classifier and mRNA classifier on GSE2034.

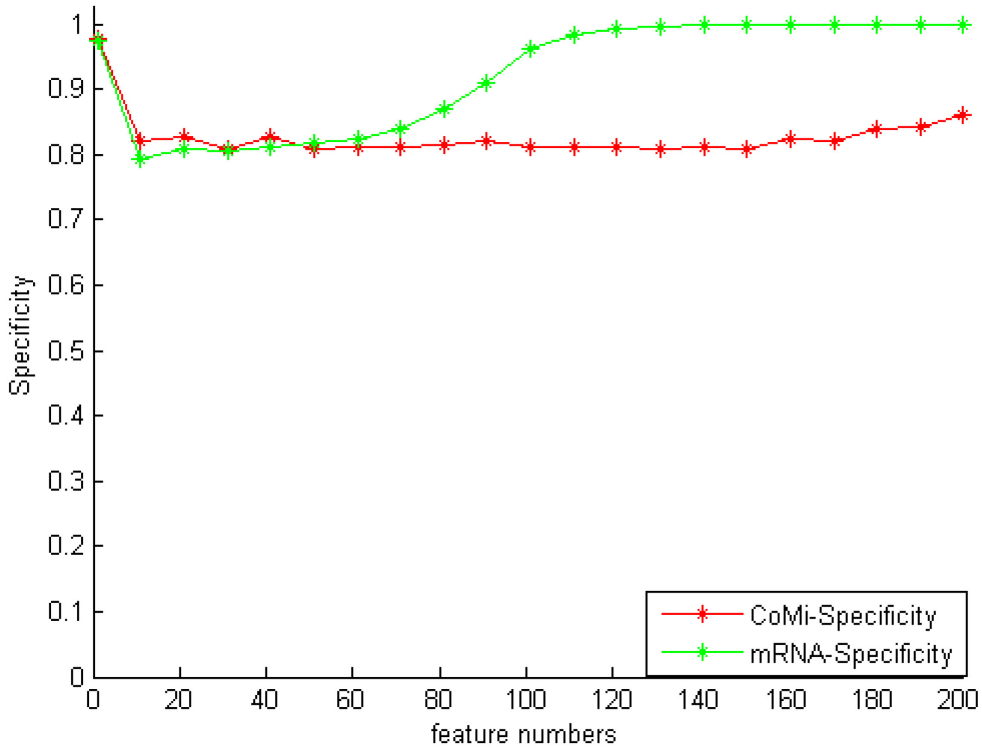

Figure 12. The specificity of the CoMi (GOBP) classifier and mRNA classifier on GSE2034.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

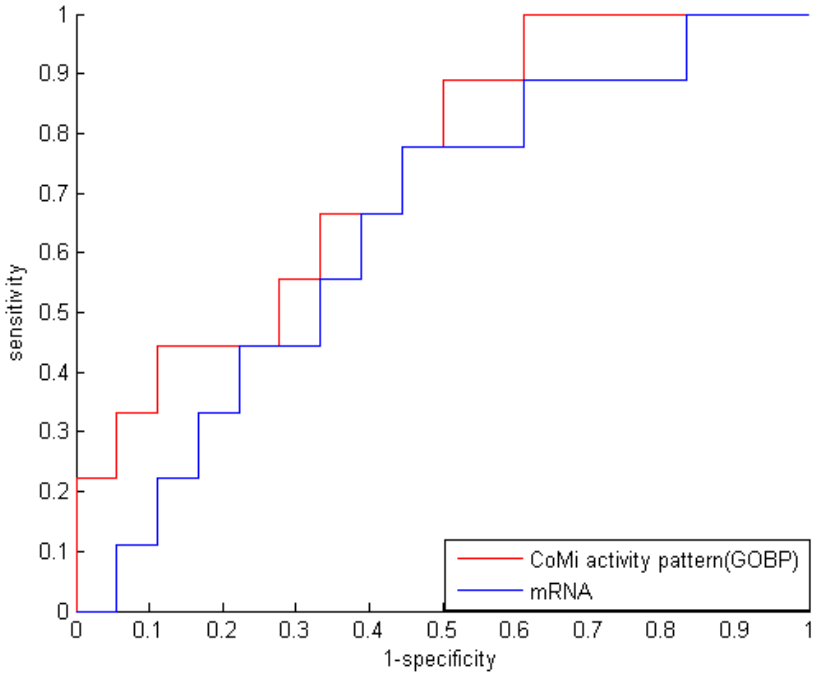

Figure 13. ROC curve of of the CoMi activity pattern classifier (red line - best performance with the number of selected features =71, AUC=0.7407) compared with the mRNA expression profile classifier (blue line - best performance with the number of selected features =41, AUC=0.6481).

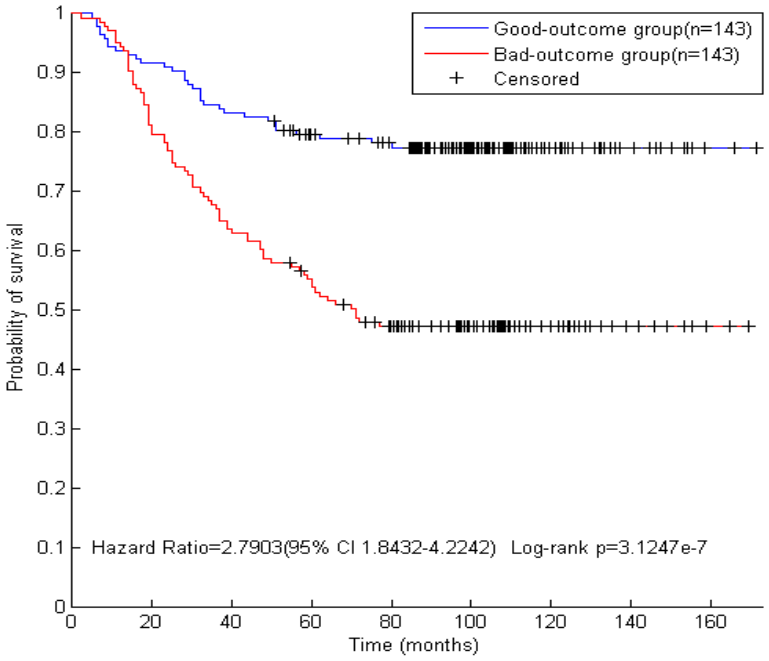

Figure 14. Survival curve of the best feature (hsa-miR-154,'response to organic cyclic sunstance') of CoMi activity pattern on GSE2034 (log-rank test) estimated by Kaplan-Meier function.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

Table 1. Sample sizes and partition by class

Data set	samples	good-outcome group (dmfs_time>=5 years ,dmfs_e=0)	bad-outcome group (dmfs_time<5 years, dmfs_e=1)	removed samples
GSE2034	286	169	93	24
GSE7390	198	128	36	34

Classification of data sets into good and bad outcome groups according to whether the patient developed distant-metastasis in 5 years. Observations (samples) were removed if they were censored before the 5-year cutoff (To removed the samples whose risk is between the low risk group and high risk group, the samples which developed distant-metastasis after 5 years also be eliminated)

Table2. miRNA target predictioon tool used in this analysis

Tool	Version	Web site
miRanda	The miRBase Targets Release Version v5	http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/
Exprtarget	Only one version	http://www.scandb.org/apps/microrna/download.html
TargetScan	Release 5.1 April 2009 (Conserved site context scores)	http://www.targetscan.org/cgi-bin/targetscan/data_download.cgi?db=vert_50
RNA22	Only one version(Human 3'UTR)	http://cbcsrv.watson.ibm.com/rna22_download_content.html

Table 3.The intersection of selected features from two different breast cacner data sets

Threshold(P-value)	Feature numbers	The number of selected feature in GSE2034	The number of selected feature in GSE7390	Size of intersection set*	P-value of intersection
0.01	6314	321	296	22	0.0252
0.02	6314	489	486	61	3.64e-05
0.03	6314	623	627	94	5.74e-06
0.04	6314	741	770	124	3.81e-05
0.05	6314	856	875	153	1.38e-04

*The features in the intersection set are significant (P-value > threshold) in both data sets and have the same direction(the same sign of the t score)

Table 4. The best 200 CoMi features that differently expressed in bad-outcome group vs good-outcome group

miRNA	Go term	P-value	t score
hsa-miR-432	cell aging	3.96121E-06	4.762315817
hsa-miR-187	apoptosis	4.18059E-06	-4.746074834
hsa-miR-15b	cell migration	9.08992E-06	4.57896588
hsa-miR-497	organ morphogenesis	1.22E-05	4.49366425
hsa-miR-154	response to organic cyclic substance	1.44032E-05	4.464775593
hsa-miR-181d	cell-cell signaling	1.50584E-05	-4.449050842
hsa-miR-133b	G-protein coupled receptor protein signaling pathway	2.11299E-05	4.366289113
hsa-miR-204	cell adhesion	2.4762E-05	4.311891737
hsa-miR-497	actin cytoskeleton organization	2.60473E-05	4.298199558
hsa-miR-15a	organ morphogenesis	3.49121E-05	4.248848264
hsa-miR-101	skeletal system development	3.65766E-05	-4.231326265
hsa-miR-375	in utero embryonic development	4.61508E-05	4.15516298
hsa-miR-138	transport	5.18851E-05	4.149373799
hsa-miR-302b	cell adhesion	6.88501E-05	4.080600137
hsa-miR-103	actin cytoskeleton organization	7.12692E-05	4.067331226
hsa-let-7d	immune response	7.92716E-05	-4.033959218
hsa-miR-149	positive regulation of transcription from RNA polymerase II promoter	8.18714E-05	4.012338633
hsa-miR-143	ion transport	0.000115154	-3.936946308
hsa-miR-144	response to organic cyclic substance	0.00012416	3.927697432
hsa-miR-107	actin cytoskeleton organization	0.000127336	3.907411986
hsa-miR-211	cell adhesion	0.000127949	3.90602206

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

hsa-miR-202	DNA recombination	0.000129739	3.901911748
hsa-miR-19a	cell surface receptor linked signaling pathway	0.000131115	3.906434358
hsa-miR-98	immune response	0.000133631	-3.895075481
hsa-miR-133b	protein amino acid phosphorylation	0.000134188	3.892785555
hsa-miR-222	regulation of apoptosis	0.000147913	-3.897221023
hsa-miR-375	response to stress	0.000157977	-3.850633731
hsa-miR-302d	cell adhesion	0.000169298	3.844606977
hsa-miR-493	negative regulation of apoptosis	0.000170211	3.84946958
hsa-miR-133a	G-protein coupled receptor protein signaling pathway	0.000175994	3.830000694
hsa-miR-212	interspecies interaction between organisms	0.000179558	-3.821836537
hsa-miR-511	ion transport	0.000203065	-3.7823137
hsa-miR-299-5p	cell cycle	0.000224462	3.749167803
hsa-miR-15b	organ morphogenesis	0.000231334	3.746314232
hsa-miR-302a	cell adhesion	0.000251611	3.743857762
hsa-miR-206	positive regulation of transcription	0.000263273	-3.711831094
hsa-miR-22	positive regulation of transcription from RNA polymerase II promoter	0.000277215	3.701994245
hsa-miR-146a	organ morphogenesis	0.000289134	3.691977505
hsa-miR-181b	cell-cell signaling	0.000300529	-3.684966333
hsa-miR-375	actin cytoskeleton organization	0.000349281	3.635290055
hsa-miR-377	protein transport	0.000373131	-3.627064197
hsa-miR-34a	cell migration	0.000378981	3.618496288
hsa-miR-512-3p	cell death	0.000390935	-3.615871201
hsa-miR-520e	negative regulation of apoptosis	0.000400686	-3.620369754
hsa-miR-520g	inflammatory response	0.000493814	3.552972641
hsa-miR-519d	negative regulation of transcription	0.000501294	3.539718453
hsa-let-7d	regulation of cell proliferation	0.000526424	3.525775804
hsa-let-7a	transmembrane transport	0.000535652	-3.525007556
hsa-miR-200a*	anti-apoptosis	0.000550266	3.507315067
hsa-miR-198	RNA splicing	0.000560855	-3.515167207
hsa-miR-424	organ morphogenesis	0.000570955	3.50006842
hsa-miR-27a	regulation of transcription	0.000596311	3.495965898
hsa-let-7i	immune response	0.000605536	-3.488048689
hsa-miR-182	small GTPase mediated signal transduction	0.000612113	3.481036307
hsa-miR-196b	negative regulation of apoptosis	0.000615729	3.486965878
hsa-miR-370	transport	0.000617809	-3.482294808
hsa-miR-382	apoptosis	0.000643086	-3.478074755
hsa-miR-183	cell-cell signaling	0.000679048	-3.459143578
hsa-miR-432	positive regulation of transcription	0.000737217	-3.427225945
hsa-miR-302b	inflammatory response	0.000754465	3.425771536
hsa-miR-141	signal transduction	0.000760681	-3.423338566
hsa-miR-433	response to drug	0.00076083	3.425064904
hsa-miR-34b	regulation of transcription from RNA polymerase II promoter	0.000779479	-3.407722605
hsa-miR-181a	anti-apoptosis	0.000799558	-3.412627372
hsa-miR-143	protein transport	0.000799597	-3.409024582
hsa-let-7f	activation of MAPK activity	0.000867984	-3.384925615
hsa-miR-302d	inflammatory response	0.000868699	3.387467362
hsa-miR-124	apoptosis	0.000875631	3.38330257
hsa-miR-520b	DNA repair	0.00088007	3.381351742
hsa-miR-15b	embryonic limb morphogenesis	0.000970328	3.352610761
hsa-miR-34b	negative regulation of cell proliferation	0.00098402	3.345228935

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

hsa-miR-515-5p	protein amino acid phosphorylation	0.000985309	-3.355247345
hsa-miR-152	cell-cell signaling	0.00098911	-3.349946752
hsa-miR-198	positive regulation of transcription, DNA-dependent	0.001003618	3.338447442
hsa-miR-223	nervous system development	0.0010044	-3.339353662
hsa-miR-526b	actin cytoskeleton organization	0.001007173	3.335109149
hsa-miR-302a	response to drug	0.001015467	-3.339937369
hsa-miR-9	cell-cell adhesion	0.001040423	3.330263084
hsa-miR-340	metabolic process	0.00105367	3.327570357
hsa-let-7b	cell cycle	0.001064662	3.316792421
hsa-miR-432	axonogenesis	0.001105321	3.31339562
hsa-miR-135b	nervous system development	0.001124916	-3.308330605
hsa-miR-432	regulation of cell growth	0.001153692	3.302767971
hsa-miR-9	cell adhesion	0.001160895	3.306759688
hsa-miR-133a	protein amino acid phosphorylation	0.001163682	3.292698852
hsa-miR-206	G-protein coupled receptor protein signaling pathway	0.001187538	3.295655291
hsa-miR-148b	apoptosis	0.001212391	-3.282897736
hsa-miR-24	regulation of apoptosis	0.001228743	-3.284531873
hsa-miR-489	immune response	0.00124972	-3.284741056
hsa-miR-452	protein ubiquitination	0.001251844	-3.272803915
hsa-miR-506	ion transport	0.001254199	3.272888822
hsa-miR-144	cell migration	0.001281527	3.27234854
hsa-miR-202	cell-cell signaling	0.001320308	-3.261152549
hsa-miR-196a	negative regulation of apoptosis	0.001370557	3.26026755
hsa-miR-143	signal transduction	0.001390137	-3.24556137
hsa-miR-493	positive regulation of I-kappaB kinase/NF-kappaB cascade	0.001437211	-3.249426565
hsa-miR-132	response to drug	0.00149568	-3.230106205
hsa-miR-145	biological_process	0.00150329	-3.225658933
hsa-miR-433	G-protein coupled receptor protein signaling pathway	0.001530499	-3.212746967
hsa-miR-19a	regulation of transcription from RNA polymerase II promoter	0.001619631	-3.205718008
hsa-miR-9	chemotaxis	0.001660713	3.189904545
hsa-miR-211	inflammatory response	0.001672787	3.189400073
hsa-miR-147	protein transport	0.001739162	-3.180398557
hsa-miR-19b	cell surface receptor linked signaling pathway	0.00177761	3.171184824
hsa-miR-1	immune response	0.001792519	3.1726194
hsa-miR-27a	response to organic substance	0.001815341	3.165218075
hsa-miR-93	G-protein coupled receptor protein signaling pathway	0.001863766	3.156962307
hsa-miR-204	transmembrane transport	0.001871599	3.149333592
hsa-miR-93	response to drug	0.001936376	3.151391263
hsa-miR-10a	response to drug	0.001986191	3.132076816
hsa-miR-196b	interspecies interaction between organisms	0.002031538	3.12541937
hsa-miR-15b	negative regulation of apoptosis	0.002087347	3.127229034
hsa-miR-204	protein amino acid phosphorylation	0.002138932	3.114046496
hsa-miR-489	response to drug	0.002140108	3.109749864
hsa-miR-302d	response to organic cyclic substance	0.002177862	3.113003845
hsa-let-7g	activation of MAPK activity	0.002187026	-3.108154008
hsa-miR-181d	anti-apoptosis	0.002221853	-3.1068147
hsa-let-7f	transmembrane transport	0.002225288	-3.099225717
hsa-miR-144	cell-cell adhesion	0.002260832	3.094042119
hsa-miR-103	mesoderm formation	0.002296366	3.08274055
hsa-miR-107	mesoderm formation	0.002296366	3.08274055

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

hsa-miR-9	response to organic cyclic substance	0.002316435	3.092881855
hsa-miR-107	immune response	0.002349499	-3.085078314
hsa-miR-221	mRNA processing	0.002357038	-3.086359489
hsa-let-7c	transmembrane transport	0.002464806	-3.07418121
hsa-miR-103	cell death	0.002474554	3.064847242
hsa-miR-98	activation of MAPK activity	0.002491039	-3.060401263
hsa-miR-515-5p	signal transduction	0.00251662	-3.062056862
hsa-let-7f	anti-apoptosis	0.002536623	-3.057031775
hsa-miR-143	intracellular protein transport	0.002540391	-3.059572052
hsa-miR-148b	cell-cell signaling	0.002558506	-3.058209541
hsa-miR-34a	biological_process	0.00256168	-3.058218359
hsa-let-7a	immune response	0.002581689	-3.05338749
hsa-miR-200b	transport	0.002599377	3.051031067
hsa-miR-19b	regulation of transcription from RNA polymerase II promoter	0.002688134	-3.0474327
hsa-miR-149	organ morphogenesis	0.002714049	3.038122784
hsa-miR-34a	negative regulation of apoptosis	0.002751504	3.04050921
hsa-miR-302a	inflammatory response	0.002873499	3.023544211
hsa-miR-485-5p	heart development	0.00291466	-3.015910391
hsa-miR-142-3p	response to estradiol stimulus	0.002921403	-3.016160615
hsa-miR-31	cellular component movement	0.00292727	3.012720474
hsa-miR-1	G-protein coupled receptor protein signaling pathway	0.00293732	3.014927842
hsa-miR-375	aging	0.003002168	-3.005061062
hsa-miR-382	innate immune response	0.003100239	-3.004290556
hsa-let-7d	response to oxidative stress	0.003100715	-2.99509693
hsa-miR-503	cellular component movement	0.003195247	2.986677277
hsa-miR-373	regulation of transcription	0.003281398	-2.974761621
hsa-miR-34a	cell death	0.003303906	2.971671032
hsa-miR-9	G-protein coupled receptor protein signaling pathway	0.003337516	2.973163251
hsa-miR-143	response to organic cyclic substance	0.003346928	2.974353333
hsa-miR-497	negative regulation of cell proliferation	0.003351688	2.969920402
hsa-miR-125b	DNA repair	0.003358748	2.967268946
hsa-let-7b	cell division	0.003397989	2.960762131
hsa-miR-363	cell proliferation	0.003409899	-2.969851575
hsa-miR-20b	cell-matrix adhesion	0.003457245	-2.963136573
hsa-miR-382	interspecies interaction between organisms	0.003462398	-2.965100592
hsa-miR-93	positive regulation of transcription	0.003487585	-2.954319108
hsa-let-7c	activation of MAPK activity	0.003499342	-2.957362351
hsa-miR-187	protein transport	0.003502399	-2.955489683
hsa-miR-34a	post-embryonic development	0.00354118	2.951935969
hsa-miR-302c*	protein transport	0.003564189	-2.955671788
hsa-miR-345	positive regulation of transcription from RNA polymerase II promoter	0.003662921	2.945384009
hsa-miR-195	cell death	0.003752201	2.929342993
hsa-miR-22	induction of apoptosis	0.003758528	2.932942574
hsa-miR-182	protein amino acid phosphorylation	0.003788682	-2.931209383
hsa-miR-222	response to oxidative stress	0.003803201	-2.93003521
hsa-miR-375	regulation of transcription	0.003939172	2.919736316
hsa-miR-181b	cell adhesion	0.003961065	-2.919618381
hsa-miR-432	response to nutrient	0.003983184	-2.915563848
hsa-miR-302c	biological_process	0.004009293	-2.910535268
hsa-miR-302a	lipid metabolic process	0.004043111	-2.913755816

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

hsa-miR-18a	transmembrane transport	0.00417635	-2.901349707
hsa-miR-512-3p	positive regulation of angiogenesis	0.004264736	-2.889593667
hsa-miR-15a	regulation of cell cycle	0.004316509	2.888290992
hsa-miR-302e	transport	0.004320468	2.89023609
hsa-miR-520a-3p	transport	0.004320468	2.89023609
hsa-miR-520c-3p	transport	0.004320468	2.89023609
hsa-miR-520d-3p	transport	0.004320468	2.89023609
hsa-miR-422a	multicellular organismal development	0.00432108	2.887763596
hsa-miR-187	transmembrane transport	0.004352332	-2.894732959
hsa-miR-15a	endocytosis	0.004391418	-2.882772793
hsa-miR-32	protein transport	0.004391439	-2.885872789
hsa-miR-302c	cell cycle	0.004403445	2.878597463
hsa-miR-520e	transport	0.00444894	2.877135639
hsa-miR-302b	response to drug	0.004465935	-2.877754373
hsa-miR-346	positive regulation of cell proliferation	0.004567523	2.870476427
hsa-miR-320a	transport	0.004591541	2.866994602
hsa-miR-320b	transport	0.004591541	2.866994602
hsa-miR-320c	transport	0.004591541	2.866994602
hsa-miR-320d	transport	0.004591541	2.866994602
hsa-miR-182*	cell adhesion	0.004596821	2.868921008
hsa-miR-23b	transport	0.00469032	-2.860601372
hsa-miR-519d	chemotaxis	0.004691776	2.860127843
hsa-miR-422a	inflammatory response	0.004705513	-2.859992668
hsa-miR-130b	response to stress	0.004723511	-2.861735596
hsa-miR-187	response to retinoic acid	0.004776236	-2.852779437
hsa-miR-154	negative regulation of apoptosis	0.004779391	2.859643707
hsa-miR-24	positive regulation of apoptosis	0.004880341	-2.850189629
hsa-miR-143	negative regulation of apoptosis	0.00489304	2.844460791
hsa-miR-130b	transcription from RNA polymerase II promoter	0.004896513	2.847280535

P-value and T score was calculated by T-test(matlab function:mattest) between the bad outcome group and good outcome group.

Table 5. The number of good features filter by FDR of Cox proportional hazards regression

Feature space	FDR<0.001	FDR<0.005	FDR<0.01
CoMi activity pattern	1	7	16
mRNA expression pattern	0	0	5

Table 6. The target of miR-154 on Go term 'response to organic cyclic substance'

Gene Id	Gene Symbol	Gene Name
551	AVP	arginine vasopressin
894	CCND2	cyclin D2
898	CCNE1	cyclin E1
960	CD44	CD44 molecule (Indian blood group)
3673	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)
5021	OXTR	oxytocin receptor
5027	P2RX7	purinergic receptor P2X, ligand-gated ion channel, 7
5037	PEBP1	phosphatidylethanolamine binding protein 1
5727	PTCH1	patched 1
6774	STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

References

[1] D. Baek, *et al.*, "The impact of microRNAs on protein output," *Nature,* vol. 455, pp. 64-71, Sep 4 2008.

[2] A. Grimson, *et al.*, "MicroRNA targeting specificity in mammals: determinants beyond seed pairing," *Mol Cell,* vol. 27, pp. 91-105, Jul 6 2007.

[3] S. Volinia, *et al.*, "A microRNA expression signature of human solid tumors defines cancer gene targets," *Proc Natl Acad Sci U S A,* vol. 103, pp. 2257-61, Feb 14 2006.

[4] C. J. Marsit, *et al.*, "MicroRNA responses to cellular stress," *Cancer Res,* vol. 66, pp. 10843-8, Nov 15 2006.

[5] C. Z. Chen, *et al.*, "MicroRNAs modulate hematopoietic lineage differentiation," *Science,* vol. 303, pp. 83-6, Jan 2 2004.

[6] E. Enerly, *et al.*, "miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors," *PLoS One,* vol. 6, p. e16915, 2011.

[7] T. D. Schmittgen, "Regulation of microRNA processing in development, differentiation and cancer," *J Cell Mol Med,* vol. 12, pp. 1811-9, Oct 2008.

[8] E. van Rooij, *et al.*, "A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure," *Proc Natl Acad Sci U S A,* vol. 103, pp. 18255-60, Nov 28 2006.

[9] N. Raver-Shapira, *et al.*, "Transcriptional activation of miR-34a contributes to p53-mediated apoptosis," *Mol Cell,* vol. 26, pp. 731-43, Jun 8 2007.

[10] S. Sassen, *et al.*, "MicroRNA: implications for cancer," *Virchows Arch,* vol. 452, pp. 1-10, Jan 2008.

[11] C. Cheng and L. M. Li, "Inferring microRNA activities by combining gene expression with microRNA target prediction," *PLoS One,* vol. 3, p. e1989, 2008.

[12] S. Volinia, *et al.*, "Identification of microRNA activity by Targets' Reverse EXpression," *Bioinformatics,* vol. 26, pp. 91-7, Jan 1 2010.

[13] I. Ulitsky, *et al.*, "Towards computational prediction of microRNA function and activity," *Nucleic Acids Res,* vol. 38, p. e160, Aug 2010.

[14] J. Xiong, *et al.*, "Pre-clinical drug prioritization via prognosis-guided genetic interaction networks," *PLoS One,* vol. 5, p. e13937, 2010.

[15] J. Li, *et al.*, "Identification of high-quality cancer prognostic markers and metastasis network modules," *Nat Commun,* vol. 1, p. 34, 2010.

[16] T. Sorlie, *et al.*, "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications," *Proc Natl Acad Sci U S A,* vol. 98, pp. 10869-74, Sep 11 2001.

[17] M. J. van de Vijver, *et al.*, "A gene-expression signature as a predictor of survival in breast cancer," *N Engl J Med,* vol. 347, pp. 1999-2009, Dec 19 2002.

[18] Y. Wang, *et al.*, "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *Lancet,* vol. 365, pp. 671-9, Feb 19-25 2005.

[19] H. Y. Chuang, *et al.*, "Network-based classification of breast cancer metastasis," *Mol Syst Biol,* vol. 3, p. 140, 2007.

[20] G. Abraham, *et al.*, "Prediction of breast cancer prognosis using gene set statistics provides signature stability and biological context," *BMC Bioinformatics,* vol. 11, p. 277, 2010.

[21] M. H. van Vliet, *et al.*, "Module-based outcome prediction using breast cancer compendia," *PLoS One,* vol. 2, p. e1047, 2007.

[22] I. W. Taylor, *et al.*, "Dynamic modularity in protein interaction networks predicts breast cancer outcome," *Nat Biotechnol,* vol. 27, pp. 199-204, Feb 2009.

[23] M. Ashburner, *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet,* vol. 25, pp. 25-9, May 2000.

[24] B. P. Lewis, *et al.*, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell,* vol. 120, pp. 15-20, Jan 14 2005.

[25] R. C. Friedman, *et al.*, "Most mammalian mRNAs are conserved targets of microRNAs," *Genome Res,* vol. 19, pp. 92-105, Jan 2009.

[26] A. J. Enright, *et al.*, "MicroRNA targets in Drosophila," *Genome Biol,* vol. 5, p. R1, 2003.

[27] S. Griffiths-Jones, *et al.*, "miRBase: tools for microRNA genomics," *Nucleic Acids Res,* vol. 36, pp. D154-8, Jan 2008.

[28] E. R. Gamazon, *et al.*, "Exprtarget: an integrative approach to predicting human microRNA targets," *PLoS One,* vol. 5, p. e13534, 2010.

[29] K. C. Miranda, *et al.*, "A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes," *Cell,* vol. 126, pp. 1203-17, Sep 22 2006.

[30] M. E. Peter, "Targeting of mRNAs by multiple miRNAs: the next step," *Oncogene,* vol. 29, pp. 2161-4, Apr 15 2010.

[31] C. Desmedt, *et al.*, "Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series," *Clin Cancer Res,* vol. 13, pp. 3207-14, Jun 1 2007.

[32] L. He, *et al.*, "A microRNA component of the p53 tumour suppressor network," *Nature,* vol. 447, pp. 1130-4, Jun 28 2007.

[33] Y. Li, *et al.*, "Regulation of microRNAs by natural agents: an emerging field in chemoprevention and chemotherapy research," *Pharm Res,* vol. 27, pp. 1027-41, Jun 2010.

[34] M. S. Nicoloso, *et al.*, "MicroRNAs--the micro steering wheel of tumour metastases," *Nat Rev Cancer,* vol. 9, pp. 293-302, Apr 2009.

[35] N. Hu, *et al.*, "miR-520b regulates migration of breast cancer cells by targeting hepatitis B X-interacting protein and interleukin-8," *J Biol Chem,* vol. 286, pp. 13714-22, Apr 15 2011.

[36] H. Oka, *et al.*, "Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis," *Cancer Res,* vol. 53, pp. 1696-701, Apr 1 1993.

[37] S. Krajewski, *et al.*, "Prognostic significance of apoptosis regulators in breast cancer," *Endocr Relat Cancer,* vol. 6, pp. 29-40, Mar 1999.

[38] J. Xu, *et al.*, "MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features," *Nucleic Acids Res,* vol. 39, pp. 825-36, Feb 2011.

[39] D. M. Dykxhoorn, *et al.*, "miR-200 enhances mouse breast cancer cell colonization to form distant metastases," *PLoS One,* vol. 4, p. e7181, 2009.

[40] M. V. Iorio, *et al.*, "MicroRNA gene expression deregulation in human breast cancer," *Cancer Res,* vol. 65, pp. 7065-70, Aug 15 2005.

[41] E. L. Goode, *et al.*, "Polymorphisms in DNA repair genes and associations with cancer risk," *Cancer Epidemiol Biomarkers Prev,* vol. 11, pp. 1513-30, Dec 2002.

[42] e. a. A. J. Lowery, "Micro-RNA expression profiling in primary breast tumours," *European Journal of Cancer,* vol. 5, p. 1, 2007.

[43] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011

[44] Y. Assenov, *et al.*, "Computing topological parameters of biological networks," *Bioinformatics,* vol. 24, pp. 282-4, Jan 15 2008.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE