BIOINFORMATICS ORIGINAL PAPER Vol. 23 no. 9 2007, pages 1141–1147
doi:10.1093/bioinformatics/btm045

Data and text mining

Discovery of microRNA–mRNA modules via population-based probabilistic learning

Je-Gun Joung[1], Kyu-Baek Hwang[2], Jin-Wu Nam[1], Soo-Jin Kim[1] and Byoung-Tak Zhang[1,3,*]

[1]Center for Bioinformation Technology, Seoul National University, Seoul 151-742, [2]School of Computing, Soongsil University, Seoul 156-743 and [3]School of Computer Science and Engineering, Seoul National University, Seoul 151-742, Korea

Received on October 30, 2006; revised on December 15, 2006; accepted on February 4, 2007

Advance Access publication March 9, 2007

Associate Editor: Satoru Miyano

ABSTRACT

Motivation: MicroRNAs (miRNAs) and mRNAs constitute an important part of gene regulatory networks, influencing diverse biological phenomena. Elucidating closely related miRNAs and mRNAs can be an essential first step towards the discovery of their combinatorial effects on different cellular states. Here, we propose a probabilistic learning method to identify synergistic miRNAs involving regulation of their condition-specific target genes (mRNAs) from multiple information sources, i.e. computationally predicted target genes of miRNAs and their respective expression profiles.

Results: We used data sets consisting of miRNA–target gene binding information and expression profiles of miRNAs and mRNAs on human cancer samples. Our method allowed us to detect functionally correlated miRNA–mRNA modules involved in specific biological processes from multiple data sources by using a balanced fitness function and efficient searching over multiple populations. The proposed algorithm found two miRNA–mRNA modules, highly correlated with respect to their expression and biological function. Moreover, the mRNAs included in the same module showed much higher correlations when the related miRNAs were highly expressed, demonstrating our method's ability for finding coherent miRNA–mRNA modules. Most members of these modules have been reported to be closely related with cancer. Consequently, our method can provide a primary source of miRNA and target sets presumed to constitute closely related parts of gene regulatory pathways.

Contact: btzhang@bi.snu.ac.kr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

MicroRNAs (miRNAs) are a class of small endogenous RNA molecules (~22 nt), which are presumed to participate in the developmental control of gene expression (Bartel *et al*., 2004). They can suppress their target genes (mRNAs) posttranscriptionally by complementary base pairing. Hence, miRNAs are related to diverse cellular processes and regarded as important components of the gene regulatory network.

Researchers have tried to elucidate the function of miRNAs in cellular processes using experimental and computational approaches (Denli *et al*., 2004; Han *et al*., 2006; Thomson *et al*., 2004). Early efforts in this area mainly focused on the identification of miRNAs and their targets (Lewis *et al*., 2005; Nam *et al*., 2006). Expression profiling techniques were also deployed for characterizing differentially expressed miRNAs according to cellular states and environmental conditions (Liu *et al*., 2004, 2005; Thomson *et al*., 2004). Correspondingly, significant amounts of data on miRNAs have now accumulated (Griffiths-Jones *et al*., 2006).

To understand the regulatory mechanism of miRNAs in complex cellular systems, it is important to identify the functional modules involved in complex interactions between miRNAs and their targets. Previously, the concept of miRNA regulatory modules (MRMs) was introduced by Yoon and De Micheli (2005). Their modules are related to only miRNA–mRNA duplexes in the sequence level without considering their expression profiles. Additional information on the expression profiles of miRNAs and mRNAs could be useful to detect the actual MRMs in specific biological processes. Recently, integrated analysis of targeting information and expression profiles was trialed to discover functional miRNA targets (Huang *et al*., 2006; Zilberstein *et al*., 2006). They reported that the utilization of expression profiles could help identify targets with high confidence.

Here we propose a population-based probabilistic method to identify coherent miRNA–mRNA modules by integrating heterogeneous information, i.e. computationally predicted target genes of miRNAs and two respective expression profiles of mRNAs and miRNAs. Here, miRNA–mRNA modules are defined as groups of miRNAs and their target mRNAs involved in similar biological processes. In our approach, a module consists of highly related miRNAs and their targets, which can be thought to have similar biological functions. Our main idea is to combine multiple information sources to extract common patterns among them, and to minimize noise and errors in each information source. Figure 1 illustrates our

*To whom correspondence should be addressed.

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

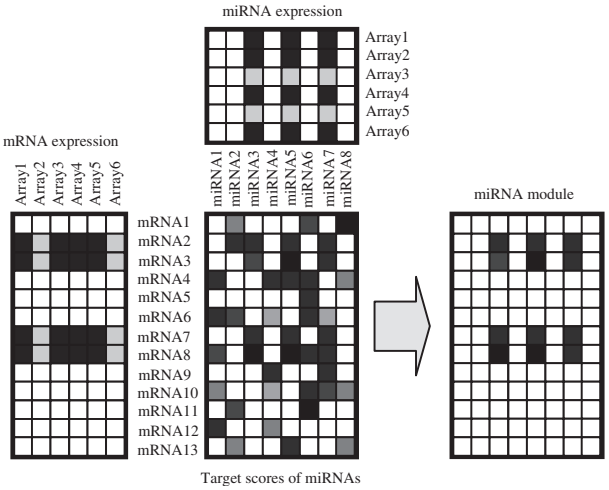

Fig. 1. Schematic view of our approach to combining multiple information sources, i.e. predicted target genes of miRNAs and two respective expression profiles of mRNAs and miRNAs, to extract coherent miRNA–mRNA modules presumed to share similar biological functions.

method of extracting coherent miRNA–mRNA modules from heterogeneous information sources.

It is not straightforward to combine heterogeneous information sources, because they have different characteristics, e.g. sizes, scales and dimensions. Our population-based probabilistic learning method is built on coevolutionary learning (Potter and De Jong, 2000) and estimation-of-distribution algorithms (EDAs) (Baluja, 1994, 1997; Larrañaga and Lozano, 2002; Zhang, 2003), addressing the problem of combining multiple data sets effectively. The fitness function is defined as a balanced aggregate of the degree of coherences between the possible pairs of miRNAs and their putative targets, i.e. miRNA–mRNA, miRNA–miRNA and mRNA–mRNA. It represents relatedness between miRNAs and mRNAs in terms of three kinds of different views. The coevolutionary learning algorithm guides the search for an optimal module by maintaining two populations consisting of miRNAs and mRNAs. EDAs are used for generating 'offspring' based on a probability model describing the current population. Our approach can identify highly correlated miRNAs and mRNAs efficiently from a high-dimensional search space through these learning strategies.

We applied our algorithm to the combined analysis of miRNA–target gene binding information and expression profiles of miRNAs and mRNAs on human cancer samples. The detected miRNA–mRNA modules were validated using correlation coefficient analysis biological functional analysis and a survey of the biomedical literature.

2 METHODS

2.1 Formulation of the problem

Our goal is to find optimal miRNA–mRNA modules based on computationally predicted miRNA–target gene information and their respective expression profiles. Let $M = \{m_1, m_2, \ldots, m_{N_m}\}$ denote the set of miRNAs and $T = \{t_1, t_2, \ldots, t_{N_t}\}$ the set of mRNAs, where N_m and N_t correspond to the number of miRNAs and mRNAs, respectively. From the total set of miRNAs and mRNAs, a subset $B = (M', T')$ can be selected as a module, where $|M'| \leq |M|$ and $|T'| \leq |T|$. The target information can be represented as an $N_m \times N_t$ scoring matrix consisting of prediction scores of target binding. The expression profiles of miRNAs (mRNAs) can also be represented as an $N_m \times N_{m_s}$ ($N_t \times N_{t_s}$) matrix, where N_{m_s} (N_{m_t}) denotes the number of microarray samples.

2.2 Definition of the fitness function

The putative target information describes the similarity between miRNAs and their target mRNAs from the viewpoint of complementary base pairing. The expression profiles of miRNAs and mRNAs provide information about their coherence in transcription across various cellular states or conditions. Taking into account these factors, the fitness of a module (M', T') can be measured as follows:

$$F(M', T') = \alpha BS_{M'T'} + \beta EC_{M'} + \gamma EC_{T'} + VOL. \quad (1)$$

Here, $BS_{M'T'}$ is the mean binding score of the subset of the target information matrix, consisting of (M', T'). $EC_{M'}$ and $EC_{T'}$ are the expression coherence (EC) scores of M' and T', respectively. We define the EC score as the mean of Pearson's correlation coefficient between all miRNAs or mRNA pairs. All the scores are normalized into $[0, 1]$. In Equation (1), α, β and γ are the parameters of the fitness function, which control the balance among the three scoring terms. In addition to the above scoring terms, a volume term is added to the fitness function to prevent finding a trivial solution consisting of one or two miRNAs and mRNAs. The volume term is given as:

$$VOL = w \cdot (w_m \cdot (N'_m/N_m) + w_t \cdot (N'_t/N_t)), \quad (2)$$

where N'_m and N'_t are the subset size, and w, w_m and w_t are control parameters similar to α, β and γ, respectively.

2.3. Description of the learning algorithm

Our algorithm for finding an optimal miRNA–mRNA module is based on coevolutionary learning and EDA approaches. Here, coevolutionary learning evolves cooperatively for two populations in the context of each other, and EDAs generate new solutions using the probability distribution. Figure 2 summarizes the probabilistic learning algorithm for miRNA–mRNA module identification. The function to optimize is defined over the binary space $\{0, 1\}^{N_m}$ for the set of miRNAs, M and $\{0, 1\}^{N_t}$ for the set of mRNAs, T. M' and T' are 1's sets in this binary space.

During the learning process, we maintain two respective populations of miRNAs and mRNAs. First, the algorithm initializes these populations by random selection. The populations X (for miRNAs) and Y (for mRNAs) can be represented as follows: $X = \{x_1, x_2, \ldots, x_\mu\}$, $Y = \{y_1, y_2, \ldots, y_\nu\}$, where μ and ν denote the respective population size. The i-th individual x_i (y_i) corresponds to a binary string representing the inclusion (1) or exclusion (0) of the miRNAs (mRNAs).

The quality of each individual x_i (y_i) in X (Y) is evaluated using Equation (1). To evaluate the fitness of an individual in a population, a corresponding individual in another population must be determined. In our approach, the individual producing the best score (Equation (1)) with any given individual is chosen. This strategy induces a coevolutionary learning effect that finds complete solutions cooperatively. Then, the best R and V ($R < \mu, V < \nu$) individuals from X and Y are selected for updating the probability distribution for each population. Specifically, the probability distribution over miRNAs (mRNAs) is

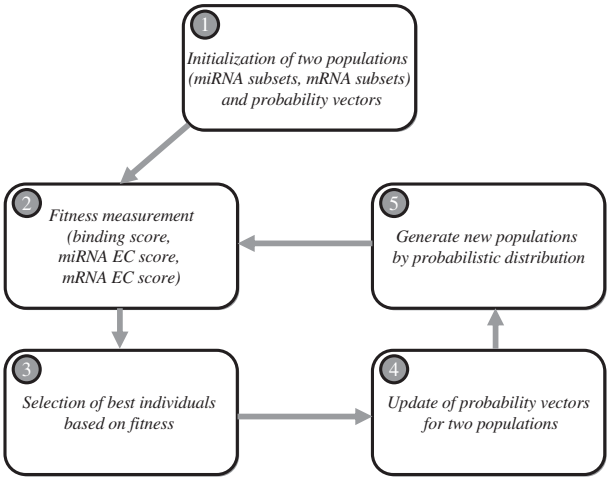

Fig. 2. Flowchart of the population-based probabilistic learning method for identifying miRNA–mRNA modules.

revised according to the number of times each miRNA (mRNA) is included in the selected set of individuals, as follows.

For the probabilistic learning, vectors of probabilities $p^m = (p_1^m, p_2^m, \ldots, p_{N_m}^m)$ and $p^t = (p_1^t, p_2^t, \ldots, p_{N_t}^t)$ are maintained. Here, p_j^m denotes the probability of choosing the jth element as a member of the individual set. The probability distribution of the $(q+1)$st generation is updated from the qth generation as

$$p_j^m(q+1) = (1-\delta_m)\cdot p_j^m(q) + \delta_m \cdot \frac{1}{R}\sum_{k=1}^{R} x_k^j$$

$$p_j^t(q+1) = (1-\delta_t)\cdot p_j^t(q) + \delta_t \cdot \frac{1}{V}\sum_{k=1}^{V} y_k^j \qquad (3)$$

where x_k^j (y_k^j) represents the jth value in the kth selected individual. The update rate is controlled by the two parameters, $\delta_m \in (0, 1]$ and $\delta_t \in (0, 1]$. When these parameter values are near zero, current probability vectors depend highly on the previous probabilities. Finally, new populations are generated based on these current probability distributions of miRNAs and mRNAs. The above procedure is iterated until the maximum number of generations is reached. Our probabilistic model for generating the next population is relatively simple. More complicated probabilistic models, considering the dependencies between miRNAs (or mRNAs), were not employed because of their extremely high computational burden. Figure 3 depicts the search procedure of this probabilistic learning algorithm for miRNA–mRNA module identification.

In the presented algorithm, two insulated populations of miRNAs and mRNAs evolve cooperatively. The performance of this type of optimization technique depends on interdependencies among the subcomponents, i.e. miRNAs and mRNAs. Each population is adapted dynamically in accordance with another population. This strategy can achieve good results via the decomposition of complex problems (Potter and De Jong, 2000; Zaritsky and Sipper, 2004). Co-adaptive changes of the two populations towards the optimum in our experiments are shown in Supplementary material.

Set $q = 0$
Initialize $X(0)$ and $Y(0)$, $p^m(0)$ and $p^t(0)$
Repeat
  Set $q = q + 1$
  For $i = 1$ to ψ // Evaluate the individuals for miRNAs
    Fitness measurement of x_i
  For $i = 1$ to ψ // Evaluate the individuals for mRNAs
    Fitness measurement of y_i
  Select R and V best individuals ($R < \psi, V < \psi$)
  For $j = 1$ to N_m // Update the probability vector for miRNAs
    $p_j^m(q+1) \longleftarrow p_j^m(q)$
  For $j = 1$ to N_t // Update the probability vector for mRNAs
    $p_j^t(q+1) \longleftarrow p_j^t(q)$
  Generate new populations $X(q+1)$ and $Y(q+1)$
Until q = max number of generations

Fig. 3. Search procedure of the population-based probabilistic learning for miRNA–mRNA module identification.

3 RESULTS

3.1 Preparation of the data sets

Expression profiles of mRNAs and miRNAs were extracted from the experimental data set of (Lu *et al.*, 2005). We used expression profiles of 217 miRNAs and about 16 063 mRNAs on 89 multiple human cancer samples. From these, we analyzed the relationships among 99 human miRNAs and 2012 mRNAs, which are linked together in miRBase Targets Version 3.0 (http://microrna.sanger.ac.uk/targets/v3/, (Griffiths-Jones *et al.*, 2006)). Of these 2012 mRNA × 99 miRNA possible binding pairs, 3982 pairs with significant binding scores (P-value $< 5 \times 10^{-3}$) were used in our experiments.

Some characteristics of the putative target-binding association between miRNAs and mRNAs were investigated. Figure 4a shows the distribution of the numbers of putative binding sites for miRNAs in the 3′ UTR of each mRNA. In our data set, one mRNA contains on average 1.97 binding sites. Figure 4b depicts the distribution of the numbers of target mRNAs of each miRNA. Here, one miRNA binds putatively to the 3′ UTR of 40.22 mRNAs on average.

3.2 Discovery of miRNA–mRNA modules

3.2.1 Parameter setting The parameters of the population-based probabilistic learning algorithm were set as follows. The population sizes, μ (for mRNAs) and ν (for miRNAs), were set to 4000 and 500, respectively. Parameters for controlling the update rate in Equation (3), δ_m and δ_t values were set as 0.8. The maximum number of iterations was set to 100. The balancing parameters in the fitness function (Equation (1)), α, β and γ were 0.6, 0.3 and 0.1, respectively. Parameters of the VOL term in the fitness functions w, w_m and w_t were set to 0.1, 0.5 and 0.5, respectively. In addition, the minimum subset size was two for miRNAs and five for mRNAs. Details of the parameter settings are given in Supplementary material.

Figure 5 shows the two best miRNA–mRNA modules found by the algorithm. The score of module I was 0.66 and that of module II was 0.78. To evaluate the significance of these scores, we estimated the distribution of scores of randomly chosen miRNA–mRNA modules. Figure 6 shows the

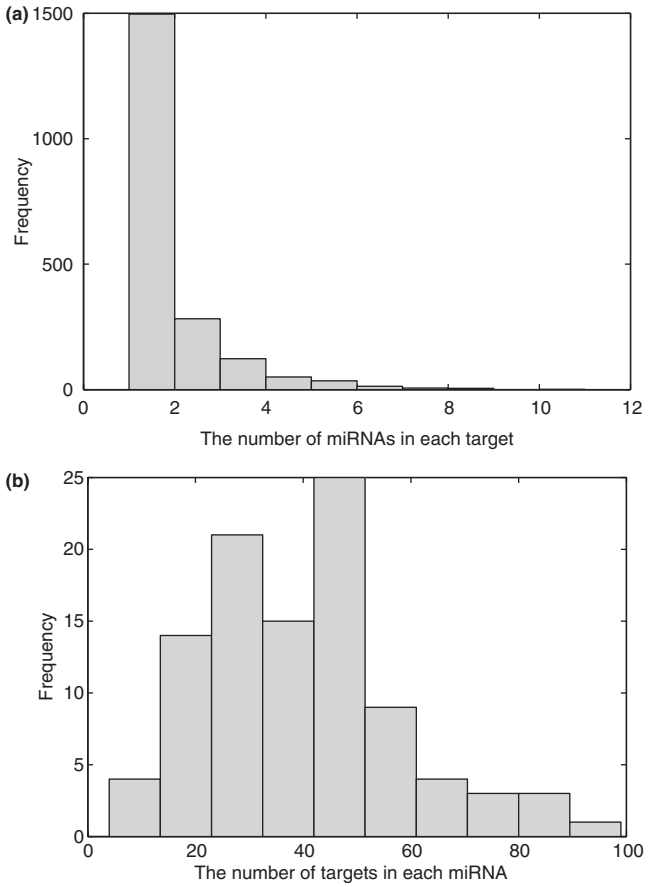

Fig. 4. Characteristics of the miRNA–target gene relationship. (**a**) Distribution of the number of miRNAs in each target mRNA. (**b**) Distribution of the number of target mRNAs in each miRNA.

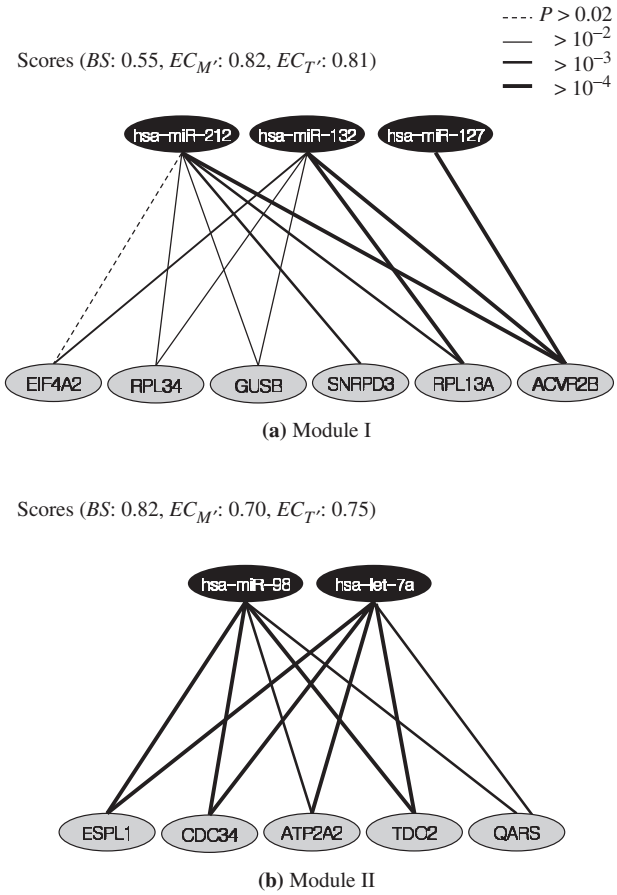

Fig. 5. Bipartite graphs representing the two best miRNA–mRNA modules. The upper (lower) ovals denote miRNAs (mRNAs). The thickness of the links corresponds to the significance of each miRNA–mRNA-binding event (*P*-values). The fitness score of each module is also shown here.

estimated distribution. Here, the scores of the two best modules are extremely high, meaning that the probability of finding these modules by chance is close to zero. Details of the predicted miRNA modules are described in Supplementary material.

3.3 Validation of the modularity

To validate the modules found, we calculated correlation coefficients between mRNAs. Our approach considers not only the similarity between mRNA expressions but also the putative miRNA–mRNA relationships and the similarity between the related miRNAs. Thus, the mRNAs included in our modules are expected to be much more correlated when the miRNAs are actively functioning, i.e. highly expressed. Differences in the correlation coefficients between the mRNA pairs included in the same module according to the miRNA expression level are contrasted in Figure 7. The *x*-axis denotes the correlation coefficients calculated using the subset of samples, where the mean expression level of the miRNAs included in the same module is in the upper 25%. The *y*-axis corresponds to the correlation coefficient across all samples. It can be shown that the expression correlation is likely to be much higher when the related miRNAs are highly expressed. The *P*-values denoting the significance of these differences were estimated by random sampling and are shown in Supplementary material.

3.4 Biological significance of the modules

In module I, three miRNAs share putative target genes. Of these, mir-212 and mir-132 are tandem-arrayed in the same chromosomal location within 300 bp and may be on a transcript (polycistronic miRNAs). They are located to a minimal loss of heterozygosity (LOH) region described in hepatocellular carcinomas (Calin *et al.*, 2004). Also, in their upstream and downstream regions, there are tumor suppressor genes, hypermethylated in cancer 1 (HIC1) and ovarian cancer gene-2 (OVCA2), respectively. Therefore, these polycistronic miRNAs may be involved in tumor suppression. The gene mir-127, which was downregulated in 75% of the human cancer cells tested, was induced strongly after treatment with chromatin-remodeling drugs (Saito *et al.*, 2006). Induction of mir-127 resulted in the downregulation of B-cell CLL/lymphoma 6 (BCL6), a known proto-oncogene. Therefore, mir-127 may have a role as a tumor suppressor, as with the clusters mir-212 and mir-132.

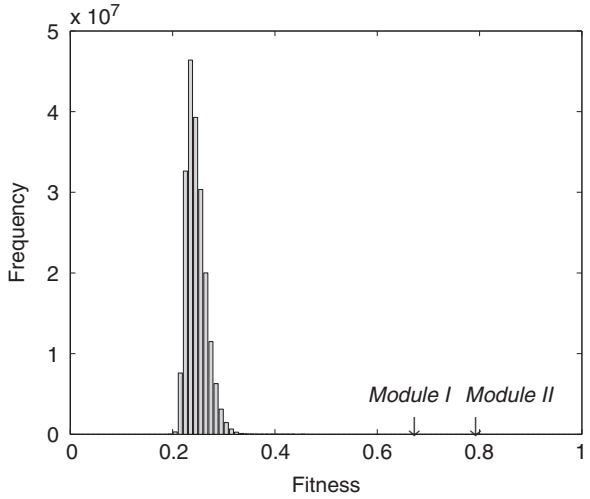

Fig. 6. Distribution of the fitness scores of randomly chosen miRNA–mRNA modules. This was estimated by 2×10^8 random samplings of modules consisting of three miRNAs and six mRNAs.

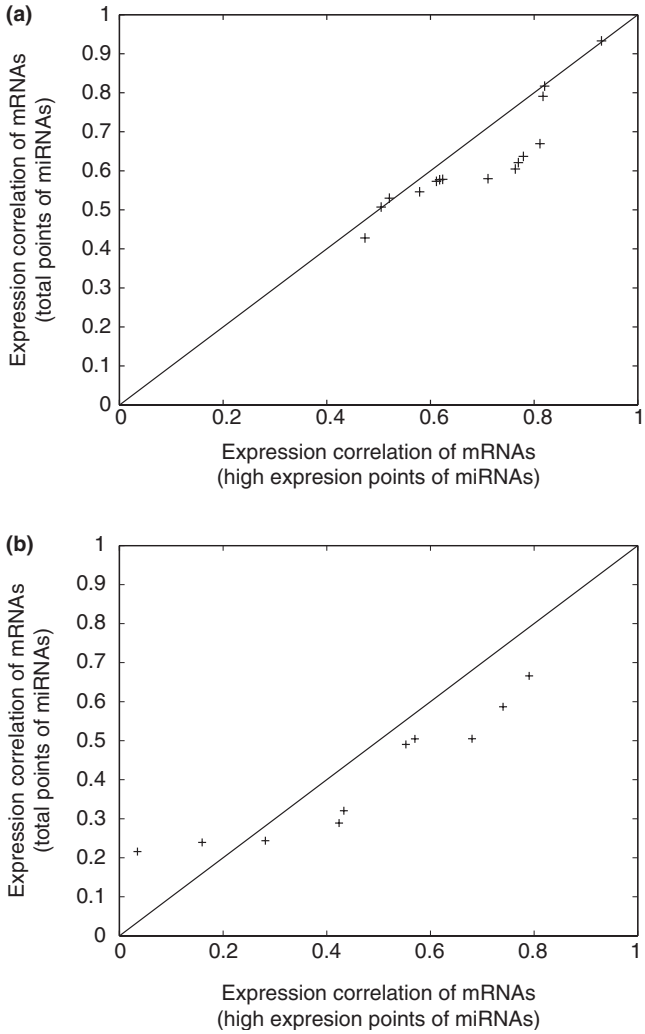

Fig. 7. Comparison of the expression correlations for the top 25% of samples with high miRNA expression (x-axis) and for all samples (y-axis). The mRNA expressions are much more highly correlated when the miRNAs are highly expressed in most cases (mRNA pairs).

Considering the miRNAs in the module as tumor suppressors, most of their target genes should be oncogenes. Indeed, among the target genes, EIF4A2, GUSB and ACVR2B are thought to be candidate oncogenes (Myllykangas *et al*., 2006). EIF4A2 is known as the partner of the BCL6 translocation (Akasaka *et al*., 2003). Overexpressed human GUSB increases the susceptibility to tumor formation (Donsante *et al*., 2001). Also, the loss of activin signaling through mutation of ACVR2 may be important in the genesis of MSI-H colorectal cancer (Jung *et al*., 2004). Although RPL34, RPL13A and SNRPD3 are considered as genes coding for structural proteins, they can show differences in expression levels between malignant and non-malignant tumor pairs. For example, the expression of RPL13A in prostate cancer tissue differs (is upregulated) significantly from the norm (Ohl *et al*., 2005). All genes in module I can be assumed to be involved in common biological functions. Although the GO annotation (Table 1) of this module is related to metabolism, the assigned genes are specifically associated with cancer in the literature.

Module II consists of two miRNAs and their five target genes. The miRNA let-7a has been reported as a tumor suppressor by negatively regulating the Ras oncogene in lung tissue (Johnson *et al*., 2005), and the other members of the let-7 family that have a similar mature sequence and expression pattern are suspected to code for cancer-related miRNAs. Both mir-98 and let-7f-2 are on an intron of the HUWE1 gene, which ubiquitinates the p53 tumor suppressor and core histones (Chen *et al*., 2005). We suspect that the mir-98 and let-7f-2, processed products of the HUWE1 transcript, may modulate or attenuate cancer development. In fact, their target genes are also involved in cancer development or the cell cycle. In terms of GO category (Table 1), they are involved in two forms of biological functions (cell cycle or metabolism).

For example, ESPL1 plays a central role in chromosome segregation by cleaving the SCC1/RAD21 subunit of the cohesin complex at anaphase (Zou *et al*., 2002) and CDC34 is involved in the late G1-to-S transition, a key regulator of the cell cycle (Pagano, 1997). Collectively, we can conclude that the predicted module II is an important cancer-related miRNA–mRNA module with strong correlation.

4 DISCUSSION AND CONCLUSIONS

Critical cellular processes can be affected by miRNAs. However, their precise functional roles are still largely unknown. Many miRNAs regulate gene expression by binding to and inhibiting mRNAs. Although this phenomenon is mostly observed in plants, our results presented in Figure 7 could serve as preliminary evidence for the posttranscriptional regulation by miRNAs in animals. Some experimental results have been reported that animal miRNAs could also guide the cleavage of endogenous targets (Yekta *et al*., 2004). Various miRNA expression profiles have been observed in human

Table 1. Biological processes enriched in each module

GO ID	*P*-value*	Biological process	Genes
Module I			
GO:0043170	4.45E−02	Macromolecule metabolism	EIE4A2, RPL34, GUSB, SNRPD3, RPL13A, ACVR2B
Module II			
GO:0000278	1.62E−02	Mitotic cell cycle	ESP1, CDC34
GO:0009308	3.18E−02	Amine metabolism	QARS1, TDO2

*Adjusted *P*-value by FDR. GO terms were extracted for the mRNAs in each module. Over-represented terms were chosen by hypergeometric testing and multiple testing adjustment using the false discovery rate (FDR) procedure ($P < 0.05$).

tumors and diseases. The advantage of using expression profiles is that modular miRNAs can be detected in a broader biological context than when only using miRNA–target duplexes. Furthermore, the information on mRNA expression profiles may be used as additional evidence that they are indeed target genes of miRNAs.

In the miRNA–mRNA modules discovered using our method, expression patterns of miRNAs as well as mRNAs were highly correlated. Moreover, expressions of the mRNAs in the same module were more strongly correlated when the miRNAs in that module were highly activated (i.e. expressed) than across the whole sample. The mRNAs included in the same module also shared similar biological functions, demonstrating the ability of our method to detect functionally related genes. The relationship between cancer and our modules was also analyzed using a literature survey, confirming the effectiveness of this proposed method for finding biologically meaningful subsets of miRNAs and mRNAs.

We proposed a population-based probabilistic search method for identifying miRNA modules from their predicted miRNA targets as well as their expression data sets. The proposed discovery method facilitates the incorporation of multiple heterogeneous information sources by adopting a balanced fitness function and coevolutionary learning strategies. As the two groups of miRNAs and mRNAs are fitted cooperatively to the best solution, the algorithm found miRNA–mRNA modules with significantly high fitness scores. The search strategy is based on the use of global statistical information contained in subsets within populations. Thus, it can search efficiently in parallel for the global optimum.

Our method is related to biclustering methods that have been noted in various biological issues (Madeira and Oliveria, 2004). The biclustering approach groups rows and columns simultaneously in a two-dimensional matrix. Similarly, our method tries to cluster miRNAs and mRNAs simultaneously. Most biclustering methods have been specified in microarray data sets (Cheng and Church, 2000; Madeira and Oliveira, 2004). They identified groups of genes that show correlations between their expression patterns under given experimental conditions. Whereas they focused on single resource, our method deals with diverse resources using well-designed objective functions.

To discover more significant biological modules summarizing multiple aspects of biological system, the computational approach needs to deal with heterogeneous biological data sets. For example, regulatory modules can be revealed by clustering rows and columns simultaneously in a two-dimensional matrix composed of heterogeneous genome-wide resources of expression profiles as well as genomic sequences (Joung *et al.*, 2006). Our method can be applied to detect miRNA modules as well as other types of biological modules from multiple resources.

In spite of the advantage obtained by utilizing diverse resources, our method needs prior setting of several parameters of the fitness function. Multi-objective optimization techniques, such as the Pareto-based ranking scheme (Deb, 2001; Fonseca and Fleming, 1993), could be exploited to reduce the number of parameters by implicit balancing between several objectives. We used expression data sets of miRNAs and mRNAs that were obtained from independent experiments. Therefore, the measurement of correlations between two expression data sets may produce additional artifacts, which can affect the analysis adversely. The regulatory effects of multiple miRNAs and mRNAs could be investigated more precisely from data sets that observe expression profiles of miRNAs and mRNAs and proteins for the same samples at the same time. Comparative analysis of homologous species (e.g. mouse) is another research direction for detecting miRNA–mRNA modules more precisely.

ACKNOWLEDGEMENTS

This research was supported in part by the National Research Laboratory Program of the Korea Ministry of Science and Technology (MOST) and in part by the Soongsil University Research Fund.

Conflict of Interest: none declared.

REFERENCES

Akasaka,T. *et al.* (2003) BCL6 gene translocation in follicular lymphoma: a harbinger of eventual transformation to diffuse aggressive lymphoma. *Blood*, **102**, 1443–1448.

Baluja,S. (1994) Population-Based incremental learning: a method for integrating genetic search Based function optimization and competitive learning. Technical Report CMU-CS-94-163, Carnegie Mellon University, Pittsburgh.

Baluja,S. (1997) Genetic algorithms and explicit search statistics. *Adv. Neural Inf. Process. Syst.*, **9**, 319–325.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

Calin,G.A. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *PNAS*, **101**, 2999–3004.

Chen,D. *et al.* (2005) ARF-BP1/Mule is a critical mediator of the ARF tumor suppressor. *Cell*, **121**, 1071–1083.

Cheng,Y. and Church,G. (2000) Biclustering of expression data. In *Proceedings of ISMB*. pp. 93–103.

Deb,K. (2001) *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester Wiley, UK.

Denli,A.M. *et al.* (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.

Donsante,A. *et al.* (2001) Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. *Gene The.*, **8**, 1343–1346.

Fonseca,C.M. and Fleming,P.J. (1993) Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In *Proceedings of ICGA,* pp. 416–423.

Griffiths-Jones,S. *et al*. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucl. Acids Res*., **34**, D140–D144.

Han,J. *et al*. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.

Huang,J.C. *et al*. (2006) Detecting microRNA targets by linking sequence, microRNA and gene expression data. In *Proceedings of RECOMB,* pp. 114–129.

Johnson,S.M. *et al*. (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–647.

Joung,J.-G. *et al*. (2006) Identification of regulatory modules by co-clustering latent variable models: stem cell differentiation. *Bioinformatics*, **22**, 2005–2011.

Jung,B. *et al*. (2004) Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. *Gastroenterology*, **126**, 654–659.

Larrañaga,P. and Lozano,J.A. (2002) *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer Academic Publishers.

Lewis,B.P. *et al*. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Liu,C.G. *et al*. (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *PNAS*, **101**, 9740–9744.

Lu,J. *et al*. (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.

Madeira,S.C. and Oliveira,A.L. (2004) Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform*., **1**, 24–45.

Myllykangas,S. *et al*. (2006) DNA copy number amplification profiling of human neoplasms. *Oncogene*, **25**, 7324–7332.

Nam,J.-W. *et al*. (2006) ProMiR II: A web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucl. Acids Res*., **34**, W455–W458.

Ohl,F. *et al*. (2005) Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? *J. Mol. Med*., **83**, 1014–1024.

Pagano,M. (1997) Cell cycle regulation by the ubiquitin pathway. *FASEB J*., **11**, 1067–1075.

Potter,M.A. and De Jong,K.A. (2000) Cooperative coevolution: an architecture for evolving coadapted subcomponents. *Evol. Comput*., **8**, 1–29.

Saito,Y. *et al*. (2006) Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell*, **9**, 435–443.

Thomson,J.M. *et al*. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nat. Methods*, **1**, 47–53.

Yekta,S. *et al*. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.

Yoon,S. and De Micheli,G. (2005) Prediction of regulatory modules comprising microRNAs and target genes. *Bioinformatics*, **21**, ii93–ii100.

Zaritsky,A. and Sipper,M. (2004). Coevolving solutions to the shortest common superstring problem. *BioSystems*, **76**, 209–216.

Zhang,B.-T. (2003) A unified Bayesian framework for evolutionary learning and optimization. In *Advances in Evolutionary Computation*, Chapter 15. Springer-Verlag, pp. 393–412.

Zilberstein,C.B.-Z. *et al*. (2006) A high-throughput approach for associating microRNAs with their activity conditions. *J. Comput. Biol*., **13**, 245–266.

Zou,H. *et al*. (2002) Anaphase specific auto-cleavage of separase. *FEBS Lett*., **528**, 246–250.