Predicting MicroRNA Targets by Integrating Sequence and Expression Data in Cancer

Naifang Su[†*], Yufu Wang[†*], Minping Qian[†‡] and Minghua Deng[†‡§]
[†]LMAM, School of Mathematical Sciences, Peking University, Beijing, P.R.China
[‡]Center for Theoretical Biology, Peking University, Beijing, P.R.China
[§]Center for Statistical Sciences, Peking University, Beijing, P.R.China
[*]These authors contributed equally to this work
Email: dengmh@pku.edu.cn

***Abstract*—Gene regulation is a key factor in gaining a full understanding of molecular biology. microRNA (miRNA), a novel class of non-coding RNA, has recently been found to be one crucial class of post-transactional regulators, and play important parts in cancer. One essential step to understand the regulatory effect of miRNAs is the reliable prediction of their target mRNAs. Typically, the predictions are solely based on sequence information, which unavoidably have high false detection rates. Here we develop a new algorithm called HCTarget, which predict miRNA targets by integrating the typical algorithm and the paired expression profiles of miRNA and mRNA. HCTarget formulates a linear model to characterize the relationship between mRNA and miRNA, and use a Markov Chain Monto Carlo algorithm to learn the target probabilities. When applying HCtarget to the expression data in multiple myeloma, we predict target genes for ten cancer related miRNAs. The experimental verification and a loss of function study of hsa-miR-16 validate our predictions. Compared with the previous approaches, our target sets have increased functional enrichment. Meanwhile, our predicted target pair hsa-miR-19b and SULF1 plays an important role in multiple myeloma. Therefore, HCtarget is a reliable and effective approach to predict miRNA target genes, and could improve our comprehensive understanding of gene regulation.**

I. Introduction

Discovering gene regulation is one of the main goals in molecular biology. Specifically, uncovering the mechanisms underlying the expression of tumor related genes is a key factor in gaining a full understanding of cancer biology [1], which is also of great therapeutic significance.

While previously a great deal of study has focused on transcriptional factors (TFs), one crucial class of regulators at the transcriptional level, the post-transcriptional regulator microRNA (miRNA) has arrested much attention recently [2]. miRNAs are a noval class of endogenous $\sim$22nt noncoding RNAs. They down regulate gene expressions through specific binding to the 3'-untranslated region of target mRNAs, which lead to their translational repression or degradation [3]. More than 1000 miRNAs have been annotated in human genome, and they are predicted to regulate up to one third of all protein-coding genes [4].

Experimental analysis has recognized that miRNAs control the key cellular processes such as growth, development and apoptosis [5][6][7]. It has been established that miRNAs make an important contribution to gene regulation in embryonic development and a growing list of human disease, especially cancer [8][9]. Previous studies have verified that miRNAs can act as tumor suppressors or oncogenes and their dysregulation is widely involved in cancer initiation and progression [10], which enable their inhibition to be a novel therapeutic strategy for cancer [11][12].

An essential step and major challenge in understanding miRNA regulatory function is the identification of their target genes [13]. Since it is infeasible to carry out high throughput experiments, only a small fraction of miRNA targets have experimental supports [14]. Typically, the target prediction is achieved by computational approaches, such as TargetScan [15][16], PicTar [17] and miRanda [18]. They mainly base on the "seed match" of miRNA and mRNA sequences, as well as miRNA's phylogenetic conservation [19]. However, these sequence based approaches have high false-positive rate, such as 22-31% for TargetScan [20]. Since the seed match complementation could not discern the real targets effectively, great deals of fake targets are confounded.

Recently, a novel strategy has been developed to predict miRNA targets by integrating experimental data to the sequence information [21][22][23][24][25][26][27]. The major experimental data is miRNA expressions, which have been putatively used to investigate the role of miRNA in many biological processes, especially cancer [6][10]. It has been verified that the expression of mRNAs targeted by highly expressed miRNAs are negative shifted compared with the background [25], which ensures that the significantly negative correlated miRNA-mRNA pairs have high potential to be the real target pairs . Therefore, the paired miRNA and mRNA expression data, which profile miRNA and mRNA expression levels simultaneously from the same sample, could serve as a suitable evidence to predict the actual miRNA targets [26].

Based on this idea, the first and most widely used approach to predict miRNA targets from sequence and expression data is GenMiR++ (Generative model for miRNA regulation) [25][26][27]. This approach characterizes mRNA expressions as a linear combination of the regulatory effects of their targeting miRNAs, and a variational Bayesian algorithm is used to learn the latent miRNA target indicators. GenMiR++ has been successfully applied on the expression data among normal and cancer tissues to predict human miRNA targets. However, it has several restrictions. First, originating from

2011 IEEE International Conference on Systems Biology (ISB)
978-1-4577-1666-9/11/$26.00 ©2011 IEEE

the experiments of different tissues, GenmiR++ characterizes miRNA's relative effects among all tissues as a constant. This assumption may not hold when considering the experiments of different cancer patients. Since patients have much more varieties, their miRNA's relative effects could not be regarded as a constant anymore. Second, GenMiR++ uses variational Bayesian algorithm to learn the parameters. Instead of leaning the real posterior distribution, variational Bayesian use a variational posterior, which restricted to the factorized form, to approximate the real one [28]. This algorithm is putatively used when the posterior distribution is difficult or time-consuming to compute directly. Its convergence rate highly depends on the form of the likelihood and priors and may be extremely slow. This restriction has been relaxed by using Metropolis-Hasting algorithm [22]. However, this approach is limited in its high computational complexity.

Here we propose a new algorithm called HCtarget (High Confident targets) to integrate expression and sequence information to detect miRNA targets. Our approach develops GenMiR++ and overcomes the above restrictions, by re-defining the parameters of miRNA effects and using a Markov chain Monte Carlo (MCMC) algorithm to learn the posterior directly. We first evaluate the performance of HCtarget by a simulation study, and then access its robustness. Furthermore, the experimental verification is extracted to compare the reliability of HCtarget and GenMiR++. Meanwhile, we refer to a loss of function study of hsa-miR-16 to validate our predictions. In addition, we also study the functional enrichment to investigate the biological significance of our predicted targets. Moreover, a specific target pair is selected to explore miRNA's role in multiple myeloma.

II. RESULTS

In HCtarget, a linear model is formulated to characterize the relationship between the paired miRNA-mRNA expressions, and we use a MCMC algorithm to estimate the parameters. The output of our model is the target probability of each miRNA-mRNA pair (See Materials and Methods).

A. Performance of HCTarget on the simulation data

We first evaluated the performance of GenMiR++ and HCTarget on the simulation data. These data have similar expression patterns with real data, with the actual target relations known (See Materials and Methods). Compared with the real targets, we computed the true positive rate and false positive rate of GenMiR++ and HCTarget with different cutoffs. Their ROC (Receiver operating characteristic) curves and AUC (the area under the ROC curve) values are shown in Figure 1, which indicate that HCTarget has higher accuracy than GenMiR++.

B. Predict miRNA targets based on cancer expression data

We then applied our HCtarget approach to detect miRNA targets in cancer. The miRNA and mRNA expression profiles of patients with multiple myeloma are extracted, and we integrated them with the computational predictions in TargetScan. Since our approach aim to discover cancer related target genes, we selected ten miRNAs and 1000 genes that specifically expressed in multiple myeloma for our prediction (See Materials and Methods). TargetScan provides 1401 target pairs for these miRNAs and genes. HCtarget cuts down these predictions to 647, while 699 target pairs are obtained by GenMiR++.

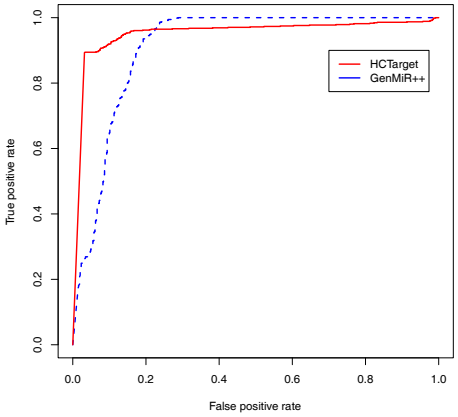

Fig. 1. **The ROC curves of HCTarget and GenMiR++ for simulation data.** Their AUC values are 0.95 and 0.91 respectively.

C. Assess the Robustness of HCtarget

We performed a series of permutation tests to assess the robustness of our approach [25]. We permuted the gene labels are permuted 100 times and generated 100 random data sets. In these sets, the relationship between miRNAs and mRNAs are destroyed and their predicted target probabilities could be regarded as background. These permutations allow us to evaluate whether our model would be affected by introducing a great deal of fake targets into the candidates. Comparing the predictions of HCtarget for both permuted and original data, we found the probabilities leaned from the real data are significantly higher than the background. The p value of one side wilcoxon test is 0.06. It illustrates that HCtarget could successfully discriminate the real target from the fake ones, which ensures its robustness in target prediction.

D. Compare with experimentally verifications

To evaluate the accuracy of our approach, we extracted experimentally supported miRNA targets from Tarbase [14](See Materials and Methods). For the multiple myeloma related miRNAs and mRNAs, three miRNAs have biological experiments, and their 17 target genes in TargetScan have experimental verifications. Nine of them are detected by HCtarget, while GenMiR++ only identifies two. The numbers of verified targets predicted by TargetScan, GenMiR++ and HCTarget as well as their precisions are listed in Table I, which show that HCtarget could identify more accurate targets than GenMiR++. For example, mir-15 has nine supported targets, seven of them are detected by HCtarget, while GenMiR++ failed to identify any

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

of them. It also indicates that HCtarget has higher precision than the original TargetScan.

TABLE I
COMPARISON WITH TARBASE

miRNA family	TargetScan	GenMiR++	HCTarget
let-7	7 (3.57%)	2 (2.02%)	2 (2.15%)
mir-15	9 (4.02%)	0 (0)	7 (6.67%)
mir-29	1 (0.51%)	0 (0)	0 (0)
total	17 (2.76%)	3 (0.95%)	9 (3.01%)

E. Validate hsa-miR-16 targets

Previous analysis suggests that hsa-miR-16 can act as a tumor suppressor in multiple myeloma [29]. A loss of function study inhibited hsa-miR-16 and identified a large number of different expressed genes as its target genes (See Materials and Methods) [29]. To validate our prediction, we compared our detected targets with these different expressed genes. TargetScan identifies 224 targets for hsa-miR-16, 34 of them have different expression levels when hsa-miR-16 is deleted (the p value of hyper-geometric test is 0.14). HCtarget, which cuts down the target genes to 105, provides 22 validated targets ($p = 0.006$) (Figure 2). This represents that HCtarget has more confirmed targets than TargetScan. In addition, GenmiR++ only detects 11 different expressed genes ($p = 0.72$), which also validates the accuracy of HCtarget.

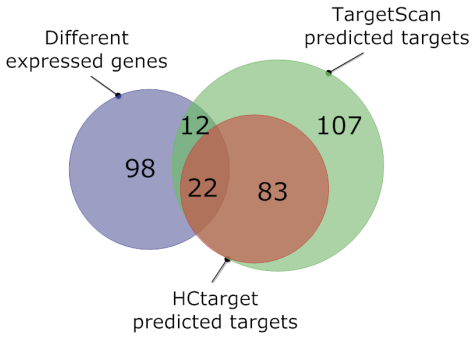

Fig. 2. **Venn diagram**. It shows the overlap of different expressed genes with the predicted targets of targetScan and HCtarget.

F. Gene Ontology enrichment analysis

To have further investigation of our predicted targets, we analyzed their function annotations in Gene Ontology (GO)(See Materials and Methods) [30]. For each target set detected by TargetScan and HCtarget respectively, we computed its GO enrichment p value using hyper geometric test. Considering multiple testing problems, these p values are corrected using FDR modification. For TargetScan, we found 107 (2.5%) functional target sets (with FDR$<$0.1). While there are 135 (3.1%) functional sets of GenmiR++ and HCtarget increases the number to 158 (3.7%). The comparison exhibits that the targets of HCtarget have significantly more consistent functional annotations.

Meanwhile, we selected the GO functions that significantly enriched (FDR$<$0.01) in hsa-miR-19b, which has been experimentally verified to be a key regulator in multiple myeloma [31]. They are: GO0034612 (response to tumor necrosis factor), GO0000723 (telomere maintenance), GO0006289 (nucleotide-excision repair), GO0006302 (double-strand break repair) and GO0045732 (positive regulation of protein catabolic process). The first annotation is significantly associated with multiple myeloma, the latter three ones are crucial functions in cell division, a key cellular process in cancer, while the last one is putative important in metabolism. These findings demonstrate that HCtarget could successfully identify the functional miRNA targets.

G. Example

Based on the above findings, we further focused on a specific target pair to discover miRNA's role in multiple myeloma. hsa-miR-19b is selected, and one of its targets detected by HCtarget is SULF1, which has been found to be a potent inhibitor of myeloma tumor growth [32]. We focused the patients with higher hsa-miR-19b expressions (with expression level larger than average), and discovered that the expression levels of SULF1 are significantly lower in these patients than in the other ones (the p values of the one side wilcoxon test is 0.1). Their cumulative distributions (Figure 3) displays that the expression of SULF1 is negatively shifted when hsa-miR-19b is highly expressed. This example further confirms the significant down regulatory effects of hsa-miR-19b, and provides us a reliable target gene SULF1. We believed that this target pair plays a crucial role in multiple myeloma and could be served as effective candidates for the therapeutic treatment.

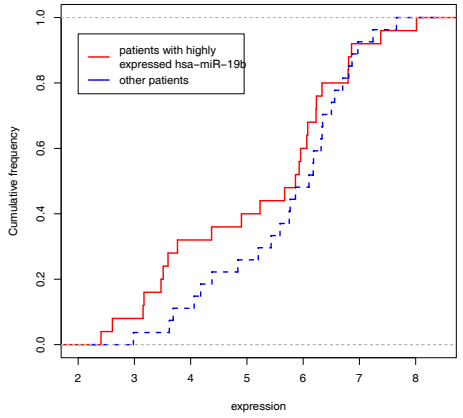

Fig. 3. **The down regulatory effect of hsa-miR-19b on SULF1.** The cumulative distributions of the expression levels of SULF1 in the sample with or without highly expressed hsa-miR-19b (red solid line and blue dashed line respectively).

III. CONCLUSION

In this paper, we propose a new algorithm, HCtarget, to predict miRNA target genes by integrating the sequence

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

information and expression profiles. The simulation study and the robustness assessment confirm the accuracy of our approach. The investigations of the expression profiles in multiple myeloma also exhibit the well performance of HCtarget.

Our model affords reliable targets of miRNA, which improve our understanding of miRNA's roles in cancer. Such as the disease related target pair, hsa-miR-19b and SULF1, is beneficial for the further discovery and clinical treatment of multiple myeloma.

Here we focus on the target prediction for miRNA and genes that specifically related with multiple myeloma, which could help our understanding of cancer biology. By selecting some other proper miRNA and mRNA expression profiles, HCtarget could be generalized to provide miRNA's whole genome target predictions, which is beneficial for the comprehensive discovering of miRNA's regulatory effects.

Although our model is designed to predict miRNA targets, it could be directly generalized to detect the target genes of TFs. In addition, previous studies demonstrated that TFs, or their cis-regulatory modules, have widely cooperation with miRNAs. Their combinatorial regulatory modules play important parts in gene regulation [33]. With accurate target predictions of miRNAs and TF, HCtarget could be an effective way to reconstruct gene regulatory network, which helps us to uncover the mechanisms underlying gene expression.

IV. Materials and Methods

In this section, we describe our HCtarget approach which integrated miRNA and mRNA expression data and sequence information to predict miRNA target gene.

A. Model

Suppose the paired miRNA and mRNA expression data profile N mRNAs and M miRNA across T samples. Let y_{it} denote the expression level of mRNA i in sample t and z_{jt} denote the expression level of miRNA j in the same sample, where $i = 1, \ldots, N, j = 1, \ldots, M$ and $t = 1, \ldots, T$.

Similar to GenMiR++, we take a linear model to formulate the relations between mRNA expressions and the regulatory effects of their targeting miRNAs [34]. A latent binary variable R is used to indicate the target relations, where $r_{ij} = 1$ if mRNA i is targeted by miRNA j, and 0 otherwise. The relationship between mRNA and miRNA expressions is formulated as:

$$y_{it} = \beta_{0t} + \sum_{j=1}^{M} r_{ij} z_{jt} \beta_{jt} + \epsilon_{it}, \epsilon_{it} \sim N(0, \sigma_t^2) \qquad (1)$$

where y_{it} and z_{jt} are the observed expressions, β_{jt} represents the regulatory effects of miRNA j at sample t (in GenMiR++, this term is factored into the product of the tissue effect and the miRNA effect $\gamma_t \lambda_j$), and β_{0t} is the background effect of sample t.

The goal of our model is to estimate the latent indicators R. Considering the known computational predictions based on mRNA and miRNA sequences, denoted as C, as $c_{ij} = 1$ if mRNA i is predicted to be targeted by miRNA j, and 0 otherwise, we suppose R have a Bernoulli distribution depend on C. That is $r_{ij} \sim bernoulli(\pi)$ in the condition of $c_{ij} = 1$, and $r_{ij} = 0$ when $c_{ij} = 0$. In the following discussion, we focus on the pair with $c_{ij} = 1$. The likelihood of R is:

$$p(R|\pi) \propto \prod_{ij} \pi^{c_{ij} r_{ij}} (1 - \pi)^{c_{ij}(1 - r_{ij})}$$

here π can be regarded as the accuracy of the sequence based predictions. This assumption enables our model to cut down the false positive rate of the previous prediction.

Let $B_t = (r_{ij} z_{jt}), A_t = [1, B_t], y_t = (y_{1t}, \ldots, y_{Nt})^T, Z = (z_{jt}), \beta_t = (\beta_{0t}, \ldots, \beta_{Mt})^T$ and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})^T$, we have the vector representation of our model:

$$Y_t = A_t \beta_t + \epsilon_t \qquad (2)$$

B. MCMC Algorithm for Statistical Inference

Based on the above model, the likelihood of the observed data $p(Y, Z, C, R|\beta, \sigma^2, \phi)$ is:

$$\prod_{i,t} e^{\left[-\frac{1}{\sigma_t^2}(y_{it} - \sum_{j=1}^{M} z_{jt} r_{ij} \beta_{jt} - \beta_{0t})^2\right]} \prod_{i,j} \pi^{c_{ij} r_{ij}} (1-\pi)^{c_{ij}(1-r_{ij})}$$

To estimate the parameters $\theta = (\beta, \sigma^2, \pi)$ and latent variables R, we apply the Bayesian methodology and a MCMC algorithm [35]. With proper prior assumptions, the posterior of R and θ have simple forms and could be directly computed using a MCMC algorithm as the following iterations [36][37]: (i) sample the parameters θ conditional on the updated latent variable; (ii) sample the latent variable R conditional on the updated parameters .

1) Updata the parameters:

Given the non-informative prior $p(\beta_t, \sigma_t^2) \propto \sigma_t^{-2}$, the posterior distributions of β_t and σ_t are

$$\beta_t|\sigma_t^2, Y \sim N(\hat{\beta}_t, (A_t^T A_t)^{-1} \sigma_t^2),\ \sigma_t^2|Y \sim v s_t^2 \chi_v^{-2} \qquad (3)$$

where $v = N - M - 1$ and

$$\hat{\beta}_t = (A_t^T A_t)^{-1} A_t^T Y_t,\ \hat{Y}_t = A_t^T \hat{\beta}_t, s_t^2 = \frac{1}{v}(Y_t - \hat{Y}_t)^T (Y_t - \hat{Y}_t).$$

While for π, with the conjugate prior $\pi \sim Beta(a_0, b_0)$, the posterior distirbution is

$$\pi \sim Beta(n_1 + a_0, n_0 + b_0) \qquad (4)$$

where $n_1 = \sum_{ij} c_{ij} r_{ij}$ and $n_0 = \sum_{ij} c_{ij}(1 - r_{ij})$

2) Updata the latent variable:

The marginal distribution of the latent variable $p(r_{ij}|c_{ij} = 1, Y, Z, \theta)$ is

$$\exp\left[-\sum_{t=1}^{T} \frac{1}{\sigma_t^2}(y_{it} - \sum_{k} z_{kt} r_{ik} \beta_{kt} - \beta_{0t})^2\right] \pi^{c_{ij} r_{ij}} (1-\pi)^{c_{ij}(1-r_{ij})}$$

Since

$$[y_{it} - \sum_{k} z_{kt} r_{ik} \beta_{kt} - \beta_{0t}]^2 = [y_{it} - \sum_{k \neq j} z_{kt} r_{ik} \beta_{kt} - \beta_{0t}]^2 + q_{ijt} r_{ij}$$

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

here q_{ijt} denotes

$$z_{jt}^2\beta_{jt}^2 - 2y_{it}z_{jt}\beta_{jt} + 2\sum_{k\neq j} z_{kt}\beta_{kt}z_{jt}\beta_{jt}r_{ik} + 2z_{jt}\beta_{jt}\beta_{0t}$$

then $p(r_{ij}|\cdot) \propto \exp(-\sum_{t=1}^{T}\frac{q_{ijt}}{\sigma_t^2}r_{ij})\pi^{c_{ij}r_{ij}}(1-\pi)^{c_{ij}(1-r_{ij})}$

that is, r_{ij} has Bernoulli marginal distribution

$$p(r_{ij}|\cdot) \sim bernoulli(p_{ij}) \tag{5}$$

with updated probability $p_{ij} = \frac{(\frac{\pi}{1-\pi})^{c_{ij}}}{(\frac{\pi}{1-\pi})^{c_{ij}} + \exp(\sum_{t=1}^{T}\frac{q_{ijt}}{\sigma_t^2})}$

3) The algorithm of HCtarget:

Based on the above discussion, we use a traditional MCMC approach to estimate the parameters and the latent variable iteratively:

1) Initial β_t, σ_t, R as $\beta_t = 1, \sigma_t = 1$ and $r_{ij}|c_{ij} = 1 \sim bernoulli(0.5)$.
2) Update σ_t^2 by sampling from $vs_t^2\chi_v^{-2}$, update β_t by sampling from $N(\hat{\beta}_t, (A_t^TA_t)^{-1}\sigma_t^2)$ and update π by sampling from $beta(n_1 + a_0, n_0 + b_0)$.
3) Given the updated parameters, sample the latent variable r_{ij} from $bernoulli(p_{ij})$.
4) Repeat the above two steps until convergence. Here the convergence is evaluated by Gelman and Rubin criteria [37].

We output p_{ij}, which represents the posterior probability that miRNA j targets mRNA i, for our final prediction. miRNA-mRNA pairs with p_{ij} larger than a certain threshold are the putative target pairs of our model. In the analysis of cancer expression data, we specify the threshold as 0.8.

C. Data sources

We selected TargetScan as the computational algorithm to predict miRNA targets based on their sequence. In all, 9448 targets of 249 miRNAs were assembled from TargetScanHuman (release 5.1) [15].

For the paired miRNA-mRNA expression data, we used the profiles from 52 patients with multiple myeloma [10]. The data was downloaded from GEO database [38] with the accession number GSE17306.

We selected multiple myeloma related miRNAs and mRNAs for our predictions. Ten miRNAs with the highest expression level are picked up, they are: hsa-let-7g, hsa-miR-142-3p, hsa-miR-148a, hsa-miR-16, hsa-miR-19b, hsa-miR-21, hsa-miR-26a, hsa-miR-29c, hsa-miR-370 and hsa-miR-494. Meanwhile 1000 mRNAs are selected, half with the highest expressions and half with the lowest expressions, since miRNA putatively repress gene expressions and may have secondary up-regulatory effects [34].

Furthermore, we downloaded the experimentally supported human miRNA targets from Tarbase (v.5c) [14], which houses 1033 miRNA-mRNA target pairs for 864 genes and 111 human miRNAs. To compare Tarbase with our predictions, miRNAs were all mapped to miRNA families using the annotations in miRBase [4].

In addition, we extracted genes' function annotations from Gene Ontology [30] (data were downloaded on April 20, 2011). GO is a rooted directed acyclic graph on three categories, and here we focused on biological process. Similar to the above approach [39], we selected the GO nodes which cover at least 50 genes. Finally 429 function annotations for our selected genes were assembled.

Profile of a loss of function study of hsa-miR-16 was extracted from GEO database (GSE24522). It provides gene expression levels before and after hsa-miR-16 deletion [29]. We focused on genes with fold change larger than 1.5 as different expressed genes. For our 1000 genes, 132 genes were selected.

D. Simulations

The simulation data is generated based on the above selected miRNAs and mRNAs from the following rule. First, the ten miRNA expression data Z are extracted from the real data from patients with multiple myeloma. Then the 1000 mRNA expressions Y are simulated from

$$y_{it} = \beta_{0t} + \sum_{j=1}^{10} r_{ij}z_{jt}\beta_{jt} + \epsilon_{it}, i = 1, \ldots 1000, t = 1, \ldots, 52$$

here β_{jt}, β_{0t} and ϵ are generated from $N(-0.3, 0.1)$,$N(1,1)$ and $N(0,1)$ respectively. The real target relations r_{ij} is obtained from $bernoulli(0.5)$ conditions on $c_{ij} = 1$, where c_{ij} represents the predictions from TargetScan.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.10871009, No.10721403), the National High Technology Research and Development of China (No.2008AA02Z306), and the National Key Basic Research Project of China (No.2009CB918503).

REFERENCES

[1] S. Volinia, G. A. Calin, C. G. Liu, S. Ambs, A. Cimmino, F. Petrocca, R. Visone, M. Iorio, C. Roldo, M. Ferracin, R. L. Prueitt, N. Yanaihara, G. Lanza, A. Scarpa, A. Vecchione, M. Negrini, C. C. Harris, and C. M. Croce, "A microRNA expression signature of human solid tumors defines cancer gene targets," *Proc Natl Acad Sci U S A*, vol. 103, no. 7, pp. 2257–61, 2006.

[2] L. Lu and J. Li, "A combinatorial approach to determine the context-dependent role in transcriptional and posttranscriptional regulation in Arabidopsis thaliana," *BMC Syst Biol*, vol. 3, p. 43, 2009.

[3] D. P. Bartel, "MicroRNAs: Genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, 2004.

[4] S. Griffiths-Jones, H. K. Saini, S. van Dongen, and A. J. Enright, "miRBase: tools for microrna genomics," *Nucleic Acids Research*, vol. 36, pp. D154–D158, 2008, sp. Iss. SI.

[5] R. Brosh, R. Shalgi, A. Liran, G. Landan, K. Korotayev, G. H. Nguyen, E. Enerly, H. Johnsen, Y. Buganim, H. Solomon, I. Goldstein, S. Madar, N. Goldfinger, A. L. Brresen-Dale, D. Ginsberg, C. C. Harris, Y. Pilpel, M. Oren, and V. Rotter, "p53-repressed mirnas are involved with e2f in a feed-forward loop promoting proliferation," *Molecular Systems Biology*, vol. 4, 2008, 229.

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

[6] L. Wang, A. L. Oberg, Y. W. Asmann, H. Sicotte, S. K. McDonnell, S. M. Riska, W. Liu, C. J. Steer, S. Subramanian, J. M. Cunningham, J. R. Cerhan, and S. N. Thibodeau, "Genome-wide transcriptional profiling reveals microRNA-correlated genes and biological processes in human lymphoblastoid cell lines," *PLoS One*, vol. 4, no. 6, p. e5878, 2009.

[7] C. Blattner, "'junk' DNA meets the p53 network," *Mol Syst Biol*, vol. 4, p. 231, 2008.

[8] A. Marson, S. S. Levine, M. F. Cole, G. M. Frampton, T. Brambrink, S. Johnstone, M. G. Guenther, W. K. Johnston, M. Wernig, J. Newman, J. M. Calabrese, L. M. Dennis, T. L. Volkert, S. Gupta, J. Love, N. Hannett, P. A. Sharp, D. P. Bartel, R. Jaenisch, and R. A. Young, "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells," *Cell*, vol. 134, no. 3, pp. 521–33, 2008.

[9] G. Wang, Y. Wang, W. Feng, X. Wang, J. Yang, Y. Zhao, Y. Wang, and Y. Liu, "Transcription factor and microRNA regulation in androgen-dependent and-independent prostate cancer cells," *BMC genomics*, vol. 9, no. Suppl 2, p. S22, 2008.

[10] Y. Zhou, L. Chen, B. Barlogie, O. Stephens, X. Wu, D. R. Williams, M. A. Cartron, F. van Rhee, B. Nair, S. Waheed, M. Pineda-Roman, Y. Alsayed, E. Anaissie, and J. Shaughnessy, J. D., "High-risk myeloma is associated with global elevation of miRNAs and overexpression of eif2c2/ago2," *Proc Natl Acad Sci U S A*, vol. 107, no. 17, pp. 7904–9, 2010.

[11] Y. Liang, D. Ridzon, L. Wong, and C. F. Chen, "Characterization of microRNA expression profiles in normal human tissues," *Bmc Genomics*, vol. 8, 2007, 166.

[12] S. Obad, C. O. dos Santos, A. Petri, M. Heidenblad, O. Broom, C. Ruse, C. Fu, M. Lindow, J. Stenvang, E. M. Straarup, H. F. Hansen, T. Koch, D. Pappin, G. J. Hannon, and S. Kauppinen, "Silencing of microRNA families by seed-targeting tiny lnas," *Nat Genet*, vol. 43, no. 4, pp. 371–8, 2011.

[13] L. Li, J. Xu, D. Yang, X. Tan, and H. Wang, "Computational approaches for microRNA studies: a review," *Mamm Genome*, vol. 21, no. 1-2, pp. 1–12, 2010.

[14] P. Sethupathy, B. Corda, and A. G. Hatzigeorgiou, "TarBase: A comprehensive database of experimentally supported animal microRNA targets," *Rna*, vol. 12, no. 2, pp. 192–7, 2006.

[15] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, no. 1, pp. 15–20, 2005.

[16] R. Friedman, K. Farh, C. Burge, and D. Bartel, "Most mammalian mrnas are conserved targets of micrornas," *Genome research*, vol. 19, no. 1, p. 92, 2009.

[17] A. Krek, D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, and N. Rajewsky, "Combinatorial microRNA target predictions," *Nat Genet*, vol. 37, no. 5, pp. 495–500, 2005.

[18] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks, "MicroRNA targets in drosophila," *Genome Biol*, vol. 5, no. 1, p. R1, 2003.

[19] C. Barbato, I. Arisi, M. Frizzo, R. Brandi, L. Da Sacco, and A. Masotti, "Computational challenges in miRNA target predictions: to be or not to be a true target," *J Biomed Biotechnol*, vol. 803069, pp. 1–9, 2009.

[20] I. Bentwich, "Prediction and validation of microRNAs and their targets," *FEBS Lett*, vol. 579, no. 26, pp. 5904–10, 2005.

[21] E. R. Gamazon, H. K. Im, S. Duan, Y. A. Lussier, N. J. Cox, M. E. Dolan, and W. Zhang, "Exprtarget: an integrative approach to predicting human microRNA targets," *PLoS One*, vol. 5, no. 10, p. e13534, 2010.

[22] Stingo, C. Francesco, Chen, A. Yian, Vannucci, Marina, Barrier, Marianne, Mirkes, and E. G. Philip, "A bayesian graphical modeling approach to microRNA regulatory network inference," *Ann. Appl. Stat*, vol. 4, pp. 2024–2048, 2010.

[23] B. Liu, J. Li, A. Tsykin, L. Liu, A. B. Gaur, and G. J. Goodall, "Exploring complex miRNA-mRNA interactions with bayesian networks by splitting-averaging strategy," *BMC Bioinformatics*, vol. 10, p. 408, 2009.

[24] J. Li, R. Min, A. Bonner, and Z. Zhang, "A probabilistic framework to improve microRNA target prediction by incorporating proteomics data," *J Bioinform Comput Biol*, vol. 7, no. 6, pp. 955–72, 2009.

[25] J. Huang, Q. Morris, and B. Frey, "Bayesian inference of microRNA targets from sequence and expression data," *Journal of Computational Biology*, vol. 14, no. 5, pp. 550–563, 2007.

[26] J. C. Huang, T. Babak, T. W. Corson, G. Chua, S. Khan, B. L. Gallie, T. R. Hughes, B. J. Blencowe, B. J. Frey, and Q. D. Morris, "Using expression profiling data to identify human microRNA targets," *Nat Methods*, vol. 4, no. 12, pp. 1045–9, 2007.

[27] J. C. Huang, B. J. Frey, and Q. Morris, "Comparing sequence and expression for predicting microRNA targets using GenMIR3," in *Pacific Symposium on Biocomputin*, 2008, pp. 52–63.

[28] H. Attias, "A variational bayesian framework for graphical models," *Advances in neural information processing systems*, vol. 12, no. 1-2, pp. 209–215, 2000.

[29] M. Gatt, J. Zhao, M. Ebert, Y. Zhang, Z. Chu, M. Mani, R. Gazit, D. Carrasco, J. Dutta-Simmons, S. Adamia *et al.*, "Micrornas 15a/16-1 function as tumor suppressor genes in multiple myeloma." *Blood*, 2010.

[30] GO. [Online]. Available: http://www.geneontology.org

[31] F. Pichiorri, S. S. Suh, M. Ladetto, M. Kuehl, T. Palumbo, D. Drandi, C. Taccioli, N. Zanesi, H. Alder, J. P. Hagan, R. Munker, S. Volinia, M. Boccadoro, R. Garzon, A. Palumbo, R. I. Aqeilan, and C. M. Croce, "MicroRNAs regulate critical genes associated with multiple myeloma pathogenesis," *Proc Natl Acad Sci U S A*, vol. 105, no. 35, pp. 12 885–90, 2008.

[32] Y. Dai, Y. Yang, V. MacLeod, X. Yue, A. C. Rapraeger, Z. Shriver, G. Venkataraman, R. Sasisekharan, and R. D. Sanderson, "HSulf-1 and HSulf-2 are potent inhibitors of myeloma tumor growth in vivo," *J Biol Chem*, vol. 280, no. 48, pp. 40 066–73, 2005.

[33] N. Su, Y. Wang, M. Qian, and M. Deng, "Combinatorial regulation of transcription factors and microRNAs," *BMC Syst Biol*, vol. 4, p. 150, 2010.

[34] K. Tu, H. Yu, Y. J. Hua, Y. Y. Li, L. Liu, L. Xie, and Y. X. Li, "Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms," *Nucleic Acids Res*, vol. 37, no. 18, pp. 5969–80, 2009.

[35] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall, 2004.

[36] N. Sun, R. J. Carroll, and H. Zhao, "Bayesian error analysis model for reconstructing transcriptional regulatory networks," *Proc Natl Acad Sci U S A*, vol. 103, no. 21, pp. 7988–93, 2006.

[37] M. A. Tanner, *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*. Springer Verlag, 1996.

[38] GEO. [Online]. Available: http://www.ncbi.nlm.nih.gov/geo/

[39] M. Deng, Z. Tu, F. Sun, and T. Chen, "Mapping gene ontology to proteins based on protein-protein interaction data," *Bioinformatics*, vol. 20, no. 6, pp. 895–902, 2004.

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