

# Integrative approaches for microRNA target prediction: combining sequence information and the paired mRNA and miRNA expression profiles

Naifang Su<sup>†\*</sup>, Minping Qian<sup>†‡</sup> and Minghua Deng<sup>†‡§</sup>

<sup>†</sup>LMAM, School of Mathematical Sciences, Peking University, Beijing, P.R.China

\*Beijing International Center for Mathematical Research, Peking University, Beijing 100871, P.R. China

<sup>‡</sup>Center for Theoretical Biology, Peking University, Beijing, P.R.China

<sup>§</sup>Center for Statistical Sciences, Peking University, Beijing, P.R.China

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-transcriptional regulators, and play important roles 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 the sequence information, which unavoidably have high false detection rates. Recently, some novel approaches are developed to predict miRNA targets by integrating the typical algorithm with the paired expression profiles of miRNA and mRNA. Here we review and discuss these integrative approaches and propose a new algorithm called HCTarget. Applying HCTarget to the expression data in multiple myeloma, we predict target genes for ten specific miRNAs. The experimental verification and a loss of function study validate our predictions. Therefore, the integrative approach is a reliable and effective way to predict miRNA targets, and could improve our comprehensive understanding of gene regulation.

Keywords: miRNA, target prediction, expression profile, integrative analysis.

## 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. miRNAs are a novel class of endogenous ~22nt noncoding RNAs. They down regulate gene expression through the following procedures. First, the primary miRNA are transcribed from “miRNA genes” or spliced from the intronic regions of their host genes. Then the primary miRNAs produce the miRNA precursors, and the final mature miRNAs. These miRNAs are combined with Argonaute (Ago) proteins to form RNA-induced silencing complexes (RISCs). RISCs bind to the 3'-untranslated region of target mRNAs, which lead to their translational repression or degradation [2]. Hundreds of

miRNAs have been annotated in human genome, and they are predicted to regulate up to one third of all protein-coding genes [3].

Experimental analysis has recognized that miRNAs control the key cellular processes such as growth, development and apoptosis [4]. It has been established that miRNAs make an important contribution to gene regulation in embryonic development and human disease, especially cancer [5], [6], [7], [8]. Previous studies have verified that miRNAs can act as tumor suppressors or oncogenes and their dysregulation is widely involved in cancer initiation and progression [9], which enable their inhibition to be a novel therapeutic strategy for cancer [10].

An essential step and major challenge in understanding miRNA regulatory function is the identification of their target genes [11]. Since it is infeasible to carry out high throughput experiments, only a small fraction of miRNA targets have experimental supports [12], [13]. Typically, the target prediction is achieved by computational approaches based on sequence analysis. A great deal of target prediction programs have been developed [14], [15], [16], [17], [18]. Among them, TargetScan [3], [19], PicTar [20] and miRanda [21] are the most common ones. Generally, they use the following principles to recognize miRNA targets: 1) seed match: the 6-8nt seed in miRNA 5' part pair to the 3'UTR region of their target mRNA; 2) thermodynamic stability: the free energy of the miRNA-mRNA hybrid is low; 3) conservation: miRNA target sites are conserved among several species. However, these sequence based approaches have high false-positive rate. It has been demonstrated that the false positive rate of TargetScan prediction is about 22-31% [22]. Since the seed match complementation could not discern the real targets effectively, great deals of fake targets are confounded.

With the development of high throughput technology, more and more miRNA and mRNA expression profiles have been achieved to investigate miRNA's role in biological processes, especially cancer [23], [9]. Previous studies have revealed that miRNA greatly repress their target mRNAs, and mRNAs have significant expression changes after miRNA transfection or

inhibition. It has also been verified that the expression of mRNAs targeted by highly expressed miRNAs are negative shifted compared with other mRNAs. Therefore, the significantly negative correlated miRNA-mRNA pairs have high potential to be the real target pairs [24].

Based on this idea, some novel strategies have been developed to predict miRNA targets by integrative analysis. They mainly use the paired miRNA and mRNA expression data, which profile miRNA and mRNA expression levels simultaneously from the same sample, to supplement the sequence prediction for the detection of actual miRNA targets.

In this article, we review and discuss the most recent integrative approaches for miRNA target predictions. We also develop a new method called HCTarget. We apply HCTarget to the expression data in multiple myeloma and evaluate the performance of our predictions.

## II. REVIEW OF PREVIOUS APPROACHES

In the recent developed integrative approaches, there are roughly three ways to incorporate miRNA and mRNA expression profiles to the sequence prediction (Table I): 1) directly consider the correlation between miRNA and mRNA expression; 2) formulate mRNA and miRNA expression with linear model with latent variables 3) use Bayesian network to model the miRNA-mRNA regulatory network.

### A. Correlation based approach

Since miRNA generally repress their target mRNAs, a straightforward way to validate miRNA targeting mRNAs is detecting whether their expressions are inversely correlated. Based on this idea, some recent approaches have been developed to integrate the correlations between miRNA and mRNA pairs for the target predictions.

MMIA [25] (miRNA and mRNA Integrated Analysis) is an integrated miRNA and mRNA analyzing web server. It incorporates the common miRNA target prediction algorithms TargetScan, PITA and PicTar, and restricts the predictions on the significantly up (down) expressed miRNAs and the corresponding down (up) expressed mRNAs. MMIA is a feasible and simple tool for integrating miRNAs and mRNA expression profiles. However, it only takes into account the significantly up and down expression features, and loses the information of their whole expression patterns and their correlations.

X. Peng et al. [26] develop this approach by considering the inverse expression relationships between miRNAs and mRNA. They calculate the Pearson correlations between every miRNA-mRNA pair, and select the significant inverse expression pairs to construct a binary miRNA-mRNA correlation network. Meanwhile, they build a miRNA-mRNA target network based on sequence analysis. Here they relax the prediction criteria to the seed match principle, without demanding phylogenetic conservation or thermodynamic stability, to provide a larger set of candidate targets. Finally, the correlation network and the target prediction network are intersected to provide an integrative miRNA-mRNA regulatory network. This approach proposes a new point of view for miRNA target prediction,

which replaces some sequence criteria by the inverse expression relationships.

G. Huang et al. provide mirConnX [27], a web interface for inferring and displaying mRNA and miRNA regulatory network. It combines five prediction algorithms including PITA, miRANDA, TargetScan, RNAhybrid and Pictar to achieve an integrative target prediction score between each miRNA-mRNA pair. The experimental verified miRNA targets [12] are also incorporated. Meanwhile, mirConnX integrates the miRNA-mRNA expression profiles by calculating the correlations (Pearson, Spearman or Kendall) between miRNA-mRNA pairs. These correlations are converted to the probabilities of association. The target scores and the association probabilities are weighted summed to the final prediction scores, with a user defined weight. mirConnX has two innovations. First, besides Pearson correlation, it considers the non-parametric coefficients (Spearman or Kendall) and converts them to probabilities. When the sample size is small or there are outliers in the expression data, this correlation is more reliable. Second, the correlation network and the target network are weighted integrated instead of the simple intersecting.

MAGIA [28] (miRNA and genes integrated analysis) is a similar web tool for the integrative analysis. It extracts the target predictions from miRanda, PITA and TargetScan, and provide four approaches to integrate miRNA and mRNA expression profiles. 1) Similar to mirConnX, compute the Pearson or Spearman correlation coefficients between each predicted miRNA-mRNA pairs, and convert them to a false discovery rate. 2) Calculate the mutual information between a miRNA expression and a mRNA expression based on nearest neighbor distance. It could be regarded as a generalization of the Pearson correlation. 3) GenmiR++, which would be described in the following part. 4) Meta-analysis when miRNA and mRNA profiles are not paired. Users could select one or several approaches and take the intersection or union to display the combined regulatory network.

S. Bandyopadhyay et al. propose a new point of view to integrate the expression data [29]. Their approach TargetMiner is a support vector machine (SVM) classifier for miRNA target prediction. It incorporates expression profiles to construct a reliable training set. Previously, the training set are putatively extracted from experimentally verified miRNA targets (from Tarbase [12] and miRecords [13]), or sequence based predictions (from miRanda, TargetScanS, PicTar and DIANA-microT). However, the number of verified targets is pretty small, and the predictions have a significant number of false positive targets. TargetMiner propose a multi-stage filtering approach to identify the non-targets in these predictions. It first identifies tissue specific miRNAs and mRNAs by analyzing miRNA and mRNA expression profiles across several tissues, and then selects mRNA as non-targets if it is over-expressed in the same tissue with its corresponding miRNA. These candidate non-targets are further filtered by removing mRNAs with feasible miRNA-mRNA duplex stability or seed-site conservation. Combining the experimentally verified miRNA targets, TargetMiner achieve an integrative training data of

miRNA targets and non-targets. A SVM classification model is built on this data, with 30 features extracted and selected from sequence site context information. The learned SVM classifier could efficiently predict miRNA targets. Generally, TargetMinner provide an integrative training data for learning a classifier. However, it only considering the expression pattern in the training procedure, without taking them as the classification features in the SVM model.

E.Gammazon et al. develop a new approach ExprTarget [30] by combining the sequence prediction approach and the expression features in the classification. Focus on a certain miRNA, ExprTarget constructs a logistic model as:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_p x_i^p + \beta_t x_i^t + \beta_m x_i^m + \beta_e x_i^e$$

Here  $p_i$  is the probability that mRNA  $i$  is a real target.  $x_i^p$ ,  $x_i^t$  and  $x_i^m$  are the target prediction scores of mRNA  $i$  from Pictar, TargetScan and miRanda respectively.  $x_i^e$  is expression feature, defined as the p value of the general linear model between mRNA  $i$  and the miRNA. Note that if the estimated coefficient in the model is positive,  $x_i^e$  is set to 1. The coefficients  $\beta$  describes the contribution weights of different prediction algorithms. Extracting the experimental validate miRNA targets as training data,  $\beta$  could be estimated using logistic regression. By this means, ExprTarget provide the target probabilities for each mRNA. ExprTarget extends TargetMinner by incorporate the expression features in the classifier. This feature  $x_i^e$  could be regarded as a generalization of the Pearson correlation, so ExprTarget is also an extension of mirConnX, with the weights learned from experimental validate targets.

Beside the above approaches, there are some other web tools that combine miRNA-mRNA expression profiles with their target predictions. miRGator [31] integrates miRanda, PicTar and TargetScan target predictions, and displays the expression correlations between miRNA-mRNA pairs. The rank list of target mRNAs sorted by their correlations with the corresponding miRNA could also be provided. MirZ [32] incorporates smiRNadb, a database containing miRNA sequencing profiles, and the EIMMo miRNA target prediction algorithm. It also integrates mRNA expression data and allow user to restrict the target prediction to specific mRNAs that expressed in a given cell type. mimiRNA [33] integrates expression data from human miRNAs and mRNAs across multiple tissues or cell types. It groups and separates miRNA or mRNA expression data into several tissues and cell types. The paired expression data could be visualized. mimiRNA also incorporates TargetScan, miRBase, RNA22 and PicTar. User could search the targets and the inverse expressed mRNAs for a given miRNA.

In addition, when miRNA expression data are not available, HOCTAR [34] (host gene oppositely correlated targets) could be employed. It considers that most human miRNAs are intragenic and are transcribed as part of their hosting transcription units, so the expression of miRNA host genes could be used as a proxy of the expression of the miRNA itself.

TABLE I  
INTEGRATIVE APPROACH FOR miRNA TARGET PREDICTION

Name	URL	Reference
Correlation based approach		
MMIA	<a href="http://cancer.informatics.indiana.edu/mmia">http://cancer.informatics.indiana.edu/mmia</a>	[25]
Peng et al.	-	[26]
mirConnX	<a href="http://www.benoslab.pitt.edu/mirconnx">http://www.benoslab.pitt.edu/mirconnx</a>	[27]
MAGIA	<a href="http://gencomp.bio.unipd.it/magia">http://gencomp.bio.unipd.it/magia</a>	[28]
TargetMinner	<a href="http://www.isical.ac.in/bioinfo_miu/">http://www.isical.ac.in/bioinfo_miu/</a>	[29]
ExprTarget	<a href="http://www.scandb.org/apps/microna/">http://www.scandb.org/apps/microna/</a>	[30]
miRGator	<a href="http://genome.ewha.ac.kr/miRGator/">http://genome.ewha.ac.kr/miRGator/</a>	[31]
MirZ	<a href="http://www.mirz.unibas.ch">http://www.mirz.unibas.ch</a>	[32]
mimiRNA	<a href="http://mimirna.centenary.org.au">http://mimirna.centenary.org.au</a>	[33]
HOCTAR	-	[34]
Linear mode approach		
GenmiR++	<a href="http://www.psi.toronto.edu/genmir/">http://www.psi.toronto.edu/genmir/</a>	[24], [35]
F.Stingo et al.	-	[36]
J.Li et al.	-	[37]
L.Lu et al.	-	[38]
Bayesian network approach		
B.Liu et al.	-	[39]

Based on this idea, HOCTAR extracts a great deal of mRNA expression profiles and provides an average inverse correlated score between each mRNA and miRNA host gene pair. These scores are then integrated with the miRanda, TargetScan, and PicTar predictions.

### B. Linear model approach

The precious approaches only consider the pairwise expression correlation between miRNA and mRNA. However, mRNA may be regulated by multiple miRNAs and its expression is affected synthetically by all the targeting miRNAs. Based on this idea, some novel methods have been developed to model miRNA's combinatorial effect on their target mRNAs.

Among them, GenmiR++ (Generative model for miRNA regulation) is the most widely used approach [24], [35]. It 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. It has been successfully applied on the paired miRNA and mRNA expression data among multiple tissues.

Let  $y_{it}$  denote the expression level of mRNA  $i$  in tissue  $t$  and  $z_{jt}$  denote the expression level of miRNA  $j$  in the same tissue, where  $i = 1, \dots, N, j = 1, \dots, M$  and  $t = 1, \dots, T$ . GenMiR++ take a linear model to formulate the mRNA expressions and the regulatory effects of their targeting miRNAs. 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:

$$y_{it} = \mu_t - \gamma_t \sum_{j=1}^M \lambda_j r_{ij} z_{jt} + \epsilon_{it}$$

$$\text{or } \vec{y}_i = \vec{\mu} - \sum_{j=1}^M \lambda_j r_{ij} \Gamma z_j + \vec{\epsilon}_i, \vec{\epsilon}_i \sim N(0, \Sigma)$$

here  $\lambda_j$  represents the regulatory effects of miRNA  $j$ ,  $\gamma_t$  accounts for the expression scaling in tissue  $t$ , and  $\mu_t$  is the background effect of tissue  $t$ .

The latent variable  $R$  indicates the target relations between miRNA and mRNA. Integrating the target predictions  $C$  from TargetScan, as  $c_{ij} = 1$  if mRNA  $i$  is predicted to be targeted by miRNA  $j$ , and 0 otherwise, GenmiR++ assign  $R$  a Bernoulli distribution depend on  $C$ . That is  $r_{ij} \sim \text{bernoulli}(\pi)$  in the condition of  $c_{ij} = 1$ , and  $r_{ij} = 0$  when  $c_{ij} = 0$ .

Assigning the prior as  $\gamma_t \sim N(1, s^2)$  and  $\lambda_j \sim \exp(\alpha)$ , GenmiR++ use a variational Bayesian algorithm to estimate the posterior distribution of  $r_{ij}$ . Since its form is complicated, instead of learning the real posterior, the variational Bayesian algorithm provide a factorized variational posterior for approximation[40]. By this means, the computation is simplified and the target probability could be achieved.

GenMiR++ has also been developed to GenmiR3 [41], with an alternative prior distribution and the parameter  $\pi$  is modified by integrating the sequence information such as the hybridization energy and context score.

GenMiR++ has been widely used to integrate the miRNA-mRNA expression data with the target predictions. However, it has several restrictions. First, originating from 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. The variational posterior may deviated from the real posterior. Its convergence rate is highly depends on the form of the likelihood and priors and may be extremely slow.

F.Stingo et. al. [36] propose a similar linear approach. Different from GenmiR++, they don't take into account the tissue effect, and consider that miRNA has distinct regulatory effect on differet mRNAs. Based on this idea, they propose a linear model to fomulate miRNA and mRNA expressions:

$$y_i = \sum_{j=1}^M \beta_{ij} r_{ij} z_j + \epsilon_i, \epsilon_i \sim N(0, \sigma_i^2), i = 1, \dots, N$$

Here  $y_i$  is the expression of mRNA  $i$  and  $z_j$  is the expression of miRNA  $j$ .  $\beta_{ij}$  represents the effect of miRNA  $j$  on mRNA  $i$ , in GenmiR++ this term is uniformed to  $\lambda_j$ . Meanwhile, the target indicator  $r_{ij}$  is assigned with Bernoulli distribution, with a modified parameter:

$$\pi_{ij} = \frac{\exp[\eta + \tau_1 c_{ij}^1 + \tau_2 c_{ij}^2 + \tau_3 c_{ij}^3 + \tau_4 c_{ij}^4 + \tau_5 c_{ij}^5]}{1 + \exp[\eta + \tau_1 c_{ij}^1 + \tau_2 c_{ij}^2 + \tau_3 c_{ij}^3 + \tau_4 c_{ij}^4 + \tau_5 c_{ij}^5]}$$

where  $c_{ij}^1, c_{ij}^2, c_{ij}^3, c_{ij}^4$  and  $c_{ij}^5$  are the prediction scores of PicTar, miRanda, aggregate TargetScan, total TargetScan and PITA respectively.

With the prior  $\beta_{ij} \sim \text{Gamma}(1, c\sigma_i)$  and  $\sigma_i^{-1} \sim \text{Gamma}(\frac{\delta+M}{2}, \frac{d}{2})$ , the posterior distribution could be estimated using Metropolis-Hasting algorithm. Thus the posterior target scores  $p(r_{ij} = 1 | \text{data})$  are achieved to construct miRNA regulatory network.

However, since  $\beta$  are distinct for different miRNAs and mRNAs, the model has a great deal of parameters. Therefore, this approach is limited in high computational complexity.

J.Li et. al. [37] also modify the model. They discretize mRNA expression to binary value  $y_{it} = 1$  or 0, which represent high or low expressions, then assume  $y_{it}$  follow a logistic model: let  $q_{it} = P(y_{it} = 1)$ ,

$$\log\left(\frac{q_{it}}{1 - q_{it}}\right) = \sum_{j=1}^M \phi_j r_{ij} z_{jt} + \epsilon_t, i = 1, \dots, N$$

Similar to GenmiR++,  $r_{ij}$  follow a bernoulli distribution depend on the TargetScan prediction  $c_{ij}$  with parameter  $\pi$ .

With the prior  $\phi_j \sim \exp(\phi)$ ,  $\phi \sim U(0, \infty)$ ,  $\epsilon_t \sim U(-50, 50)$  and  $\phi \sim \text{beta}(1, 1)$ , the posterior could be estimated using Gibbs sampling. They also apply the similar approach to study the relation between miRNA expression and protein abundance.

In this approach, the binary mRNA value lose the information of the whole expression profile.

The above approaches use Bayesian methodology for parameter estimation. On the other hand, Y. Lu et. al.[38] incorporate a lasso regression model to predict miRNA targets. Moreover, they pay attention to the role of RISCs and assume that mRNA expression follow a linear model with its targeting RISCs. The RISC level could be obtained through the expression of its comprising miRNA and Ago proteins. There are four Ago proteins in human, and Ago2 is the essential one. Therefore, the model is:

$$y_i = \beta_{i0} + \sum_{j=1}^M c_{ij} \beta_{ij} z_j \text{Ago}_2 + \sum_{j=1}^M c_{ij} \phi_{ij} z_j \text{Ago}_{134} + \epsilon_i$$

Here  $y_i$  is the expression of mRNA  $i$ ,  $z_j$  is the expression of miRNA  $j$ ,  $\text{Ago}_2$  is the expression of Ago2 mRNA and  $\text{Ago}_{134}$  is the combined expression of Ago1, Ago3 and Ago4 mRNA.  $c_{ij}$  indicates the target prediction relation from TargetScan and PicTar.

Then a multi-run lasso regression procedure is produced, and miRNAs are ranked by their estimated coefficients. With these ranked scores, the targeting miRNAs could be achieved.

However, this approach produce lasso regression for each mRNA separately. It will be time consuming when applying to a great deal of mRNAs.

### C. Bayesian network approach

Beside the linear model approach, some novel studies are developed to model the whole miRNA-mRNA regulatory network. Bayesian network, a probabilistic graphical model, has been widely used to discover the structure of gene networks [42]. It could also be applied to study the regulation between miRNA and mRNA [43]. Liu et al. [39] develop a new

approach which use Bayesian network to learn the miRNA-mRNA regulatory network by integrating miRNA target prediction and expression profiles.

Denote miRNA and mRNA as nodes and their target relations as directed edges, the regulatory network could be modeled as a discrete Bayesian network  $G$ . The miRNA and mRNA expressions  $X$  are discretized to binary values 1 and 0, indicating high and low expressions. Let  $N_{ijk}$  be the observed times that mRNA  $X_i$  is in state  $k$  ( $k = 1, \dots, r_i$ , here  $r_i = 2$ ) with its parent miRNAs in state  $j$  ( $j = 1, \dots, q_i$ ), then  $X$  follow a multinomial distribution with parameter  $\theta_{ijk} = P(X_i = k | \text{parent}(X_i) = j)$ :

$$p(X|\theta, G) \propto \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}}$$

Assigning the Dirichlet prior to  $\theta$  as

$$p(\theta|G) \propto \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1}$$

the Bayesian score of the network  $P(X|G)$  is given by [44]:

$$P(X|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(N_{ij} + \alpha_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk} + \alpha_{ijk})}{\Gamma(\alpha_{ijk})}$$

Here  $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ ,  $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$  and  $\alpha_{ijk} = N/r_i q_i$ ,  $N$  is the sample size.

Network with the maximum score is selected as the learned Bayesian network, which is putatively achieved by exhaustive searching algorithm such as hill climbing. The searching space could be reduced by constraining the target relations within miRBase, PicTar and TargetScan predictions. By this means, Liu et al. analyze miRNA-mRNA expression profiles from multiple cell types and build Bayesian network for each cell type. These networks are then integrated to provide the significant miRNA-mRNA target relations.

Bayesian network is a reliable and accurate model for the regulatory network [42]. However, its learning algorithm has high computational complexity and is time consuming. Therefore, Bayesian network could not be applied to learn large-scale networks.

### III. HCTARGET METHOD

Based on the above discussion, we propose a new algorithm called HCTarget (High Confident targets) to integrate expression and sequence information to detect miRNA targets. Our approach extends GenMiR++ and overcomes its restrictions in the following two ways. First, GenMiR++ characterizes miRNA's relative effects among all tissues as a constant. We improved this constrain by re-defining the parameters of miRNA effects. Second, GenMiR++ uses variational Bayesian algorithm to approximate of the real posterior. Its convergence rate may be slow and the estimation is not stable. We use a classical Markov chain Monte Carlo (MCMC) algorithm to learn the posterior directly.

#### A. Model

Incorporating the notations in GenmiR++, we propose a linear model to formulate the relations between mRNA expressions and the regulatory effects of their targeting miRNAs as:

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

Here  $\beta_{jt}$  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  $\beta_{0t}$  is the background effect of sample  $t$ .

The goal of our model is to estimate the latent indicators  $R$ . Similarly, it follow a Bernouli distribution depend on the sequence prediction  $C$ . 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  $\pi$  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  $B_t = (r_{ij} z_{jt})$ ,  $A_t = [1, B_t]$ ,  $y_t = (y_{1t}, \dots, y_{Nt})^T$ ,  $Z = (z_{jt})$ ,  $\beta_t = (\beta_{0t}, \dots, \beta_{Mt})^T$  and  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})^T$ , we have the vector representation of our model:

$$Y_t = A_t \beta_t + \epsilon_t$$

#### 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:

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

To estimate the parameters  $\theta = (\beta, \sigma^2, \pi)$  and latent variables  $R$ , we apply the Bayesian methodology and a MCMC algorithm [45]. With proper prior assumptions, the posterior of  $R$  and  $\theta$  have simple forms and could be directly computed using a MCMC algorithm as the following iterations [46], [47]: (i) sample the parameters  $\theta$  conditional on the updated latent variable; (ii) sample the latent variable  $R$  conditional on the updated parameters .

##### 1) Update the parameters:

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

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

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  $\pi$ , with the conjugate prior  $\pi \sim \text{Beta}(a_0, b_0)$ , the posterior distribution is

$$\pi \sim \text{Beta}(n_1 + a_0, n_0 + b_0) \quad (2)$$

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

### 2) Update the latent variable:

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

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

Since

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

here  $q_{ijt}$  denotes

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

The first term doesn't contain  $r_{ij}$ , so

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

that is,  $r_{ij}$  has Bernoulli marginal distribution

$$p(r_{ij}|\cdot) \sim \text{bernoulli}(p_{ij}) \quad (3)$$

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

### 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  $\beta_t, \sigma_t, R$  as  $\beta_t = 1, \sigma_t = 1$  and  $r_{ij}|c_{ij} = 1 \sim \text{bernoulli}(0.5)$ .
- 2) Update  $\sigma_t^2$  by sampling from  $vs_t^2 \chi_v^{-2}$ , update  $\beta_t$  by sampling from  $N(\hat{\beta}_t, (A_t^T A_t)^{-1} \sigma_t^2)$  and update  $\pi$  by sampling from  $\text{beta}(n_1 + a_0, n_0 + b_0)$ .
- 3) Given the updated parameters, sample the latent variable  $r_{ij}$  from  $\text{bernoulli}(p_{ij})$ .
- 4) Repeat the above two steps until convergence. Here the convergence is evaluated by Gelman and Rubin criteria [47].

We output  $p_{ij}$ , which represents the probability that miRNA  $j$  targets mRNA  $i$  given the data, 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, so that our selected miRNA targets covered nearly 50% of the sequence-based predictions, and they are comparable with GenMiR++ targets.

## IV. RESULTS

We applied HCTarget to study miRNA's role in cancer. The computational predictions were extracted from TargetScanHuman (release 5.1). Several paired miRNA-mRNA expression datasets, such as breast cancer data (GSE19783), prostate cancer data (GSE7055) and multiple myeloma data (GSE17306) were downloaded from GEO database [48]. Since their results are similar, we took the multiple myeloma data as an example

in our analysis. It profiled miRNA and mRNA expressions from 52 patients with multiple myeloma [9].

We selected multiple myeloma related miRNAs and mRNAs for our predictions. Ten miRNAs with the highest expression level were 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 were 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 [49].

### A. Performance of HCTarget on the simulation data

First, we generated a simulation data to compare the performance of GenMiR++ and HCTarget. The ten miRNA expression data  $Z$  were extracted from the real data from patients with multiple myeloma, where 1000 mRNA expressions  $Y$  were simulated from

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

here  $\beta_{jt}, \beta_{0t}$  and  $\epsilon$  were generated from  $N(-0.3, 0.1)$ ,  $N(1, 1)$  and  $N(0, 1)$  respectively. The real target relations  $r_{ij}$  was obtained from  $\text{bernoulli}(0.5)$  conditions on  $c_{ij} = 1$ , where  $c_{ij}$  represents the predictions in TargetScan.

Applying GenmiR++ and HCTarget on the simulation data, we computed their true positive rate and false positive rate 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++.

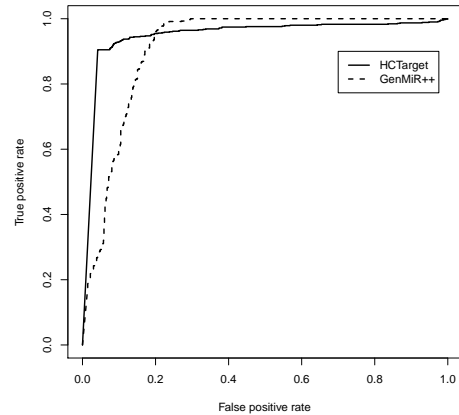


Fig. 1. The ROC curves of HCTarget and GenMiR++ for simulation data. Their AUC values are 0.95 and 0.91 respectively.

### B. Predict miRNA targets based on cancer expression data

We then applied our HCTarget approach to the real miRNA-mRNA expression data to detect miRNA targets in cancer. TargetScan provides 1401 target pairs for our selected miRNAs

and genes. HCTarget cuts down these predictions to 647, while 699 target pairs are obtained by GenMiR++.

To assess the robustness of HCTarget, we performed a series of permutation tests [24]. We permuted the gene labels 1000 times and generated 1000 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 0.1. In addition, the proportion of the probabilities bigger than 0.8 for the real data (46.2%) is higher than permuted data (44.1%). It illustrates that HCTarget could successfully discriminate the real target from the fake ones, which ensures its robustness in target prediction.

Furthermore, we extracted experimentally supported miRNA targets from Tarbase (v.5c) [12] to evaluate the accuracy of our approach. To compare Tarbase with our predictions, miRNAs were all mapped to miRNA families using the annotations in miRBase [50]. For the multiple myeloma related miRNAs and mRNAs, three miRNAs and their 17 target genes have biological 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 II, 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 of them. It also indicates that HCTarget has higher precision (2.78%, 18 out of 647) than the original TargetScan (2.43%, 34 out of 1401).

TABLE II  
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%)

### C. Validate hsa-miR-16 targets

Previous analysis suggests that hsa-miR-16 can act as a tumor suppressor in multiple myeloma [51]. We extracted a loss of function study profile of hsa-miR-16 from GEO database (GSE24522). It provided gene expression levels before and after hsa-miR-16 deletion [51]. We focused on genes with fold change larger than 1.5 as different expressed genes. For our 1000 genes, 132 genes were selected.

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 hypergeometric 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.

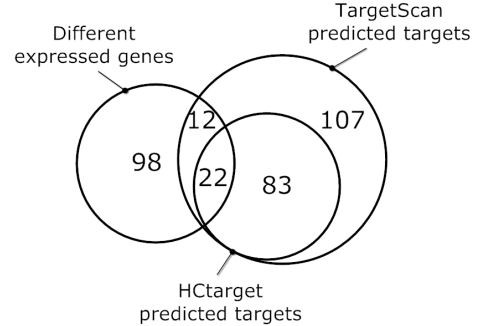


Fig. 2. **Venn diagram.** It shows the overlap of different expressed genes with the predicted targets of targetScan and HCTarget.

### D. Gene Ontology enrichment analysis

To have further investigation of our predicted targets, we analyzed their function annotations in Gene Ontology (GO)[52], [53]. For each miRNA 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 were 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 [54]. 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.

### E. 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 was selected, and one of its targets detected by HCTarget is SULF1, which has been found to be a potent inhibitor of myeloma tumor growth [55]. 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 believe that this target pair plays a crucial role in multiple myeloma and could be served as effective candidates for the therapeutic treatment.

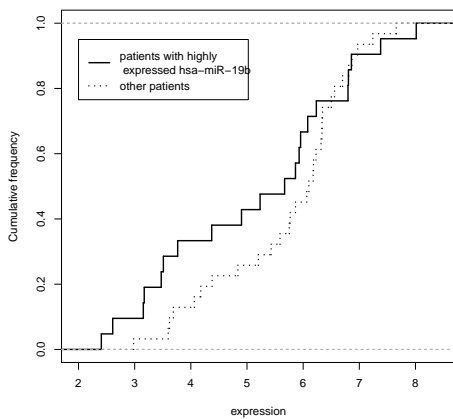


Fig. 3. **The down regulatory effect of hsa-miR-19b on SULF1.** The cumulative distributions of the expression levels of SULF1 in the sample with or without highly expressed hsa-miR-19b (red solid line and blue dashed line respectively).

## V. CONCLUSION

In this paper, we review and discuss the integrative approach that predict miRNA target genes by combining the sequence information and expression profiles.

We also propose a new algorithm, HCtarget. 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. Furthermore, selecting some other proper miRNA and mRNA expression profiles, HCtarget could be generalized to provide miRNA's whole genome target predictions, which is helpful for the comprehensive discovering of miRNA's regulatory effects.

Generally, the integrative approaches improve miRNA target predictions. They 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 [56]. With accurate

target predictions of miRNAs and TF, the integrative approaches could effectively construct gene regulatory network, which helps us to uncover the mechanisms underlying gene expression.

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.31171262, No.11021463), 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] D. P. Bartel, "MicroRNAs: Genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, 2004.
- [3] 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.
- [4] 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.
- [5] 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, A. 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.
- [6] G. Calin and C. Croce, "MicroRNA signatures in human cancers," *Nature Reviews Cancer*, vol. 6, no. 11, pp. 857–866, 2006.
- [7] A. Care, D. Catalucci, F. Felicetti, D. Bonci, A. Addario, P. Gallo, M. Bang, P. Segnalini, Y. Gu, N. Dalton *et al.*, "MicroRNA-133 controls cardiac hypertrophy," *Nature medicine*, vol. 13, no. 5, pp. 613–618, 2007.
- [8] X. Wang, S. Tang, S. Le, R. Lu, J. Rader, C. Meyers, and Z. Zheng, "Aberrant expression of oncogenic and tumor-suppressive microRNAs in cervical cancer is required for cancer cell growth," *PLoS One*, vol. 3, no. 7, p. e2557, 2008.
- [9] 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.
- [10] 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 Inas," *Nat Genet*, vol. 43, no. 4, pp. 371–8, 2011.
- [11] 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.
- [12] 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.
- [13] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li, "mirecords: an integrated resource for microRNA-target interactions," *Nucleic acids research*, vol. 37, no. suppl 1, pp. D105–D110, 2009.
- [14] 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.
- [15] D. Bartel, "MicroRNAs: target recognition and regulatory functions," *Cell*, vol. 136, no. 2, pp. 215–233, 2009.



- [16] H. Min and S. Yoon, "Got target?: computational methods for microRNA target prediction and their extension," *Experimental & molecular medicine*, vol. 42, no. 4, p. 233, 2010.
- [17] T. Saito and P. Sætrom, "MicroRNAs—targeting and target prediction," *New biotechnology*, vol. 27, no. 3, pp. 243–249, 2010.
- [18] Y. Dai and X. Zhou, "Computational methods for the identification of microRNA targets," *Open Access Bioinformatics*, vol. 2, pp. 29–39, 2010.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] I. Bentwich, "Prediction and validation of microRNAs and their targets," *FEBS Lett*, vol. 579, no. 26, pp. 5904–10, 2005.
- [23] 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.
- [24] 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.
- [25] S. Nam, M. Li, K. Choi, C. Balch, S. Kim, and K. Nephew, "MicroRNA and mRNA integrated analysis (mmia): a web tool for examining biological functions of microRNA expression," *Nucleic acids research*, vol. 37, no. suppl 2, pp. W356–W362, 2009.
- [26] X. Peng, Y. Li, K. Walters, E. Rosenzweig, S. Lederer, L. Aicher, S. Proll, and M. Katze, "Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers," *Bmc Genomics*, vol. 10, no. 1, p. 373, 2009.
- [27] G. Huang, C. Athanassiou, and P. Benos, "mirconnx: condition-specific mRNA-microRNA network integrator," *Nucleic acids research*, vol. 39, no. suppl 2, p. W416, 2011.
- [28] G. Sales, A. Coppe, A. Bisognin, M. Biasiolo, S. Bortoluzzi, and C. Romualdi, "Magia, a web-based tool for miRNA and genes integrated analysis," *Nucleic acids research*, vol. 38, no. suppl 2, pp. W352–W359, 2010.
- [29] S. Bandyopadhyay and R. Mitra, "Targetminer: microRNA target prediction with systematic identification of tissue-specific negative examples," *Bioinformatics*, vol. 25, no. 20, p. 2625, 2009.
- [30] E. R. Gamazon, H. K. Im, S. Duan, Y. A. Lussier, N. J. Cox, M. E. Dolan, and W. Zhang, "Exptarget: an integrative approach to predicting human microRNA targets," *PLoS One*, vol. 5, no. 10, p. e13534, 2010.
- [31] S. Nam, B. Kim, S. Shin, and S. Lee, "mirgator: an integrated system for functional annotation of microRNAs," *Nucleic Acids Research*, vol. 36, no. suppl 1, pp. D159–D164, 2008.
- [32] J. Hausser, P. Berninger, C. Rodak, Y. Jantscher, S. Wirth, and M. Zavolan, "Mirz: an integrated microRNA expression atlas and target prediction resource," *Nucleic acids research*, vol. 37, no. suppl 2, pp. W266–W272, 2009.
- [33] W. Ritchie, S. Flamant, and J. Rasko, "mimirna: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets," *Bioinformatics*, vol. 26, no. 2, pp. 223–227, 2010.
- [34] V. Gennarino, M. Sardiello, R. Avellino, N. Meola, V. Maselli, S. Anand, L. Cutillo, A. Ballabio, and S. Banfi, "MicroRNA target prediction by expression analysis of host genes," *Genome research*, vol. 19, no. 3, pp. 481–490, 2009.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] Y. Lu, Y. Zhou, W. Qu, M. Deng, and C. Zhang, "A lasso regression model for the construction of microRNA-target regulatory networks," *Bioinformatics*, vol. 27, no. 17, pp. 2406–2413, 2011.
- [39] 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.
- [40] H. Attias, "A variational bayesian framework for graphical models," *Advances in neural information processing systems*, vol. 12, no. 1-2, pp. 209–215, 2000.
- [41] 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.
- [42] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using bayesian networks to analyze expression data," *Journal of computational biology*, vol. 7, no. 3-4, pp. 601–620, 2000.
- [43] 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.
- [44] R. Neapolitan, *Learning bayesian networks*. Pearson Prentice Hall Upper Saddle River, NJ, 2004.
- [45] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall, 2004.
- [46] 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.
- [47] M. A. Tanner, *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*. Springer Verlag, 1996.
- [48] NCBI. [Online]. Available: <http://www.ncbi.nlm.nih.gov/geo/>
- [49] 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.
- [50] 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.
- [51] 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.
- [52] GO. [Online]. Available: <http://www.geneontology.org>
- [53] 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.
- [54] 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.
- [55] 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.
- [56] N. Su, Y. Wang, M. Qian, and M. Deng, "Combinatorial regulation of transcription factors and microRNAs," *BMC Syst Biol*, vol. 4, p. 150, 2010.