Bioinformatics Advance Access published September 10, 2015

Systems Biology

Inferred miRNA activity identifies miRNA-mediated regulatory networks underlying multiple cancers

Eunjee Lee[1,2], Koichi Ito[3], Yong Zhao[1], Eric E. Schadt[1,2,4], Hanna Y. Irie[3,4], and Jun Zhu[1,2,4,*]

[1]Department of Genetics and Genomic Sciences; [2]Icahn Institute of Genomics and Multiscale Biology; [3]Department of Medicine, Hematology and Medical Oncology; [4]The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029

Associate Editor: Dr. Janet Kelso

ABSTRACT

Motivation: MicroRNAs (miRNAs) play a key role in regulating tumor progression and metastasis. Identifying key miRNAs, defined by their functional activities, can provide a deeper understanding of biology of miRNAs in cancer. However, miRNA expression level can't accurately reflect miRNA activity.

Results: We developed a computational approach, *ActMiR*, for identifying active miRNAs and miRNA-mediated regulatory mechanisms. Applying ActMiR to four cancer datasets in The Cancer Genome Atlas (TCGA), we showed that (1) miRNA activity was tumor subtype specific; (2) genes correlated with inferred miRNA activities were more likely to enrich for miRNA binding motifs; (3) expression levels of these genes and inferred miRNA activities were more likely to be negatively correlated. For the four cancer types in TCGA we identified 77~229 key miRNAs for each cancer subtype and annotated their biological functions. The miRNA-target pairs, predicted by our ActMiR algorithm but not by correlation of miRNA expression levels, were experimentally validated. The functional activities of key miRNAs were further demonstrated to be associated with clinical outcomes for other cancer types using independent datasets. For $ER^-/HER2^-$ breast cancers, we identified activities of key miRNAs let-7d and miR-18a as potential prognostic markers and validated them in two independent $ER^-/HER2^-$ breast cancer data sets. Our work provides a novel scheme to facilitate our understanding of miRNA. In summary, inferred activity of key miRNA provided a functional link to its mediated regulatory network, and can be used to robustly predict patient's survival.

Availability: the software is freely available at http://research.mssm.edu/integrative-network-biology/Software.html.

1 INTRODUCTION

MicroRNAs (miRNAs) have been shown to control cell growth, differentiation and apoptosis; consequently, impaired miRNA expression has been implicated in tumorigenesis (Iorio, et al., 2005; Jansson and Lund, 2012). In recent years, miRNA expression signatures have been used to classify cancers and to predict favorable prognosis including breast cancers (Calin, et al., 2005; Iorio, et al., 2005). However, these studies did not provide direct mechanistic links between miRNAs and their mRNA targets, therefore, whether and how the identified miRNAs play key regulatory roles in post-transcriptional regulation is not clear.

A conventional way for identifying key regulatory miRNAs, which potentially regulate expression levels of a large number of genes, is to explore miRNA-mRNA relationships based on their expression levels. By integration of mRNA and miRNA expression levels, key cellular pathways related with miRNA signatures are shown to associate with cancer progression (Dvinge, et al., 2013). However, the association between miRNA and mRNAs does not imply that the miRNA causally regulates these genes. It may be due to pleiotropic effect of upstream regulators. Furthermore, miRNA expression level is not equivalent to its functional activity (Mullokandov, et al., 2012). There are proteins or RNAs that can mediate the influence of miRNAs on target genes, such as RISC complex (Krol, et al., 2010). The relative abundance of miRNAs to its target genes determines the functional activity levels of miRNAs (Ebert, et al., 2007). Thus, to infer key miRNAs and understand their gene regulatory networks, it is critical to accurately quantify the regulatory activity of miRNAs by considering their effectiveness on target genes

By applying systematic computational models integrated with prior information on the regulatory sequence of miRNAs, multiple miRNA-mediated pathways have been identified. For example, key miRNAs are shown to affect genes' expression levels using an integrative model (Setty, et al., 2012). Different approaches for identifying key miRNAs have been proposed such as testing miRNAs binding sites enrichment among sets of co-expressed genes (Gennarino, et al., 2012). Some studies infer the miRNA activity using mRNA expression data of genes enriched for miRNA target sites (Arora and Simpson, 2008; Madden, et al., 2010) or average difference in expression levels of miRNA targets versus non-targets (Cheng, et al., 2009). However, these methods do not explicitly consider the role of miRNAs as a *post-transcriptional* regulator. The amount of expression level variation attributed to miRNA regulation corresponds to the changes from the total transcribed mRNA level (or baseline mRNA expression level). The model to consider pre-processed mRNA level is essential to accurately infer miRNA activities.

Here we present a novel computational method **ActMiR** for explicitly inferring the activity of miRNAs based on the changes in the expression levels of target genes due to post-transcriptional regulation. The inferred miRNA activity is further used to identify key miRNAs that regulate expression levels of a large number of genes and may drive tumor progression. Key miRNAs identified from its inferred activity can be used as biomarkers for predicting

*To whom correspondence should be addressed.

prognosis or as therapeutic targets for cancer treatment. Our ActMiR method for inferring miRNA activity is based on two assumptions about miRNA function: (i) the baseline expression levels of target genes (when a miRNA has no impact) is approximated by their levels in samples with low expression level of the corresponding miRNA (as sufficient miRNA concentration is essential for its function); (ii) the suppression of target genes by a miRNA depends on the expression levels of the mRNAs (Doench and Sharp, 2004).

We applied ActMiR to four cancer types in The Cancer Genome Atlas (TCGA). ActMiR identified 77~229 key miRNAs that play a key role in transcriptional regulations for each cancer subtype. We demonstrated that our inferred miRNA activity could be further used for identifying functional target genes and as prognostic biomarkers. Notably, we identified let-7d and let-7g as key tumor suppressors in $ER^-/HER2^-$ breast cancers, targeting on the expression level of collagen and extracellular matrix genes. The predicted putative causal relationships between let-7d and let-7g and target genes based on the inferred activity were experimentally validated. Additionally, the predicted activities of let-7d and let-7g were associated with survival. We further demonstrated in two independent cohorts that predicted activities of two key miRNAs (i.e. let-7d and miR-18a) in $ER^-/HER2^-$ breast cancers were consistently associated with survival, while none of expression levels of these key miRNAs was significantly associated with survival. The phenomenon that only miRNA activity predicted by our ActMiR method was robustly associated with patient survival was observed in multiple cancer types. Taken together, our proposed ActMiR method does not only identify key miRNAs, but also provides direct mechanism links between miRNAs and their functional mRNA targets which in turn act together to affect clinical phenotypes.

2 METHODS

2.1 The Cancer Genome Atlas (TCGA) data

To identify key miRNAs that regulate a large subset of transcriptome and in turn affect tumor progression, we integrated mRNA expression data with miRNA expression data for four cancer types in TCGA: Breast invasive carcinoma (BRCA), Colon adenocarcinoma (COAD), Glioblastoma (GBM), and Ovarian serous cystadenocarcinoma (OV). We selected these four cancer types because there were publically available independent cohort data sets consisting of both mRNA and miRNA expression profiles as validation sets. We further considered subtypes of each cancer that was previously classified by gene expression levels or clinical features (Brennan, et al., 2013; Cancer Genome Atlas, 2012; Verhaak, et al., 2010). For GBM and COAD, we classified tumor samples based on gene expression levels. For BRCA, the samples were classified into ER^+ and $ER^-/HER2^-$ subtypes based on its clinical features (**Supplementary Table 1**). See details in the Supplementary Methods section.

2.2 ActMiR for inferring activity of miRNAs from expression levels of miRNA and its target genes

We developed ActMiR, a method for inferring miRNA activity based on expression levels of miRNAs and their predicted target genes. **Figure 1** presents an overview of ActMiR for inferring miRNA activity. Three pieces of information were used: (i) miRNA expression levels of samples; (ii) mRNA expression levels of samples; (iii) the predicted target lists of each miRNA. For the predicted target list of miRNAs, we used a collection of predicted target genes for 1537 unique mature miRNAs from TargetScan (www.targetscan.org) that considers all conserved miRNA binding sites inherited from 23-way alignments of UTR sequences (Grimson, et al., 2007). We filtered out miRNAs whose number of target genes is smaller than 10. We further focused on miRNAs whose predicted target genes' expression levels and their own expression levels were available.

The ActMiR method consists of three steps. First, for each miRNA, we estimated the "baseline" expression levels of miRNA's target genes at the state where the miRNA had no impact. As sufficient miRNA concentration is essential for its functional activity, we defined baseline expression level of the target gene t of miRNA φ as $y^b_{\varphi,t} = E(y_{\varphi,t} \mid \mathrm{e}(\varphi) \to 0)$, which is the average expression level of the samples with low miRNA expression level (**Figure 1A**) (see Supplementary Methods for estimating the baseline expression level). Next, we defined the "degradation" levels as the difference between the observed expression levels of targeted genes for each sample, which is affected by the miRNA, and the baseline expression level, which is unaffected by the miRNA. For each sample S, degradation levels $y^d_{\varphi,t,s}$ of predicted target t of miRNAs φ is determined as follows: $y^d_{\varphi,t,s} = y^b_{\varphi,t} - y_{t,s}$, where $y_{t,s}$ is the observed expression level of the predicted target. The expression degradation level allows us to measure how much expression level change of the target gene is potentially affected by each miRNA. Finally, based on the assumption that the impact of a miRNA on its target genes depends on its expression level (Doench and Sharp, 2004), we used a linear model representing the relationship between the degradation levels and baseline expression levels of target genes for each sample, $Y^d_{\varphi,t\in T,s} = \alpha_{\varphi,s} Y^b_{\varphi,t\in T,s}$ in which the coefficient $\alpha_{\varphi,s}$ represents the activity of miRNA φ in sample S, where T is the collective set of all targets of miRNA φ (**Figure 1B**). Not all predicted target genes with seed sequences are functionally regulated by miRNAs (Wu et al. 2015). To take account for the probability of a predicted target gene being a functional target, we used an iteratively reweighted least squares (IRLS) regression method to estimate $\alpha_{\varphi,s}$, assuming that the higher anti-correlation between miRNA activity and a gene's expression level across samples indicates the higher possibility of being a functional target (shown in **Supplementary Figure S1** and Supplementary Methods).

2.3 Identifying key miRNAs and their functional target genes

A key miRNA was defined as the miRNAs whose activity is *significantly associated with a large number* of mRNAs' expression levels. To determine key miRNAs, we first computed associations between miRNA activity and mRNAs' expression levels using the Pearson correlation. The significance of associations between activity of miRNA and mRNA's expression level was assessed by permutation tests (see Supplementary Methods for details). Secondly, we counted the number of genes whose expression levels were significantly correlated with each miRNA activity. We determined the threshold of number of genes correlated to each miRNA that was significantly larger than numbers from the permuted data sets.

For each key miRNA, we examined the functional target genes, whose expression levels were correlated with miRNA activity, among predicted target genes of the miRNA based on TargetScan (Grimson, et al., 2007). We further annotated function of key miRNAs by comparing their direct target genes with 1320 canonical pathways from BioCarta, KEGG and Reactome (Subramanian, et al., 2005), identified biological pathways over-represented in the functional target gene set of each miRNA using the Fisher's exact test. The FDR was computed based on the 50 randomized data sets, as the ratio of the average number of pathways with a p-value below threshold across permuted data sets, and the number of pathway with p-value below threshold.

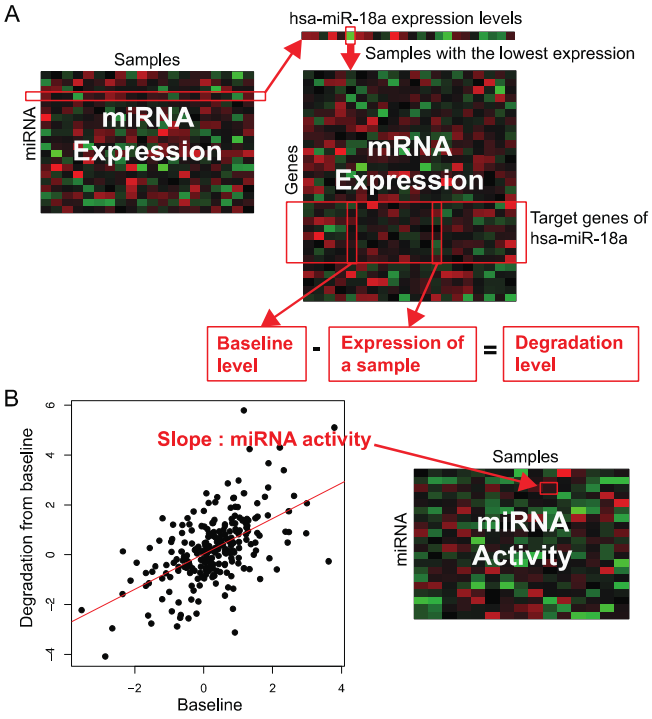

Figure 1. The overview of our ActMiR procedure. (A) The miRNA and mRNA expression matrix are used to infer the regulatory activity of miRNA for each sample and for each miRNA. For each miRNA, we estimated the baseline expression levels of target genes from their expression levels within samples with the lowest miRNA expression level. For each sample, the degradation levels of target genes were computed as the difference between expression levels in the sample and baseline expression levels. (B) For each sample independently, genome-wide linear regression of baseline expression levels on degradation levels was performed. The coefficient from this linear fitting, estimated using an IRLS regression (shown in Supplementary Figure S1), represents miRNA activity.

2.4 Validating functional target genes of key miRNAs.

HCC1187 cells (obtained from ATCC) were cultured in RPMI1640 media supplemented with 10% fetal bovine serum and penicillin/streptomycin. Forward transfection of mirVana inhibitors against miR-18a-5p, let-7d-5p, or let-7g-5p was performed. After 72h incubation, total RNA was extracted. For cDNA synthesis reaction, 1 μg of the total RNA was used with TaqMan Reverse Transcription Reagents. Quantitative PCR (qPCR) was performed by using Power SYBR Green Master Mix. Details were shown in Supplementary Materials and **Supplementary Table S2**.

2.5 Identifying prognostic key miRNAs based on their activities

To further explore functional relevance of key miRNAs, we identified prognostic key miRNAs. We tested whether there was association between overall survival time and miRNA activities for each cancer type in TCGA using a univariate Cox regression model. Additionally, survival prognosis by individual miRNAs was tested using a multivariate Cox proportional hazard model including covariates such as age at diagnosis, lymph-node status, tumor grade if available. To determine the statistical significance, we randomly permutated the activity for each miRNA for 1000 times, then used the resulting empirical null distribution to compute a FDR. For comparison, we applied the same procedure based on the expression of miRNA instead of the activity of miRNA.

2.6 Validating prognostic key miRNAs in independent data sets

We further validated our predicted prognostic key miRNAs in five independent data sets consisting of both miRNA and mRNA expression profiles, including two data sets for BRCA and one for each remaining cancer type. All independent data sets we used were summarized in **Supplementary Table S3**. For GBM and COAD, we clustered samples based on gene expression levels, and for BRCA, we used clinical information to classify samples (Details are shown in Supplementary Methods). For each data set, we inferred miRNA activity from miRNAs and mRNAs expression levels using the procedure described above, and then tested association between the inferred activity of each key miRNA and overall survival to identify prognostic key miRNAs. For Buffa *et al.*'s breast cancer data set (Buffa, et al., 2011), because the overall survival is not available, we used distant relapse free survival time instead.

3 RESULTS

3.1 miRNA-mRNA correlations were cancer subtype specific.

Due to its molecular and therapeutic heterogeneity, each cancer type is generally categorized into subtypes according to gene expression patterns or clinical features (Cancer Genome Atlas, 2012; Cancer Genome Atlas, 2012; Gruvberger, et al., 2001; Verhaak, et al., 2010), and each subtype is of significantly distinct molecular portrait, response to therapies, and survival rate (Cancer Genome Atlas, 2012; Verhaak, et al., 2010). The miRNA-mRNA correlation structure for each cancer subtype was different (**Supplementary Figures S2A, S3A, and S4A**). In particular, strong miRNA-mRNA associations based on all samples might be due to miRNA and mRNA expression level differences among cancer subtypes instead of true association between miRNA and mRNA (**Supplementary Figures S2B, S3B, and S4B**). Additionally, subtype-specific miRNA-mRNA relationships (see purple in **Supplementary Figures S2A, S3A, and S4A**) might be missed when considering all samples as a whole. These results suggested that miRNA regulatory mechanisms were subtype specific and each subtype should be studied individually, consistent with recent reports that transcriptional regulations of target genes by miRNAs were cancer subtype specifics (Farazi, et al., 2014; Pecot, et al., 2013; Song, et al., 2013).

Therefore, for GBM and COAD, we classified tumor samples based on previous classification (Brennan, et al., 2013; Cancer Genome Atlas, 2012; Verhaak, et al., 2010) (**Supplementary Table 1**). For BRCA, expression levels of many miRNAs (i.e. 211 out of 682 miRNAs tested) were strongly associated with ER status (**Supplementary Figure S2C**). Among ER^- tumors, 21 samples were $HER2^+$, which was another important clinical feature to determine effective therapies. Expression levels of 18 miRNA were strongly associated with HER2 status (**Supplementary Figure S2D**). Therefore, we classified samples into ER^+ and $ER^-/HER2^-$ subtypes. We focused on each subtype in all further analyses. All subtypes used were listed in **Supplementary Table 1**.

3.2 Expression level of miRNAs was not sufficient for identifying causal relationship with target genes

We first evaluated relationships between miRNA and mRNAs expression levels in individual cancer subtypes using Pearson correlations. For the most subtypes of cancers, the distributions of interconnectivity between miRNA and mRNA followed a scale-free distribution (**Supplementary Figure S5**), which is the common property of biological networks (Chen, et al., 2008). This observation suggested that there were a small number of miRNAs associating with a large number of genes' expression levels, denoted as *hub miRNAs*. Hub miRNAs were not equivalent to *key miRNA* regulators, which likely functionally regulated a large number of mRNAs (**Supplementary Figure S6A**). Indeed, predicted miRNA binding motifs, based on TargetScan (Grimson, et al., 2007) were not enriched among these correlated genes for most cancer subtypes (x-axis in **Figure 2A** and **Supplementary Figure S7A, S8A and S9A**). Furthermore, for OV and BRCA, the mRNA-miRNA correlations were dominated by positive associations (x-axis in **Figure 2B** and **Supplementary Figure S9B**), indicating that the correlation between expression levels of miRNAs and mRNAs was likely due to pleiotropic effect of upstream regulators. Together, our results suggest that the expression level of miRNAs was not sufficient to identify functional relationships between miRNA and mRNA nor key miRNAs for cancers. This motivates us to infer "*functional*" activity of a miRNA by integrative analysis of expression levels of miRNA and its target genes.

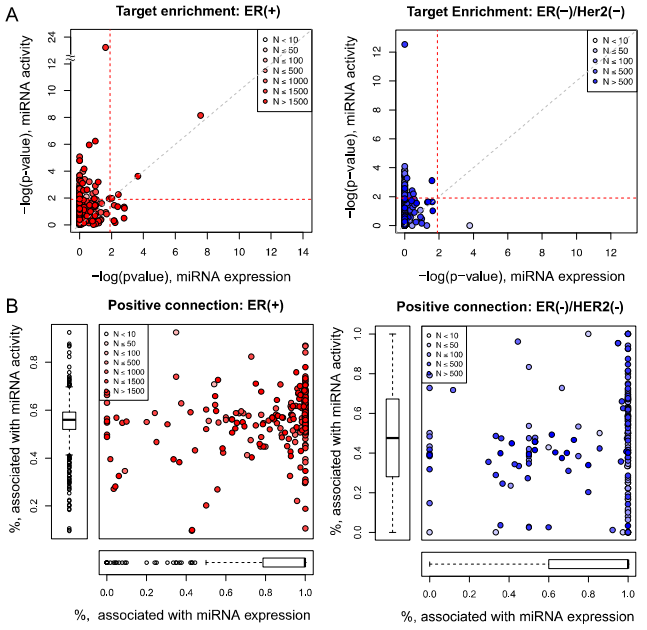

Figure 2. (A) The enrichment of miRNA binding motifs. The x-axis represents $-\log_{10}$ (p-values of the Fisher's exact test) for enrichment of miRNA binding motifs among genes whose expression levels were correlated with each miRNA expression levels. The y-axis represents the corresponding $-\log_{10}$ (p-values) based on correlations with miRNA activities. The color strength indicates the total number of genes whose expression levels are correlated with each miRNA activity levels. **(B) The percentage of positive miRNA-mRNA correlations.** The percentage of positively correlated mRNAs for each miRNA based on miRNA expression levels (x-axis) and that based on miRNA activities (y-axis) were shown. Each dot in the plots represents one miRNA.

3.3 Inferred miRNA activities were subtype specific

As the accessibility of essential miRNA machinery or relative abundance of miRNA targets might affect the activity of miRNA (Krol, et al., 2010), the inferred miRNA activities did not always correlate with miRNA expression levels (**Supplementary Figure S10**). In general, the correlation between activity and expression level of miRNA was positive corresponding to its role in target degradation, therefore, the distribution of their correlation coefficients was positively skewed (**Supplementary Figure S10**). It is worth to note that inferred miRNA activities were not sensitive to platforms used for profiling miRNA expression (**Supplementary Figure S11**). For the TCGA data sets, there were more miRNAs with high correlation between its activity and expression value for BRCA and OV. There were also subtype-common and subtype-specific miRNAs. For example, miR-18a was one of miRNAs, whose activity was the most significantly correlated with its expression levels in both BRCA subtypes (r= 0.62 and 0.54 for $ER^-/HER2^-$ and ER^+, respectively), whereas, the correlation between miR-200b expression and its activity was significant only for $ER^-/HER2^-$ samples (r= 0.42) but not for ER^+ samples (r= -0.004)

To distinguish potential causal relationships from associations between miRNA activities and their correlated mRNAs (**Supplementary Figure S6),** we investigated the enrichment of predicted target genes among genes associated with each miRNA activity. We showed that genes whose expression levels correlated with miRNA activities were more significantly enriched for miRNA target genes predicted based on TargetScan (Grimson, et al., 2007) than genes correlated with miRNA expression levels (y-axis in **Figure 2A** and **Supplementary Figures S7A, S8A and S9A**). Furthermore, for OV and BRCA, the dominance of positive correlations between mRNA and miRNA was not detected when miRNA activities instead of miRNA expression levels were used (y-axis in **Figure 2B** and **Supplementary Figure S9B**). For example, the distributions of the percentage of positive associations based on miRNA activity and miRNA expression were significantly different for both $ER^-/HER2^-$ (the Wilcoxon rank sum test p-value $< 10^{-15}$) and ER^+ subtypes (p-value $< 10^{-15}$). This observation suggested that the activity of a miRNA implies its potential of functional regulation on target mRNAs, and the activity of miRNA could be used to infer key miRNAs, which are defined as miRNAs that might causally regulate a large number of mRNAs.

3.4 Identifying key miRNAs leads to novel insights into breast cancer biology

We defined key miRNAs as the miRNAs whose activities associated with expression levels of a number of genes larger than expected by chance. For the BRCA data set, 85 and 96 key miRNAs were identified for $ER^-/HER2^-$ and ER^+ subtypes, respectively (**Supplementary Tables S4** and **S5** for ER^+ and $ER^-/HER2^-$ subtypes, respectively). Among them, 44% and 39% of key miRNAs for $ER^-/HER2^-$ and ER^+, respectively, were common. Key miRNAs included well-known miRNAs whose causal roles in breast cancers have been previously studied, such as miR-203, miR-495 and miR-125b as ER^+ specific key miRNAs (Zhang, et al., 2011; Hwang-Verslues, et al., 2011; Scott, et al., 2007), miR-9, let-7d, let-7g, miR-200b, and miR-200a as $ER^-/HER2^-$ specific ones (Zhou, et al., 2012; Yu, et al., 2007; Gregory, et al., 2008) and miR-27a, miR-17, and miR-20a, miR-141, miR-106b, miR-301a, miR-339-5p,

miR-429, miR-200c, miR-222 as common key miRNAs (Mertens-Talcott, et al., 2007; Hossain, et al., 2006; Yu, et al. 2008; Smith, et al., 2012; Shi, et al., 2011; Wu, et al., 2010; Zhao, et al., 2008).

To understand function of key miRNAs, we compared the functional target genes of each key miRNA with 1320 canonical pathways from BioCarta, KEGG and Reactome (Subramanian, et al., 2005), identified biological pathways significantly enriched in functional target gene set of each miRNA at FDR < 1% corresponding to the Fisher's exact test p-value $<1x10^{-4}$ (**Supplementary Figures S12-15**). For the ER$^-$/HER2$^-$ BRCA, we found total 23 biological pathways that were significantly enriched in target sets of at least one of 85 key miRNAs and the 11 key miRNAs with at least one enriched pathway were clustered into two groups according to their inferred function (**Figure 3**). The first cluster contained five miRNAs (miR-20a, miR-17, miR-93, miR-106a, and miR-106b), which were strongly associated with cell cycle related pathways (**Figure 3**), consistent with previous findings that miR-17 and miR-20a were regulators of cell cycle transition in breast cancer and cancer cell proliferation (Yu, et al., 2008). The functional targets of these miRNAs included known cell cycle regulators such as *E2F3* and *CDC25A*. The second cluster contained two miRNAs (let-7g and let-7d) that were enriched for several pathways, such as the ECM receptor interaction, and the collagen formation (**Figure 3**). The functional target genes within these pathways include twelve collagen genes and ten molecules associated with extracellular matrix receptor (e.g. *THBS1, THBS2, FN1, ITGA11*, and *ITGB5*) (**Supplementary Figure S16**). The activities of these miRNAs were significantly anti-correlated with expression levels of their target genes, representing degradation of transcripts of these genes by miRNAs. A recent study showed that high levels of collagen in breasts were associated with breast cancer metastasis (Zhang, et al., 2013), which is common in triple negative breast cancers. In particular, let-7g was shown to promote breast cancer invasion and metastasis (Qian, et al., 2011). These together suggested the importance of these two key miRNAs (let-7d and let-7g) in ER$^-$/HER2$^-$ breast cancers metastasis.

For ER$^+$ BRCA, 27 pathways were significantly enriched (FDR <0.01) in target gene sets of at least one key miRNA (**Supplementary Figure S12A**). Among the 27 pathways, 6 pathways including cell cycle pathway were significantly enriched for both ER$^-$/HER2$^-$ and ER$^+$. In particular, the genes involved in the cell cycle pathway were regulated by four key miRNAs (miR-106a, miR-93, miR-17, and miR-20a), consistent in ER$^-$/HER2$^-$ subtype.

3.5 Functional activities of key miRNAs were associated with survival rate

To further explore functional relevance of key miRNAs, we tested whether there was association between overall survival and the activities of each miRNA based on TCGA data. We identified 3 and 30 significant prognostic key miRNAs at 5% FDR (p-value $<1.2x10^{-2}$) for ER$^+$ and ER$^-$/HER2$^-$ subtype of BRCA, respectively (**Table 1**). It is worth to note that miRNA activities were more significantly associated with clinical outcome than their corresponding expression levels for most key miRNAs (**Figure 5A**).

There were only 3 prognostic key miRNAs for ER$^+$ breast cancer. One reason for lack of prognostic key miRNAs might be that downstream genes modulated by miRNAs were highly overlapped (**Supplementary Figure S17A**), suggesting that regulatory networks of ER$^+$ breast cancer were regulated together by many key miRNAs. Therefore, perturbing individual miRNAs was unlikely to have a significant impact on regulatory networks of ER$^+$ breast cancer, resulting in less prognostic power of individual miRNA.

In contrast, we identified 30 prognostic key miRNAs for ER$^-$/HER2$^-$ breast cancer, whereas only 11 non-key miRNAs were significantly associated with clinical outcomes (p-value$<1.2 \times 10^{-2}$) (**Figure 5A** and **Supplementary Table S6**). The activities of well-known miRNAs whose causal roles in breast cancers have been studies, including miR-200c, miR-200b (Gregory, et al., 2008) and miR-106b, miR-17, miR-20a (Hossain, et al., 2006; Smith, et al., 2012; Yu, et al., 2008), were associated with clinical outcomes.

We found the prognostic miRNAs with particular interests. First, the functional target genes of three prognostic miRNAs including miR-106b, miR-17, miR-20a were highly enriched for mitotic cell cycle (**Figure 3**). Secondly, let-7d and let-7g were detected as a regulator of metastasis (**Figure 3**) as well as the prognostic miRNAs. Tumors with over-active let-7d and let-7g were associated with better overall survival rate in ER$^-$/HER2$^-$ breast cancers (p-value<0.001) (**Supplementary Figure S18BC**). These observations were consistent with the previous study showing association of let-7g depletion with poor prognosis and its effect on tumor metastasis (Qian, et al., 2011). Furthermore, miR-18a was the key miRNA of prognostic power in ER$^-$/HER2$^-$ breast cancer, corresponding to its role in induction of malignancy (Mouw, et al., 2014). Even though the activity and the expression level of has-miR-18a were significantly correlated (r=0.62) in ER$^-$/HER2$^-$ breast cancer, its activity was significantly associated with survival (p-value<0.003) while its expression level was not (**Figure 5A**). This suggests the activity of miRNA instead of expression level of miRNA was functionally significant in breast cancer prognosis.

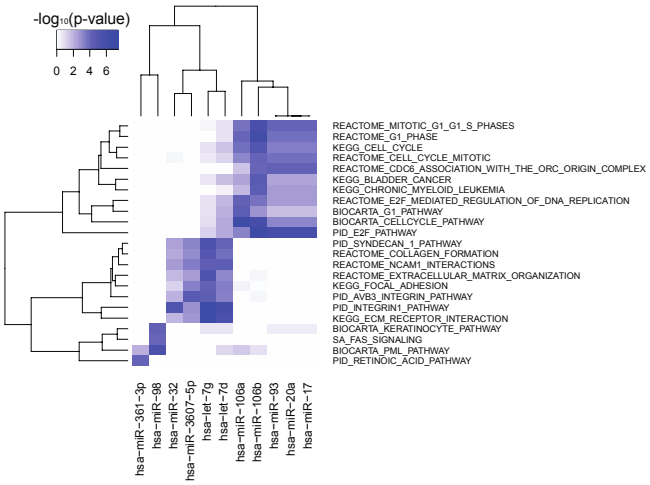

Figure 3. Functional annotation of key miRNAs' functional targets. Heatmap of pathway enrichment of functional target genes of each key miRNA for ER$^-$/HER2$^-$ group of BRCA. The displayed pathways were significantly enriched for target genes of at least one key miRNA.

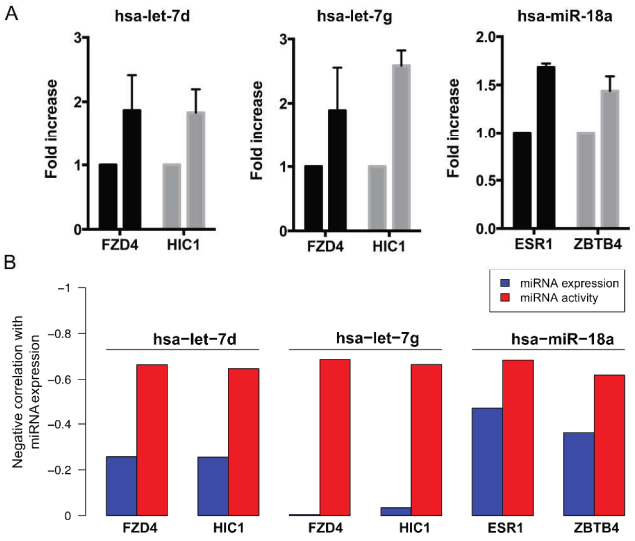

Figure 4. Experimental validation for functional target genes of key miRNAs. (A) Fold increase of each miRNA's predicted target genes using the control and siRNA treatment of each miRNA were shown. (B) The Pearson correlation between target genes of each miRNAs and miRNA expression (blue) or miRNA activity (red). The miRNA activities of let-7d and let-7g significantly correlated with the expression levels of *FZD4* and *HIC1*, whereas the expression of let-7d and let-7g did not. Knocking down the expression levels of *let-7d* and *let-7g* in a triple negative breast cancer cell line HCC1187 led to up-regulation of *FZD4* and *HIC1*. Similarly, up-regulation of *ESR1* and *ZBTB4* by knocking down miR-18a suggested *ESR1* and *ZBTB4* were functional target genes of miR-18a.

3.6 Experimental validation of key miRNA-target pairs in vitro

Inference of miRNA activity using our ActMiR method relies on expression level changes of miRNA target genes. To validate potential causal relationships between prognostic miRNAs and their target genes, which affect biological processes related to patient survivals, we perturbed prognostic miRNAs let-7g, let-7d and miR-18a in a triple negative breast cancer cell line and measured responses of their target genes. For each miRNA, we measured expression changes of two functional targets among its target genes, whose expression levels were most correlated with miRNA activity in the TCGA data set. Comparison of qPCR results of the control and siRNA of each miRNA confirmed that each miRNA regulated the expression of its predicted target genes (**Figure 4A**). As expected, *HIC1* and *FZD4* mRNA expression increased in anti-miRNA treatment. It is worth to note that *HIC1* and *FZD4* were predicted as functional targets of let-7d and let-7g only based on miRNAs' activities but not based on miRNAs' expression levels (**Figure 4B**), further validating the importance of the miRNA activity.

3.7 Validating prognostic value of key miRNAs in independent cohorts

To validate prognostic potential of key miRNAs' activities, we collected 2 independent breast cancer cohorts (Buffa, et al., 2011; Dvinge, et al., 2013) with miRNA and gene expression profiles **(Supplementary Table S3)**. For each ER$^-$/HER2$^-$ subtype of validation dataset, we performed the same procedure and identified 53

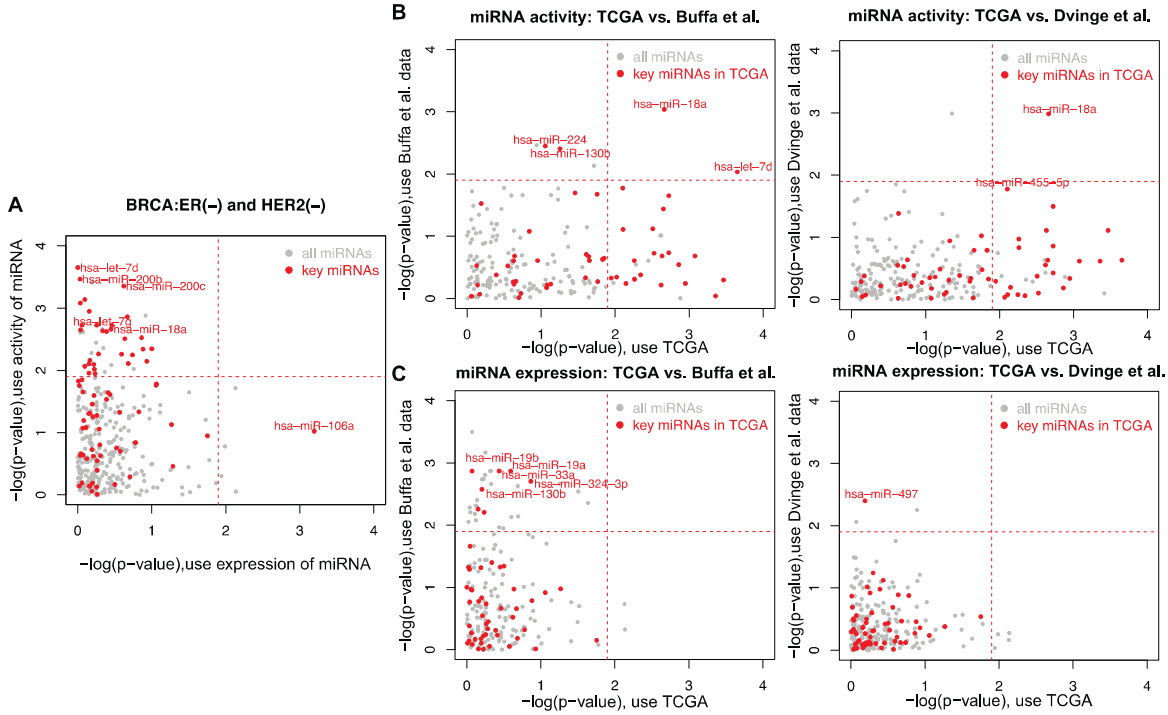

Figure 5. Activities of key miRNAs identified in TCGA data sets are of robust prognostic value in independents cohort data sets for ER$^-$/HER2$^-$ breast cancer. (A) Survival prognosis by miRNA activity (y-axis) and expression level (x-axis), using a likelihood ratio test was shown. Analysis based on ER$^-$/HER2$^-$ subtype of BRCA of TCGA samples. The red dot represented key miRNAs. (B) Survival prognosis by miRNA expression based on two independents cohort data sets (y-axis), Buffa *et al.* data set (Buffa, et al., 2011) and Dvinge *et al.* data set (Dvinge, et al., 2013), were compared to the results based on TCGA data (x-axis). (C) Survival prognosis by miRNA activity based on two independents cohort data sets (y-axis), Buffa *et al*. data set (Buffa, et al., 2011) and Dvinge *et al*. data set (Dvinge, et al., 2013), were compared to the results based on TCGA data (x-axis).

and 36 key miRNAs for Buffa *et al.* and Dvinge *et al*. dataset, respectively. Encouragingly, these key miRNAs significantly overlapped with the ones derived from the TCGA BRCA data set (p-value < $2x10^{-5}$, and 0.01 for Buffa *et al.* and Dvinge *et al*. dataset, respectively). For the $ER^-/HER2^-$ breast cancer in Buffa *et al.* dataset, the activities of let-7d, miR-18a, hsa-miR-130b, and miR-224 were associated with metastatic outcomes ($p<10^{-2}$) (**Figure 5C**). Two of them (let-7d and miR-18a) were prognostic key miRNAs in TCGA samples (**Figure 5C** and **Table 1**). Based on Dvinge *et al*. data set, the activity of miR-18a was consistently associated with survival of patients with the most advanced tumors (Stage III) (**Figure 5C**). It is worth to note that the expression of let-7d, miR-18a were not associated with overall survival in any data sets (**Figure 5B** and **Supplementary Figure S18AB**). Furthermore, the functional target genes of let-7d and let-7g from TCGA dataset significantly overlapped with the sets from the validation datasets (for let-7d, p-values $<2x10^{-22}$, and $<1x10^{-27}$, and for let-7g, p-values $<1x10^{-36}$, and $<1x10^{-60}$ for Buffa et al. and Dvinge et al. dataset, respectively). The target genes of miR-18a based on were TCGA and Buffa et al. samples also significantly overlapped (p-value$<1x10^{-27}$). Together, our results suggest the miRNA activity can robustly predict patient survival in independent validation sets while miRNA expression levels can't.

Table 1. Summary of the key miRNAs for each cancer subtype.

Type	Subtypes	Key[1]	Prognosis[2]	Robust prognostic keys[3]
BRCA	ER^+	96	3	miR-500a
	$ER^-/HER2^-$	85	30	let-7d, miR-18a
COAD	Invasive	141	3	-
	CIN	175	0	-
	MSI/CIMP	229	32	miR-301b, miR-519a, miR-548b-5p
GBM	Classical	184	0	-
	Mesenchymal	77	0	-
	Neural	159	0	-
	Proneural	102	4	miR-29a
OV		180	16	miR-519d, miR-520d*, miR-9

[1] Number of key miRNAs; [2] Number of prognostic key miRNAs; [3] Robust prognostic key miRNAs that are significant in both TCGA and the independent cohort data set.

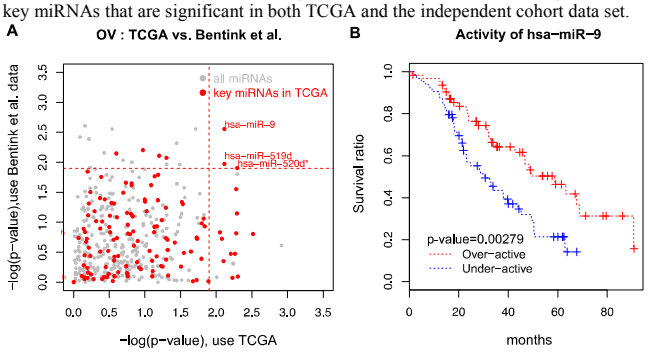

Figure 6. Activities of key miRNAs identified in TCGA data sets are of robust prognostic value in independents cohort data sets for OV. (A) Survival prognosis by miRNA activity based on the independents cohort data set (y-axis), Bentink *et al.* ovarian cancer data set (Bentink, et al., 2012), were compared to the results from TCGA data (x-axis). (B) Kaplan-Meier survival curve based on the activity of miR-9 based on Bentink *et al.* data set (Bentink, et al., 2012). The blue and red curve represents under and over active group, respectively

3.8 Identifying key miRNAs in other types of cancers

To ask whether our approach can be applied to cancer bioloy in general, we applied our ActMiR approach above to the GBM, COAD, and OV data sets and identified 77~229 key miRNAs for each caner subtype (**Table 1**). For GBM, about 40% of key miRNAs for each subtype (e.g. 39%, 32%, 38% and 43% for classical, mesenchymal, neural and proneural subtypes, respectively) were subtype specific miRNAs, whereas for COAD, 60% of total key miRNAs were common for at least two subtypes. We also compared the direct functional target genes of each key miRNA with 1320 canonical pathways from BioCarta, KEGG and Reactome (Subramanian, et al., 2005), and biological pathways regulated by each miRNA at FDR < 1% (**Supplementary Figures S13-15**). Interestingly, for the COAD, we detected that most pathway-miRNA pairs were subtype specific (**Supplementary Figure S13**). On the other hand, the neurotransmitter related pathways were common among subtypes of GBM (**Supplementary Figure S14**).

We compared miRNA activity and patient survival information, and identified 16, 32, and 4 significant prognostic key miRNAs at 5% FDR (corresponding to p-value $<1.2x10^{-2}$) for OV, MSI/CIMP subtype of COAD, and Proneural subtype of GBM, respectively (**Supplementary Figures S19A, S20A, and S21A, Table 1,** and **Supplementary Table S6**). To test whether these prognostic key miRNAs' activities can robust predict patient's survival, we assessed their prognostic potentials in independent cohorts with miRNA and gene expression profiles (**Supplementary Table S3).** Results were shown in **Supplementary Figures S19B, S20B, and S21B.**

For OV, the activities of three prognostic key miRNAs (miR-9, miR-519d, and miR-520d*) in TCGA data were also significantly associated with clinical outcome in the independent cohort data set (Bentink, et al., 2012) (**Figure 6A**). It has been shown that miR-9 inhibits cell growth of ovarian cancer (Guo, et al., 2009). Our result showed that the over-active group of miR-9 had better survival rate (**Figure 6B**), consistent with the previous study, while the expression level of miR-9 were not associated with the clinical outcomes.

DISCUSSION

We presented a novel computational approach ActMiR for inferring miRNA activity and identifying key miRNAs that might causally regulate a large number of genes and applied it to multiple cancer types. Our approach explicitly infers miRNA activities instead of using miRNA expression levels as a surrogate. By applying this method to mRNA and miRNA expression profiles in four different cancer types, we identified 77~229 key miRNAs for each tumor subtype. In particular, when applied our approach to the TCGA breast cancer data set, we identified 85 and 96 key miRNAs for $ER^-/HER2^-$ and ER^+ breast cancers, respectively. It is worth to note that results based on clinical subtyping (ER/HER2/PR status) and molecular subtyping (PAM50 classification) were similar (**Supplementary Figure S22A** and detailed in Supplementary Results). Key miRNAs of prognostic potentials were still significantly associated with survival after adjusting clinical/pathological parameters (**Supplementary Figure S23**). $ER^-/HER2^-$ especially triple negative breast tumors are frequently invasive and metastatic involving biological processes such as

extracellular matrix remodeling. Interestingly, we identified two miRNAs (let-7d and let-7g) that regulated genes involved in extracellular matrix. Furthermore, we showed that activities of let-7d were consistently associated with survival in the independent breast cancer studies. Our results not only support metastasis is associated with survival of triple negative breast cancers, but also show that let-7d, and let-7g are putative key regulators of metastasis. Additionally, the activity of miR-18a was prognostic in all three independent cohorts. In addition to as potential prognostic biomarkers, the inferred activities of these key miRNAs in ER^-/$HER2^-$ breast cancers can also be valuable therapeutics targets as there are very limited therapeutic options for triple negative breast cancers.

Our findings based on an unbiased systematic approach were supported by the previous observations based on experimental approach in great details (Mouw, et al., 2014; Qian, et al., 2011). Our analyses suggested the role of let-7g in tumor metastasis (Qian, et al., 2011) and miR-18a as a malignancy promoting factor in breast tumors (Mouw, et al., 2014) by the miRNA activity followed by its functional annotation and prognosis test. Furthermore, our result also suggested the novel role of let-7d in metastasis and prognosis, whereas the previous study failed to identify it because their study was based on the expression levels of let-7 miRNA family instead of miRNA activity (Qian, et al., 2011). This suggests our approach based on the inferred miRNA activity will facilitate to find more tumor-promoting or prognostic miRNAs that may not be identified based on their expression levels alone.

Different types of survival outcomes were used in the TCGA, Dvinge et al, and Buffa et al. breast cancer data sets. By overlapping results based on these data sets, we reduced potential false positives, but might also miss many true positives. However, survival analyses based on both miRNA expression level and miRNA activities suffered the same drawback. Our result showed that there was no prognostic miRNA based on expression level consistent in multiple independent data sets while there were multiple prognostic miRNAs based on activities consistent in multiple independent data sets, suggesting our approach based on the miRNA activity is robust.

In an effort to uncover prognosis biomarkers of cancers, miRNAs would be a suitable candidate due to their small size resulting in resistant to RNase degradation (Lawrie, 2008). Furthermore, we showed that the effect of miRNAs for prognosis was robust. We measured the effectiveness and robustness of miRNAs vs. CNV as prognostic markers using two independent cohorts of OV (**Supplementary Figure S24** and detailed in Supplementary Results). CNVs are predominant functional genomic alterations in OV (Ciriello, et al., 2013). Interestingly, we found only one consistent prognostic CNV factor or associated mRNA factor, whereas 3 consistent prognostic miRNAs based on activity in OV (**Supplementary Figure S24**), suggesting a potential of miRNA activities as effective and robust biomarkers.

It is worth to note that our ActMiR approach aims to infer miRNA activities, not to comprehensively identify miRNA targets. The ActMiR procedure for inferring miRNA activities is based on regression models. Too few candidate targets of a miRNA included in the ActMiR procedure may result in less robustness of miRNA activity estimation. On the other hand, too many low confident candidate targets included will bias miRNA activity estimation to zero. There were multiple miRNA target prediction data bases in addition to TargetScan, such as miRanda (Betel, et al. 2010) and PITA (Kertesz, et al., 2007). Compared to other databases, TargetScan covered more miRNAs (**Supplementary Figure S25A**). Also, target genes based on TargetScan were more consistent with experimentally derived targets (such as by CLASH containing both canonical and non-canonical targets (Helwak, et al., 2013), and PAR-CLIP containing canonical targets (Farazi, et al., 2014) methods) than other database (**Supplementary Figure S25BC**).

Unlike TargetScan which predicted canonical miRNA targets, experimental approaches such as CLASH (Helwak, et al., 2013) revealed both canonical and non-canonical miRNA targets. However, miRNA-target gene interactions are subtype/context-specific. Most experimental methods covered only limited number of conditions, resulting large numbers of false positives or false negative, which in turn affect accuracy of miRNA activity inference based on the regression model. Also, experimentally derived target sets were only available for a fraction of miRNAs. For example, CLASH dataset (Helwak, et al., 2013) consisted of 399 miRNAs and PAR-CLIP dataset (Farazi, et al., 2014) consisted of only 68 miRNAs. It is hard to make fair comparison of miRNA activities across all miRNAs if target genes for some miRNAs contain experimentally derived targets while others do not. Thus, we used only TargetScan but excluding other experimentally derived target databases in our current study.

The target genes from the TargetScan database were mainly predicted based on seed sequences so that miRNAs in the same miRNA family share targets. To test whether considering miRNAs in the same miRNA family together enhance the ability to identify potential miRNA functional targets, we compared the percentage of negative correlations between the predicted target genes' expression levels with the sum of the expression levels of miRNAs in a family or the expression levels of individual miRNAs. Our result (**Supplementary Figure S26)** showed that considering all members in a miRNA family as a whole did not improve regulatory potentials of miRNAs. On the other hand, many studies (Boyerinas, et al., 2010; Elefant, et al., 2011; Helwak, et al., 2013; Wu, et al., 2015) suggested that miRNAs in the same miRNA family have different binding patterns to their target genes even though they have the same seed sequences. Also, only a fraction of predicted target genes were regulated by a miRNA, reflecting to idiosyncratic patterns of miRNAs in the miRNA family. We used WLS regression to account for binding affinity differences to different target genes when inferring a miRNA activity. As a result, different miRNAs in the same miRNA family, which had the same set of target genes based on TargetScan, had different sets of functional targets. For example, activities of let-7a and let-7d correlated with different genes among the same set of target genes (**Supplementary Figure S27**).

A miRNA can post-transcriptionally regulate many target genes. In our validation experiment, we only tested two targets whose expression levels most correlated corresponding miRNA activities. Validation of a few target genes may not guarantee that other target genes will be regulated by miRNAs nor related pathways regulated miRNAs. Instead of genome-wide target gene validation and func-

tional validation, which are needed in future studies, we validated miRNA-target relationships by showing that miRNA functional target sets based on independent data sets were significantly overlapped.

Our regression procedure for inferring miRNA activity fully accounts for post-transcriptional regulation by miRNA activity in contrast to directly use the expression level of either the miRNA or targeted genes as a surrogate for miRNA regulatory activity. Indeed, the target genes of miRNAs that were predicted based on the activity and were experimentally validated in this study could not be detected as target genes based on corresponding miRNA expression. There are multiple experimental approaches for estimating miRNA activities. Mullokandov *et al*. (Mullokandov, et al., 2012) proposed to experimentally measure miRNA activity by quantifying its target genes in high-throughput manners. They demonstrated that some highly expressed miRNAs might exhibit relatively weak activity, which in some cases correlated with a high target-to-miRNA ratio or increased nuclear localization of the miRNA. Our results were consistent with their results: only one third of miRNAs analyzed show significant (FDR<1%) correlation between their expression levels and activities. However, our method is different from their method in several aspects: (1) our inferred activity is cell context specific instead of synthetic environments; (2) our method is based on degradation levels of all genes with predicted miRNA binding motifs instead of expression levels of a few genes with conserved binding motifs.

Similarly, Pecot et al. (Pecot, et al., 2013) and other papers showed that genes and associated pathways functionally regulated by a miRNA were cancer subtype specific. Farazi et al. (Farazi, et al., 2014) combined experimentally defined miRNA binding sites and TargetScan predicted miRNA binding sites to define miRNA target genes and used difference of median correlations of miRNA target genes or miRNA non target genes with miRNA expression level to indicate miRNA activity. Farazi et al. showed that only a fraction of target genes predicted by TargetScan were repressed by a miRNA and miRNA activity was cancer subtype specific. Even though some conclusions were similar, our approach significantly differed from Farazi et al's approach in multiple ways. First, our iterative weighted regression approach for inferring miRNA activity explicitly identified functional target genes of miRNAs in each cancer subtype based on in vivo data. In contrast, Farazi et al. inferred subtype specific target genes based on in vitro data of one single cell line of other cancer subtype, which contradicted with their own assumption that miRNA target genes were cancer subtype specific. Second, even though the choice of using correlation between gene expression and miRNA activity and target gene enrichment test to infer functionally active miRNAs in our approach or correlation between gene expression and miRNA expression level to reflect miRNA function activity used by Farazi et al. is subjective, our approach is applied to multiple breast cancer data sets, indicating that our ActMiR approach for inferring miRNA activities is robust. More importantly, our approach can be readily applied to any cancers without the need of experimentally deriving miRNA targets in the specific cancer type.

Taken together, our results underscore the value of inferring miRNA activity from a systems biology perspective as a promising strategy for investigation of its causal effects on target genes, and furthermore survival outcomes of cancer patients. There are several directions in which this approach can be extended. First, the integration of miRNA, mRNA, CNV, and other high throughput data into Bayesian causal models (Zhu, et al., 2012) may shed light on how key miRNAs are regulated and lead to further refine miRNA-mediated regulatory networks. It will also be interesting to uncover the mechanisms underlying the miRNA functions by comparison of regulators between activity and expression of miRNAs using CNV or mutation data. Furthermore, a model considering several miRNAs together is needed to investigate the cooperative or synergetic effects between miRNAs function. Finally, it is worth to further assess potentials of key miRNAs as prognosis biomarkers and therapeutic agents experimentally.

ACKNOWLEDGEMENTS

Funding: The work was partially supported by the National Institute of Health (CA163772, AG046170, CA172460, and HG008451).

REFERENCES

Arora, A. and Simpson, D.A. (2008) Individual mRNA expression profiles reveal the effects of specific microRNAs, *Genome biology*, **9**, R82.

Bentink, S., *et al.* (2012) Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer, *PloS one*, **7**, e30269.

Betel, D., et al. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, *Genome biology*, 11, R90

Boyerinas, B., et al. (2010) The role of let-7 in cell differentiation and cancer, *Endocrine-related cancer*, 17, F19-36.

Brennan, C.W., *et al.* (2013) The somatic genomic landscape of glioblastoma, *Cell*, **155**, 462-477.

Buffa, F.M., *et al.* (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer, *Cancer research*, **71**, 5635-5645.

Calin, G.A., *et al.* (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia, *The New England journal of medicine*, **353**, 1793-1801.

Cancer Genome Atlas, N. (2012) Comprehensive molecular characterization of human colon and rectal cancer, *Nature*, **487**, 330-337.

Cancer Genome Atlas, N. (2012) Comprehensive molecular portraits of human breast tumours, *Nature*, **490**, 61-70.

Chen, Y., *et al.* (2008) Variations in DNA elucidate molecular networks that cause disease, *Nature*, **452**, 429-435.

Cheng, C., *et al.* (2009) mRNA expression profiles show differential regulatory effects of microRNAs between estrogen receptor-positive and estrogen receptor-negative breast cancer, *Genome biology*, **10**, R90.

Ciriello, G., *et al.* (2013) Emerging landscape of oncogenic signatures across human cancers, *Nature genetics*, **45**, 1127-1133.

Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression, *Genes Dev*, **18**, 504-511.

Dvinge, H., *et al.* (2013) The shaping and functional consequences of the microRNA landscape in breast cancer, *Nature*, **497**, 378-382.

Ebert, M.S., Neilson, J.R. and Sharp, P.A. (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells, *Nature methods*, **4**, 721-726.

Elefant, N., Altuvia, Y. and Margalit, H. (2011) A wide repertoire of miRNA binding sites: prediction and functional implications, *Bioinformatics*, 27, 3093-3101.

Farazi, T.A., et al. (2014) Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets, *Genome biology*, 15, R9.

Gennarino, V.A., *et al.* (2012) Identification of microRNA-regulated gene networks by expression analysis of target genes, *Genome research*, **22**, 1163-1172.

Gregory, P.A., *et al.* (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1, *Nature cell biology*, **10**, 593-601.
Grimson, A., *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing, *Mol Cell*, **27**, 91-105.
Gruvberger, S., *et al.* (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns, *Cancer research*, **61**, 5979-5984.
Guo, L.M., *et al.* (2009) MicroRNA-9 inhibits ovarian cancer cell growth through regulation of NF-kappaB1, *The FEBS journal*, **276**, 5537-5546.
Helwak, A., et al. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding, *Cell*, 153, 654-665.
Hwang-Verslues, W.W., et al. (2011) miR-495 is upregulated by E12/E47 in breast cancer stem cells, and promotes oncogenesis and hypoxia resistance via downregulation of E-cadherin and REDD1, *Oncogene*, 30, 2463-2474
Hossain, A., Kuo, M.T. and Saunders, G.F. (2006) Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA, *Molecular and cellular biology*, **26**, 8191-8201.
Iorio, M.V., *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer, *Cancer research*, **65**, 7065-7070.
Jansson, M.D. and Lund, A.H. (2012) MicroRNA and cancer, *Molecular oncology*, **6**, 590-610.
Kertesz, M., et al. (2007) The role of site accessibility in microRNA target recognition, *Nature genetics*, 39, 1278-1284.
Krol, J., Loedige, I. and Filipowicz, W. (2010) The widespread regulation of microRNA biogenesis, function and decay, *Nature reviews. Genetics*, **11**, 597-610.
Lawrie, C.H. (2008) MicroRNA expression in lymphoid malignancies: new hope for diagnosis and therapy?, *Journal of cellular and molecular medicine*, **12**, 1432-1444.
Madden, S.F., *et al.* (2010) Detecting microRNA activity from gene expression data, *BMC bioinformatics*, **11**, 257.
Mertens-Talcott, S.U., *et al.* (2007) The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells, *Cancer research*, **67**, 11001-11011.
Mouw, J.K., *et al.* (2014) Tissue mechanics modulate microRNA-dependent PTEN expression to regulate malignant progression, *Nature medicine*, **20**, 360-367.
Mullokandov, G., *et al.* (2012) High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries, *Nat Methods*, **9**, 840-846.
Pecot, C.V., et al. (2013) Tumour angiogenesis regulation by the miR-200 family, *Nature communications*, 4, 2427.
Qian, P., *et al.* (2011) Pivotal role of reduced let-7g expression in breast cancer invasion and metastasis, *Cancer research*, **71**, 6463-6474.
Scott, G.K., *et al.* (2007) Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b, *The Journal of biological chemistry*, **282**, 1479-1486.
Setty, M., *et al.* (2012) Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma, *Molecular systems biology*, **8**, 605.
Shi, W., *et al.* (2011) MicroRNA-301 mediates proliferation and invasion in human breast cancer, *Cancer research*, **71**, 2926-2937.
Smith, A.L., *et al.* (2012) The miR-106b-25 cluster targets Smad7, activates TGF-beta signaling, and induces EMT and tumor initiating cell characteristics downstream of Six1 in human breast cancer, *Oncogene*, **31**, 5162-5171.
Song, S.J., et al. (2013) MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling, *Cell*, 154, 311-324.
Subramanian, A., *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc Natl Acad Sci U S A*, **102**, 15545-15550.
Verhaak, R.G., *et al.* (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1, *Cancer cell*, **17**, 98-110.
Wu, Z.S., *et al.* (2010) MiR-339-5p inhibits breast cancer cell migration and invasion in vitro and may be a potential biomarker for breast cancer prognosis, *BMC cancer*, **10**, 542.
Wu, L., et al. (2015) MicroRNA let-7g and let-7i inhibit hepatoma cell growth concurrently via downregulation of the anti-apoptotic protein B-cell lymphoma-extra large, Oncology letters, 9, 213-218.
Yu, F., *et al.* (2007) let-7 regulates self renewal and tumorigenicity of breast cancer cells, *Cell*, **131**, 1109-1123.
Yu, Z., *et al.* (2008) A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation, *The Journal of cell biology*, **182**, 509-517.
Zhang, Z., et al. (2011) Epigenetic Silencing of miR-203 Upregulates SNAI2 and Contributes to the Invasiveness of Malignant Breast Cancer Cells, *Genes & cancer*, 2, 782-791.
Zhang, K., *et al.* (2013) The collagen receptor discoidin domain receptor 2 stabilizes SNAIL1 to facilitate breast cancer metastasis, *Nature cell biology*.
Zhao, J.J., *et al.* (2008) MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer, *The Journal of biological chemistry*, **283**, 31079-31086.
Zhou, X., *et al.* (2012) MicroRNA-9 as potential biomarker for breast cancer local recurrence and tumor estrogen receptor status, *PloS one*, **7**, e39011.
Zhu, J., *et al.* (2012) Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation, *PLoS Biol*, **10**, e1001301.