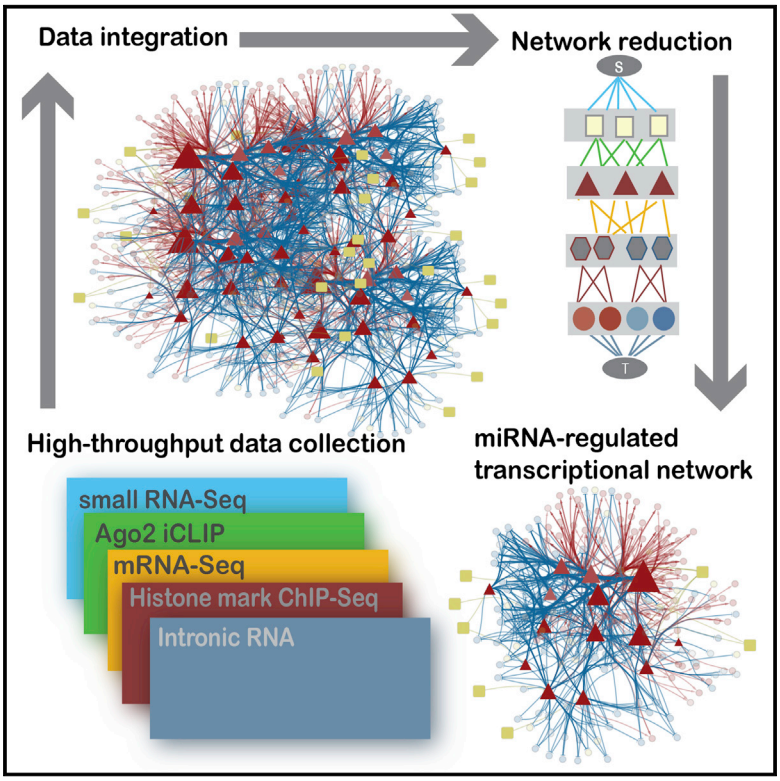
Article

Cell Reports

Elucidating MicroRNA Regulatory Networks Using Transcriptional, Post-transcriptional, and Histone Modification Measurements

Graphical Abstract



Authors

Sara J.C. Gosline, Allan M. Gurtan, Courtney K. JnBaptiste, ..., Yoon S. Yap, Phillip A. Sharp, Ernest Fraenkel

Correspondence

sharppa@mit.edu (P.A.S.), fraenkel-admin@mit.edu (E.F.)

In Brief

Gosline et al. systematically profile microRNAs, mRNAs, and histone modifications to decouple transcriptional and post-transcriptional regulatory modes in a model of global miRNA loss. They then apply a network modeling approach to identify miRNA-regulated transcription factors that give rise to the changes in mRNA levels observed upon miRNA loss.

Highlights

- miRNA depletion causes global rewiring of transcriptional regulatory network
- The resulting changes are dominated by indirect, transcriptional effects
- Network modeling identifies transcription factors that execute miRNA programs

Accession Numbers

GSE61035
GSE61033
GSE61034
GSE61031
GSE44159
GSE45828

Gosline et al., 2016, Cell Reports *14*, 310–319
January 12, 2016
http://dx.doi.org/10.1016/j.celrep.2015.12.031

OPEN ACCESS
CellPress

Cell Reports
Article

Elucidating MicroRNA Regulatory Networks Using Transcriptional, Post-transcriptional, and Histone Modification Measurements

Sara J.C. Gosline,[1,4] Allan M. Gurtan,[2,4] Courtney K. JnBaptiste,[2,3] Andrew Bosson,[2,3] Pamela Milani,[1] Simona Dalin,[1] Bryan J. Matthews,[1] Yoon S. Yap,[1] Phillip A. Sharp,[2,3,*] and Ernest Fraenkel[1,*]
[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]David H. Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA 02139, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Co-first author
*Correspondence: sharppa@mit.edu (P.A.S.), fraenkel-admin@mit.edu (E.F.)
http://dx.doi.org/10.1016/j.celrep.2015.12.031

SUMMARY

MicroRNAs (miRNAs) regulate diverse biological processes by repressing mRNAs, but their modest effects on direct targets, together with their participation in larger regulatory networks, make it challenging to delineate miRNA-mediated effects. Here, we describe an approach to characterizing miRNA-regulatory networks by systematically profiling transcriptional, post-transcriptional and epigenetic activity in a pair of isogenic murine fibroblast cell lines with and without Dicer expression. By RNA sequencing (RNA-seq) and CLIP (crosslinking followed by immunoprecipitation) sequencing (CLIP-seq), we found that most of the changes induced by global miRNA loss occur at the level of transcription. We then introduced a network modeling approach that integrated these data with epigenetic data to identify specific miRNA-regulated transcription factors that explain the impact of miRNA perturbation on gene expression. In total, we demonstrate that combining multiple genome-wide datasets spanning diverse regulatory modes enables accurate delineation of the downstream miRNA-regulated transcriptional network and establishes a model for studying similar networks in other systems.

INTRODUCTION

MicroRNAs (miRNAs) are ~22-nucleotide regulatory RNAs that guide the RNA-induced silencing complex (RISC) to the 3′ UTR of mRNAs to inhibit translation and promote degradation (Baek et al., 2008; Guo et al., 2010; Selbach et al., 2008). miRNA activity is pleiotropic, with each miRNA repressing numerous targets that can be identified computationally using sequence features of mRNAs (Garcia et al., 2011; Grimson et al., 2007; Pasquinelli, 2012) or experimentally by individual nucleotide crosslinking followed by immunoprecipitation (iCLIP) of Argonaute, a member of the RISC (Chi et al., 2009; König et al., 2012; Sugimoto et al., 2012). Misregulation of miRNAs can lead to strong phenotypes in development (Chen et al., 2004) and disease (Lu et al., 2005; Mendell and Olson, 2012), despite the finding that most direct targets are only modestly (~2-fold) repressed (Baek et al., 2008).

Recent studies have found that miRNAs can have more profound effects when acting within larger regulatory networks, either alongside other miRNAs or together with transcription factors (Gurtan and Sharp, 2013; Herranz and Cohen, 2010; Schmiedel et al., 2015). When miRNAs regulate transcription factors, they can affect cellular phenotype, as demonstrated by miR-134 regulation of differentiation through interactions with mRNAs encoding Nanog and LRH1 transcription factors (Tay et al., 2008), let-7 regulation of HMGA2 (Mayr et al., 2007), or miR-145 regulation of SOX9 (Rani et al., 2013). Some studies have suggested that miRNAs preferentially target transcription factors (Lewis et al., 2003) and cause widespread changes in transcriptional activation (Gurtan et al., 2013). Additionally, miRNAs are often found within network motifs containing transcription factors, suggesting that they act alongside transcription factors to buffer gene expression (Gerstein et al., 2012; Shalgi et al., 2007; Tsang et al., 2007).

Despite the known biological importance of studying miRNA-transcription factor interactions, to date, it is still challenging to distinguish direct miRNA-mediated effects from transcriptional effects by measuring mRNA alone, via arrays or with RNA sequencing (RNA-seq). While there are both experimental (Chi et al., 2009; Wen et al., 2011) and computational (Agarwal et al., 2015; Chiu et al., 2015; Garcia et al., 2011) methods to identify miRNA targets, identifying miRNA-regulated transcriptional changes is more challenging. Numerous computational approaches have used computational target prediction algorithms with transcription factor binding prediction tools to model the downstream effects of miRNAs through transcription factors (Afshar et al., 2014; Bisognin et al., 2012; Friard et al., 2010; Naeem et al., 2011; Tu et al., 2009). Recent advances in RNA-seq efforts have enabled the use of total RNA measurements to capture both intronic and exonic changes. While this has been used as an additional way to identify genes that show

evidence of post-transcriptional rather than transcriptional regulation (Du et al., 2014; Gaidatzis et al., 2015), it can still conflate transcriptional and post-transcriptional regulation.

Recently, the use of epigenetic data such as DNase I hypersensitivity assays (Song and Crawford, 2010) and histone post-translational modification marks (Ernst and Kellis, 2010) has improved characterization of transcriptional regulatory changes. These assays can measure specific changes to chromatin configuration near transcription start sites, providing accurate identification of genes with altered transcriptional regulation in a condition of interest. Incorporating these data into transcription factor binding predictions can improve the identification of genes that are transcriptionally regulated (Heintzman et al., 2009), as well as the transcription factors that are regulating the genes (Cuellar-Partida et al., 2012; Pique-Regi et al., 2011). To date, however, measurement of epigenetic perturbations alongside miRNA perturbation has been studied only in context of general changes (Gurtan et al., 2013) and not used to characterize miRNA regulatory networks.

In this work, we describe a comprehensive approach to study the relationship between miRNAs and transcription factors through integrative analysis of epigenetic, transcriptional, and post-transcriptional changes. We apply this approach to immortalized Dicer$^{f/f}$ (wild-type; WT) and Dicer$^{-/-}$ (knockout; KO) murine fibroblast cell lines (Gurtan et al., 2013) to study the impact of global miRNA loss in a stable system. We collect and analyze miRNA expression, RNA expression, and epigenetic data in both cell lines to fully quantify the contribution of transcriptional regulatory changes compared to post-transcriptional regulation. Then, we introduce a network-based computational approach that takes advantage of these diverse high-throughput measurements to enable the identification of transcription factors likely to contribute to miRNA-mediated changes. Given the widespread availability of epigenetic and transcriptional data across various diseases, tissues, and cell line models, this approach is highly applicable to study the effect of miRNAs in many different contexts.

RESULTS

Decoupling Post-transcriptional and Transcriptional Regulation Reveals Global Changes in Transcription upon miRNA Loss

We first collected RNA from an isogenic clonal pair of immortalized Dicer$^{f/f}$ WT and Dicer$^{-/-}$ KO murine fibroblast cell lines (Gurtan et al., 2013) to quantify the changes in gene expression observed upon global miRNA loss. We sequenced two distinct RNA libraries: (1) ribo-depleted total-RNA libraries from WT and KO cells (see Experimental Procedures) to compare changes in exonic reads upon Dicer KO (Table S1A) with intronic read changes (Table S1B), which are unaffected by direct miRNA-mRNA interactions, and (2) poly(A)-selected libraries (see Experimental Procedures; Table S1C). The exonic read changes from the total RNA library were highly correlated with reads from the poly(A) library (Figure S1A; Spearman's ρ = 0.88) as well as individually selected low-throughput targets via qPCR (see Experimental Procedures; Figure S1B; Pearson's r = 0.96).

We then used intronic reads to estimate how much of the observed changes in mature mRNA expression were caused by changes in transcription. Since introns are spliced out of transcripts before export to the cytoplasm, comparisons of intronic and exonic read changes have been used as a way to isolate post-transcriptional changes of interest, as changes in intronic reads, which represent changes in pre-mRNA expression, can be "subtracted" from mature mRNA changes represented by exonic reads (Du et al., 2014; Gaidatzis et al., 2015). We first confirmed that the intronic measurements of gene expression changes were accurate in a low-throughput manner using qPCR (Figures S1C and S1D). Then, we used the intronic measurements to ask what fraction of mature mRNA changes observed upon Dicer loss could be attributed to changes in transcription. We found that changes in exonic reads were highly correlated with intronic reads within the same library (Figure 1A; Spearman's ρ = 0.94). We repeated the same analysis with gene expression changes measured in poly(A)-selected library and found a statistically significant strong correlation (Figure 1B; Spearman's ρ = 0.83). Together, these correlations strongly suggest that most significant mRNA expression changes observed after miRNA perturbation can be explained by changes in gene transcription and not miRNA-mediated degradation.

miRNA Target Identification

To identify those genes that exhibited evidence of post-transcriptional regulation, we used previously published iCLIP data (Gurtan et al., 2013) to identify bound targets of stably expressed Flag-hemagglutinin (HA)-tagged Ago2 in Dicer WT fibroblasts (Bosson et al., 2014; Zhang and Darnell, 2011) and small RNA-seq measurements (Gurtan et al., 2013) (see Experimental Procedures) from the same cells. Using these two datasets, we identified high-confidence miRNA targets as those that showed evidence of a significant (q < 0.05) iCLIP binding event in the 3′ UTR as well as a 7-mer or 8-mer seed match of an expressed miRNA, for a total of 2,754 miRNA-targeted genes in the poly(A) data and 2,729 miRNA-targeted genes in the ribo-depleted data (see Experimental Procedures for more details; Table S2). As expected, these biochemically identified targets were enriched in genes that were upregulated upon Dicer loss (p = 1.16e-11 in the poly(A) data, p = 1.66e-40 in the ribo-depleted data) and had a statistically significant impact on global mRNA expression change in both libraries (Figures 2A and 2B). Furthermore, iCLIP activity was positively correlated with an increase in expression of those targets upon Dicer KO in both RNA libraries (Pearson's r = 0.24, p = 2.36e-38 in the poly(A) data; Pearson's r = 0.27, p = 4.09e-48 in the ribo-depleted data; Figures S2B and S2C).

We evaluated the post-transcriptional gene expression changes of biochemically identified miRNA targets by comparing exonic read changes (Δexon, defined as $\log_2$ Exon$_{WT}$/Exon$_{KO}$) and intronic read changes (Δintron, defined as $\log_2$ Intron$_{WT}$/Intron$_{KO}$) between Dicer WT and KO cells. This approach was recently introduced (Gaidatzis et al., 2015) and uses a generalized linear model using DESeq2 (Love et al., 2014) to assess the changes between exonic and intronic reads in the same sample (see Experimental Procedures). Genes with a greater difference between exonic and intronic changes are likely to be altered post-transcriptionally; therefore, this metric

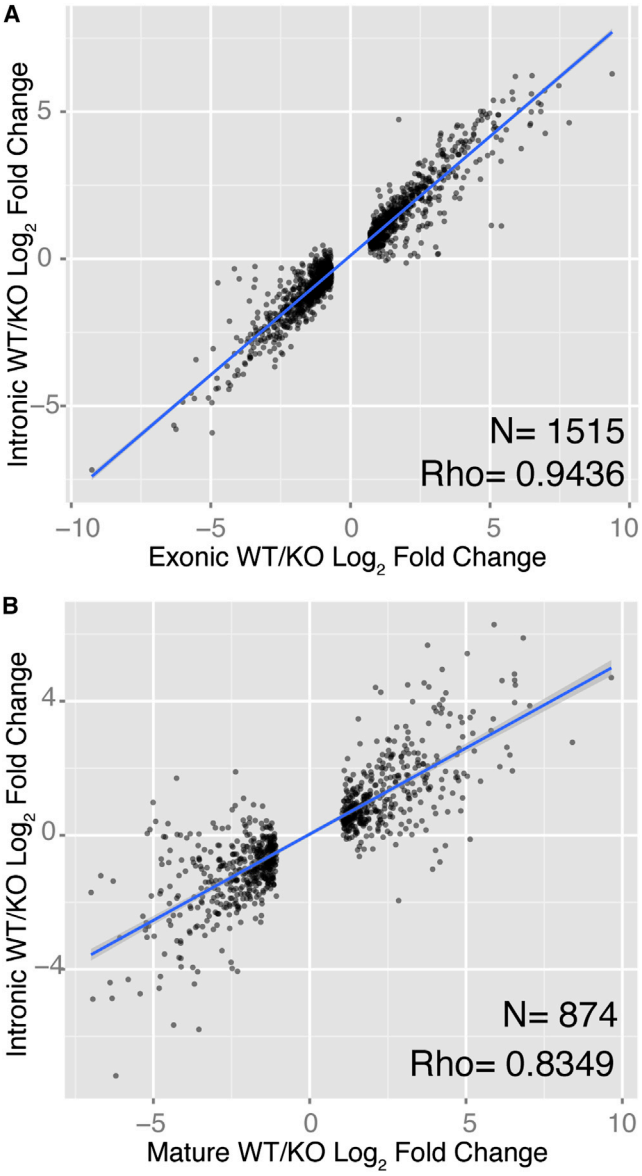


Figure 1. Transcription Drives Gene Expression Changes following miRNA Loss

(A) Log_2 fold change of genes that exhibit significant ($q < 0.05$) change in ribo-depleted exon-aligned reads from WT versus KO cells (x axis) compared to fold change of intronic reads aligned to same genes (y axis); regression line drawn in blue.

(B) Log_2 fold change of poly(A)-collected reads from WT versus KO cells (x axis) from significantly ($q < 0.05$) changing genes compared to changes in reads aligned to introns from the ribo-depleted libraries (y axis).

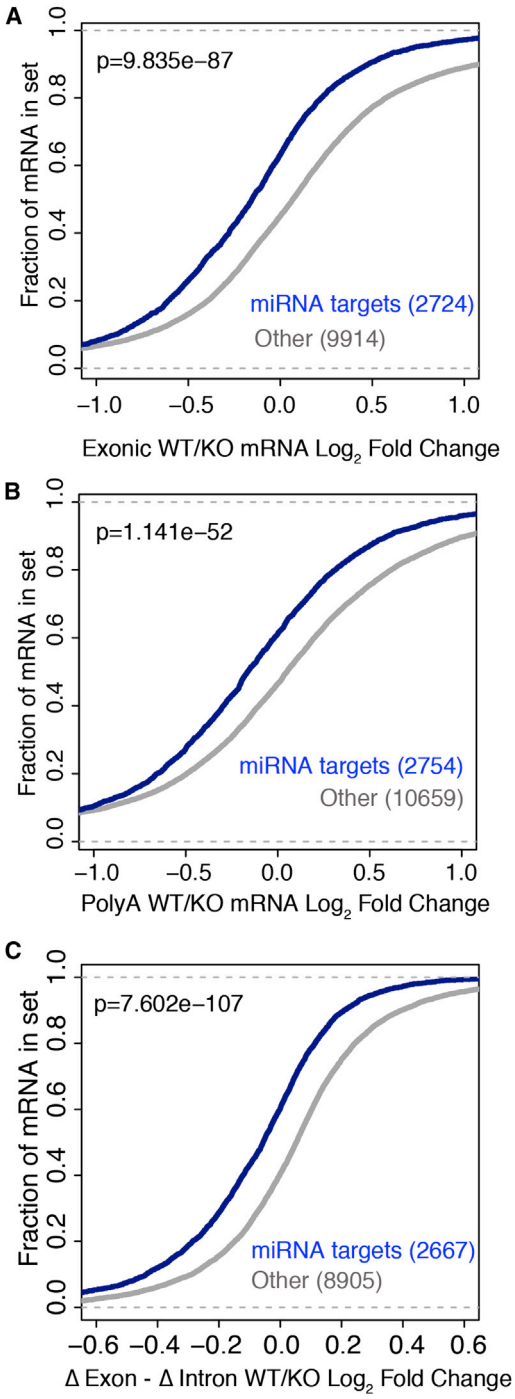


Figure 2. CDFs of miRNA Targets upon Global miRNA Loss

(A) Mature mRNA expression changes according to exonic reads of total RNA of direct miRNA targets (blue) compared to non-targets (gray) upon Dicer KO.

(B) Mature mRNA expression of direct miRNA targets by poly(A)-tagged mRNA.

(C) Δexon − Δintron values of direct miRNA targets representing differences in exonic measurements versus intronic measurements of miRNA targets (blue) compared to non-targets (gray).

can be used to assess miRNA-mediated repression of transcripts. mRNAs that are post-transcriptionally repressed by miRNAs will exhibit greater repression at the exonic level compared to the intronic level causing the Δexon − Δintron values of these genes to be negative. As expected, Δexon − Δintron values of genes that are miRNA targets are significantly ($p = 7.60e-107$) more negative than non-targets, as depicted in

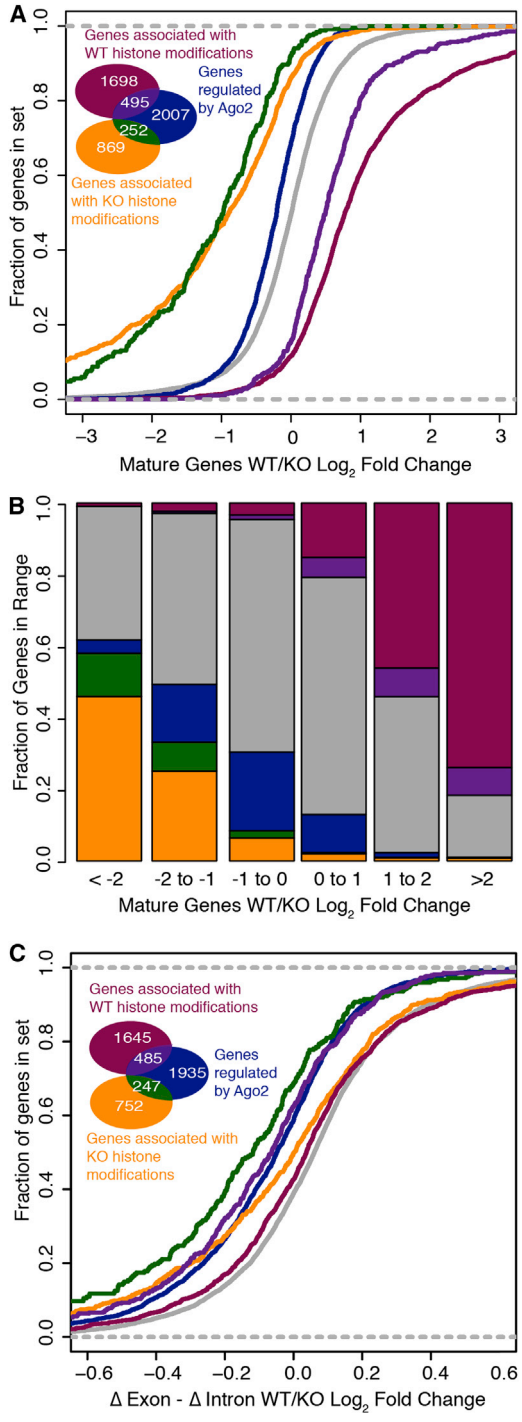


Figure 3. Histone Marks Illustrate Collaboration between Transcriptional and Post-transcriptional Regulatory Modes with Distinct Impacts of Gene Expression

(A) Cumulative distribution plot of groups of genes defined by their mode of regulation, with colors and counts indicated by inset Venn diagram; gray represents genes without evidence of transcriptional or miRNA regulation.

(B) Fraction of genes within each fold change bracket belonging to each class.

(C) Distribution of Δexon − Δintron $\log_2$ fold change values of genes according to mode of regulation.

Figure 2C. This shift confirms the post-transcriptional effect of global Dicer loss on gene expression of direct miRNA targets.

Epigenetic Data Integration Identifies Transcriptional Regulatory Changes

To identify genes that were transcriptionally modulated upon Dicer deletion, we measured histone modifications, which are altered during transcription factor activity (Ernst and Kellis, 2010). We analyzed reads from previously collected histone 3 lysine 4 tri-methylation (H3K4me3) marks (Gurtan et al., 2013), present on active promoters, and histone 3 lysine 36 tri-methylation (H3K36me3) marks (Gurtan et al., 2013), present on active gene bodies. Additionally, we collected data from histone 3 lysine 27 acetylation (H3K27ac), a mark associated with transcriptional promoters and enhancers (Creyghton et al., 2010). We used all three marks to identify regions that showed significant ($q < 0.05$ for each mark) enrichment in WT (WT specific) or KO (KO specific) cells (see Experimental Procedures; Tables S3A–S3F). By pairing changes in histone marks to nearby genes (see Experimental Procedures), we identified genes with gain or loss of transcriptional activity in the KO (1,187 and 2,259 genes respectively; Figures S3A and S3B). After eliminating the 66 genes that showed evidence of gain of activation with one mark and loss of activation with another mark, we found a total of 3,314 genes with evidence of altered transcriptional regulation, representing ~25% of the total number of expressed genes. We confirmed that each of the three histone marks represent changes in transcriptional activity by measuring correlations between changes in histone modifications and mature mRNA expression of proximal genes (Pearson's r = 0.61, 0.68, and 0.79 for H3K4me3, H3K27ac, and H3K36me3, respectively), shown in Figures S3C–S3E, as well as intronic expression of proximal genes, shown in Figures S3F–S3H (Pearson's r = 0.56, 0.67, and 0.66 for H3K4me3, H3K27ac, and H3K36me3 respectively).

Then, we compared impact of transcriptional regulation with post-transcriptional regulation by dividing the gene population according to its mode of regulation (transcriptional, post-transcriptional, or both; Figures 3A and 3C, insets) and then computing the cumulative distribution functions (CDFs) of the mRNA log fold change of each set. The CDFs of all five sets of genes are depicted in Figure 3A, together with the genes that show no evidence of either transcriptional or post-transcriptional regulation (Figure 3A, gray curve). While genes regulated only post-transcriptionally exhibit a statistically significant shift in distribution (Figure 3A, blue curve; p = 1.64e-63), much greater shifts were observed for the CDFs of genes that are activated only transcriptionally (Figure 3A, yellow curve) or are both regulated post-transcriptionally and activated transcriptionally (Figure 3A, green curve). Approximately 60% of mRNAs exhibiting a >4-fold increase in expression in the KO cells show evidence of transcriptional activity (Figure 3B, yellow and green bars), which dwarfs the impact of miRNAs, whose targets constitute fewer than 5% of genes showing a >4-fold increase in expression (Figure 3B). Thus, transcriptional changes are far greater in both magnitude and number than post-transcriptional changes.

Lastly, we also compared Δexon − Δintron measurements changed among genes that were transcriptionally regulated and

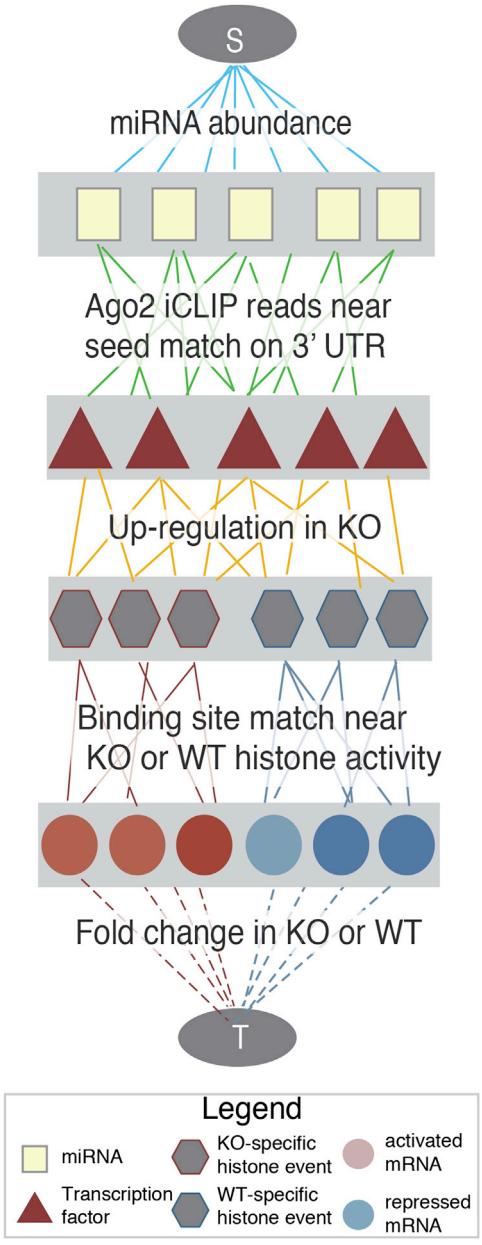


Figure 4. Implementation of Hierarchical Network Algorithm
The network structure weights edges according to five types of data representing changes incurred upon miRNA loss. Legend is inset. S, source; T, sink.

post-transcriptionally regulated. As we described earlier, genes that are post-transcriptionally repressed in the WT will have lower Δexon values than Δintron values, which would cause the distribution of Δexon − Δintron values to be more negative. We plotted these values in cumulative distribution in Figure 3C for each of the same groups of genes described in Figure 3A. Indeed, genes with evidence of iCLIP activity without transcriptional activity (Figure 3C, blue curve) exhibit a negative shift in cumulative distribution of Δexon − Δintron values compared to genes without any evidence of transcriptional or post-transcriptional regulation (Figure 3C, gray curve). Additionally, we see changes in Δexon − Δintron between genes that are both transcriptionally and post-transcriptionally regulated (Figure 3C, green and purple curves) compared to those that are only transcriptionally regulated (Figure 3C, yellow and magenta curves). The distinct distribution of genes that are co-regulated by miRNAs and transcription factors both in Figure 3C and Figure 3A (green and purple curves) underscore the importance of characterizing the transcriptional regulatory changes that occur downstream of miRNA perturbation, as miRNA and mRNA measurements alone fail to fully characterize the impact that miRNAs can have on regulatory networks. Measuring epigenetic changes that occur upon global Dicer loss greatly increases the ability to characterize the broader impacts of post-transcriptional regulation.

Hierarchical Network Algorithm Integrates All Data to Characterize Transcriptional Programs Activated upon miRNA Loss

We then used the epigenetic information provided by the histone marks to enumerate the transcriptional regulatory network activated upon miRNA perturbation. Specifically, we built an algorithm that explicitly modeled transcriptional activity in a network framework together with miRNA abundance and binding activity to identify which miRNA-regulated transcription factors best explain the observed global expression changes. The algorithm consists of two primary steps: assembling the diverse high-throughput data into a graph (Figure S4) and reducing the graph to the smallest set of nodes and edges that best explains the observed data (see Experimental Procedures for details).

The graph structure, summarized in Figure 4, consists of nodes and edges that represent the individual datasets measuring changes between WT and KO fibroblasts. The nodes of the graph represent miRNAs (squares), transcription factors (triangles), predicted transcription factor binding regions (hexagons), mRNA (circles), and two dummy nodes dubbed the "Source" (S) and the "Sink" (T). Each edge is weighted by the likelihood of an interaction between two of the nodes in the network, and every possible path between the source and the sink represents a putative way in which miRNAs can affect mRNA changes (as measured at intronic level; see Experimental Procedures; Figure S4). Thus, if a miRNA affects transcription of an mRNA, a green edge is shown between that miRNA and a transcription factor, a gold edge is shown between that transcription factor and a binding site upstream of the gene that encodes the mRNA, and a red or blue edge is shown between the binding site and the mRNA.

To reduce the space of thousands of putative interactions between miRNAs, mRNAs that encode transcription factors, DNA-binding proteins, and DNA binding sites, we applied a graph reduction step that uses the SAMNet constrained optimization algorithm (Gosline et al., 2012) to select the minimum number of edges in the graph that connect the source to the sink while ensuring to select the combination of edges with the highest total weight. SAMNet uses a "network flow" approach that attempts to find the best path from the Source node to the Sink node using the fewest total edges while maximizing the sum of the weight on all the edges. Once a suitable solution is found from the source to the sink, no additional nodes are selected.

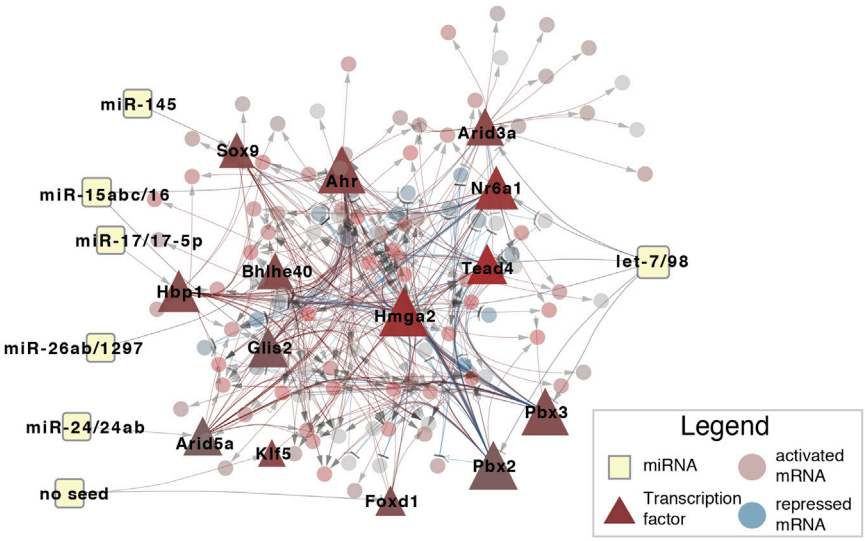


Figure 5. The Predicted Network Implicates 14 Transcription Factors in Activated mRNA and Repressed mRNA Downstream of Expressed miRNAs

Transcription factors are indicated by triangles; activated RNA is indicated by red circles; repressed mRNA is indicated by blue circles; and expressed miRNAs are indicated by squares. Legend is inset.

The resulting network, depicted in Figure 5, maps a subset of the observed intronic RNA changes (85 activated genes and 26 repressed genes) via six miRNAs and 14 distinct transcription factors. Given the algorithmic goal of minimizing the selection of nodes and edges while maximizing total weight, only the mRNAs that exhibit the largest absolute fold change are selected. As such, we focused our analysis on selected transcription factors (Figure 5, triangles), their predicted number of targets (indicated by size), and upregulation in the KO (indicated by the degree of red coloring), described in Table S4. As a control to address the possibility that the Argonaute protein complex can affect mRNA expression without a precise miRNA seed match (Chi et al., 2012), we allowed the algorithm to consider the possibility that a transcription factor can be repressed without the presence of an exact seed match ("No Seed" in Figure 5). In this analysis, we identified two putatively active transcription factors, Foxd1 and Klf5, but each had few of their own targets predicted (Table S4), suggesting that these factors are less biologically relevant than those with seed evidence of miRNA binding. The complete list of transcription factors agrees with previous miRNA-transcription factor studies: let-7 represses Hmga2 (Lee and Dutta, 2007; Mayr et al., 2007), and Nr6a1 (Gurtan et al., 2013) and miR-145 has been shown to regulate Sox9 (Rani et al., 2013).

To validate the robustness of the algorithm to the method by which miRNA targets were selected, we explored the possibility of applying the network approach with miRNA target predictions rather than iCLIP data. Given the steady improvements in the accuracy of miRNA target prediction tools such as TargetScan (Agarwal et al., 2015) and the difficulty of executing the iCLIP protocol, it was important to evaluate the performance of the network algorithm using computational predictions. To do so, we applied the network approach using the TargetScan 6.2 mouse miRNA Context+ scores as weights on the edges between miRNA and transcription factor nodes (see Experimental Procedures). The resulting network is depicted in Figure S5. The network identified 28 transcription factors regulated by seven miRNAs. Of these 28 transcription factors, 10 were found in the original iCLIP-derived network (that predicted only 14 transcription factors; Figure 5). The ten common transcription factors include those that were experimentally validated, as discussed later. The increased number of transcription factors in the network using TargetScan is likely due to a combination of false-negatives in the iCLIP data and false-positives in the TargetScan predictions.

Model Assessment and Validation

To assess the predictions made by the model, we applied both a computational approach and an experimental approach. The computational approach ensured that the predictions made by the algorithm were due to the experimental data and not based on other biases in the prediction algorithm. The experimental validation showed that the transcription factors selected can partially recover the transcriptional changes observed upon Dicer deletion.

To computationally assess the predictions, we re-ran the algorithm on 1,000 different graphs, each graph comprising the same nodes as those in the original network but with shuffled weights on each of the edges (see Experimental Procedures). If a node in the graph that represents a transcription factor is frequently identified in a random network, it reduces our overall confidence that the transcription factor is truly represented by the data. Therefore, we favor transcription factors that show up with lower frequency in the random networks. The frequency of each transcription factor selected by the model is shown in black in Figure 6A. These results suggest that the transcription factors selected by the network algorithm are specific to the experimental data, as they all appear in less than 5% of the graphs in which data were perturbed.

We also used the network randomization to assess precisely which type of data contributed to each transcription factor prediction. To do this, we perturbed only one type of data by randomizing the edge weights and then measured how frequently the resulting transcription factors appeared in the random networks. For example, when only miRNA-targeting data are randomized, the network algorithm cannot predict most transcription factors with a high degree of accuracy (Figure 6A, orange points), suggesting that miRNA target information is critical to these predictions. When other types of data are perturbed, however, nodes are still selected by the network,

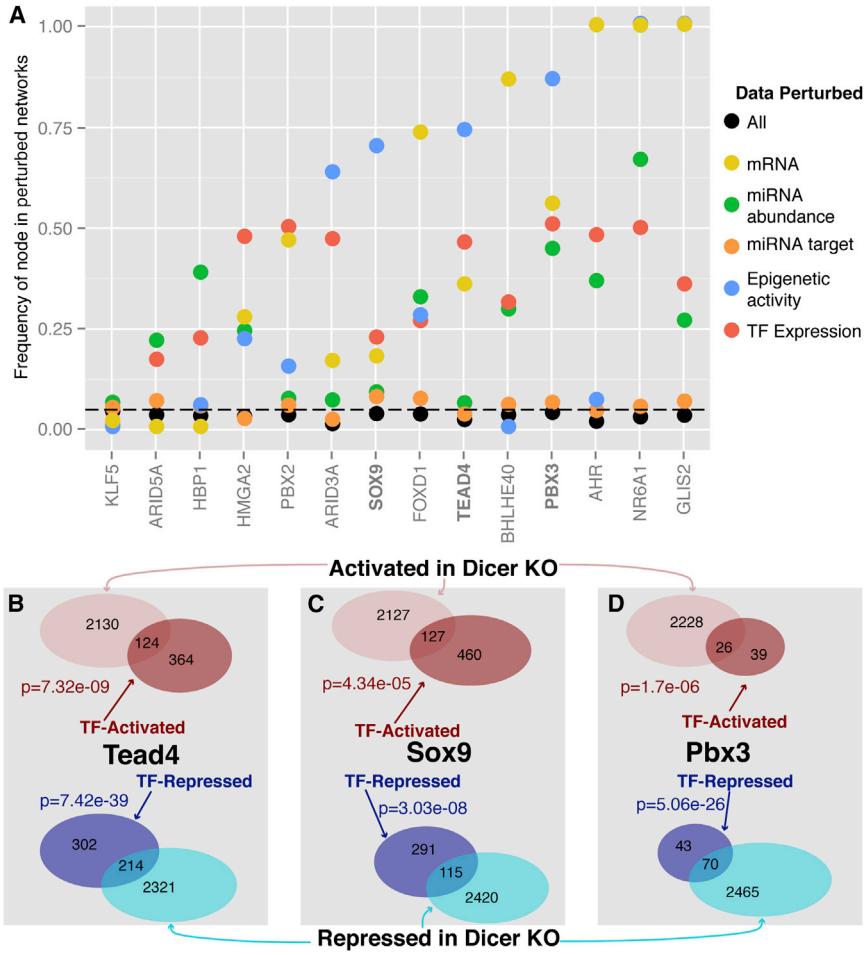


Figure 6. Computational and Experimental Validation of Selected Transcriptional Factors

(A) Results of computational network perturbation indicating how frequently the algorithm-selected transcription factors showed up when either individual (colored) data sources were perturbed or all (black) data sources were perturbed.

(B–D) Genes with significant ($q < 0.05$) changes in intronic RNA levels that are activated (red) and repressed (blue) upon overexpression of Flag-HA-Tead4 (B), Flag-HA-Sox9 (C), and Flag-HA-Pbx3 (D) are significantly enriched in genes with intronic $\log_2$ RNA changes >0.5 (pink) or < −0.5 (cyan) in the Dicer KO. The p values were computed via Fisher's exact test.

suggesting that the predictions in the final network do not rely on each type of data equally. Specific examples of this include predictions of Ahr, Nr6a1, and Glis2, which rely heavily on miRNA targeting data (Figure 6A, orange points) but not on mRNA expression data (Figure 6A, yellow points). Therefore, we can use this data-specific perturbation to assess the quality of the predictions made in the final network.

To confirm the efficacy of the network algorithm in identifying transcription factors that regulate miRNA-mediated response, we validated specific transcription factors supported by the computational model and subsequent randomizations. Specifically, we selected Pbx3, Tead4, and Sox9 for overexpression in Dicer WT cells to observe changes in gene expression. We transduced Dicer WT cells with N-terminally Flag-HA-tagged retroviral expression constructs for Tead4, Sox9, Pbx3, or an empty vector negative control. We isolated ribo-depleted total RNA and carried out RNA-seq in duplicate for each construct (see Experimental Procedures; Figure S6A). Then, we measured changes in intronic regions to assay transcriptional changes (Table S6) without the confounding post-transcriptional effect of endogenous miRNAs (Khan et al., 2009) and then compared the genes that were activated and repressed in each of the overexpression experiments to those genes activated and repressed in the Dicer KO cells. We found the overlap to be statistically significant according to Fisher's exact test, as shown in Figures 6B–6D, confirming that each transcription factor significantly contributed to the transcriptional changes measured in the Dicer KO cells.

DISCUSSION

In this work, we present a comprehensive approach to deconvoluting the impact of miRNAs on gene expression by the identification of miRNA-regulated transcription factors. We use ribo-depleted RNA-seq, poly(A) mRNA-seq, small RNA-seq, iCLIP, and histone chromatin immunoprecipitation sequencing (ChIP-seq) to delineate the impact of global miRNA loss on both transcriptional and post-transcriptional regulation. Changes in mRNA expression levels are highly correlated with changes in total RNA-seq reads mapping to introns, indicating that most genes that change in expression after Dicer loss are altered transcriptionally. We showed experimentally that the magnitude of changes caused by transcription, as indicated by epigenetic measurements of histone marks, is greater than changes caused by miRNA-mediated repression. Then, we introduced a robust computational method to identify the transcription factors that explain these transcriptional changes downstream of miRNA loss and experimentally validate three of these transcription factors as amplifying the effects of miRNAs.

Given the pronounced role of transcription factors in mediating miRNA-mediated effects, the graphical modeling approach introduced here enables reverse engineering of the regulatory network from gene expression and epigenetic data. This approach advances the field of miRNA analysis by leveraging valuable epigenetic data to deconvolute the pleiotropic effects of miRNAs and filter for miRNAs consequential to gene expression. By incorporating miRNA activity upstream and intronic RNA changes downstream of transcription, our approach also builds upon transcription factor prediction tools that use

epigenetic and expression data (Foat et al., 2006; Sherwood et al., 2014). The algorithm presented here is flexible and can be applied widely to any matched miRNA/mRNA/epigenetic such as those in large repositories such as the NIH Roadmap Epigenomics Mapping Project (Bernstein et al., 2010) or ENCODE (Rosenbloom et al., 2013), together with miRNA target prediction algorithms.

miRNAs have now been implicated in a stunningly wide range of biological processes and diseases (Zadran et al., 2013; Chen et al., 2004) and lead to large global changes in mRNA expression (Garcia et al., 2011) while causing only moderate repression of most direct targets. This study demonstrates that decoupling transcriptional changes from post-transcriptional changes and integrating them with epigenetic alterations in a computational framework can elucidate the transcriptional network that tunes and amplifies the effect of miRNA loss. The computational framework introduced here may benefit studies of miRNAs by shifting emphasis to the rewired transcriptional networks that cause the majority of the transcript-level changes.

EXPERIMENTAL PROCEDURES

miRNA Target Identification

iCLIP reads were collected from GSE45828 and aligned to mm9 using Bowtie (Langmead and Salzberg, 2012). iCLIP events were assigned significance using GEM v1.1 (Guo et al., 2012), together with a custom read distribution derived from reads surrounding let-7 binding sites. Significant ($q < 10e-5$) events were then filtered for the presence of a 7-mer or 8-mer seed match of a miRNA family that represented at least 1% of the reads from the small RNA population of the Dicer WT cells from GSE44156 as described in the Supplemental Experimental Procedures. To compare the effect of iCLIP-defined targets, we used Context+ scores from TargetScan 6.2 (Garcia et al., 2011) from http://www.targetscan.org/mmu_61/.

RNA Expression Measurements

This work included a total of six RNA-seq datasets, each collected in duplicate. Mature mRNA was collected from untreated KO and WT fibroblasts via traditional poly(A)-collected library preparation, and total RNA was collected from the same cells using the Illumina Ribo-Zero Kit. Total RNA was also collected from Dicer WT fibroblasts transduced retrovirally with vector control, Flag-HA-PBX3, Flag-HA-TEAD4, or Flag-HA-SOX9 (see Supplemental Experimental Procedures). DESeq v1.10.1(Anders and Huber, 2010) was used for all data normalization and differential expression calls—a minimum of two DESeq-normalized reads (in both conditions) was required to call a gene expressed. The poly(A) data contained 13,413 genes expressed in both KO and WT cells, while the total RNA data contained 12,638 genes in both genotypes at the exonic level and 14,487 genes in both genotypes at the intronic level. qPCR measurements of seven genes were used to confirm the exon and intron reads from the total RNA measurements, described in the Supplemental Experimental Procedures. Δexon − Δintron values were computed using DESeq 2 (Love et al., 2014) according to the exon-intron split analysis (EISA) method previously described (Gaidatzis et al., 2015).

Histone Data Collection and Event Calling

ChIP assays for the H3K27ac mark were performed as previously described (Macisaac and Fraenkel, 2010), and other marks were collected from previously published data (Gurtan et al., 2013), available under GEO accession GSE: 44159. For H3K27ac and H3K4me3 marks, custom read distributions were used to call significant ($q = 0.05$) events between WT and KO marks using GEM v1.1 (Guo et al., 2012), as described in the Supplemental Experimental Procedures while the default read distribution was used for the H3K36me3 marks. H3K4me3 and H3K27ac marks were associated with a gene if they fell within 10 kb of a transcription start site, while H3K36me3 marks were associated with a gene if they fell within the gene body. See the Supplemental Experimental Procedures for more details.

Network Integration

Small RNA expression levels, iCLIP binding levels, mature mRNA expression levels, ChIP-seq binding data, and intronic RNA expression changes were encoded in a graphical model depicted in Figures 4 and S4 that was reduced using a version of the SAMNet algorithm (Gosline et al., 2012), as described in great detail in the Supplemental Experimental Procedures. The Garnet module of the Omics Integrator package (http://fraenkel.mit.edu/omicsintegrator) was used to predict transcription factor binding sites using the histone data. Network inputs and additional details are described in the Supplemental Experimental Procedures, and the code used to implement the algorithm is freely available at http://github.com/sgosline/topaz.

Experimental Validation and Target Identification

N-terminally Flag-HA-tagged Pbx3, Sox9, and Tead4 were PCR amplified from mouse cDNA generated from Dicer KO fibroblasts. Transduced cells were sequenced in duplicate, together with a vector control, and DESeq v1.1 was used to compare intronic reads between conditions. Genes that were significantly ($p < 0.05$) upregulated upon transfection that were also up-regulated in the Dicer KO cells were considered to be activated by the transcription factor. Genes that were significantly ($p < 0.05$) downregulated upon transcription that were also downregulated in the Dicer WT were considered repressed. Details are described in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Data presented in this study are available under the GEO SuperSeries accession number GEO: GSE61035. Total RNA-seq data are available under the accession number GEO: GSE61033, the H3K27ac data are available under the accession number GEO: GSE61034, the poly(A) RNA-Seq data are available under the accession number GEO: GSE61031. This work also references previously collected ChIP-seq data with the accession number GEO: GSE44159 and previously collected iCLIP data with the accession number GEO: GSE45828.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.12.031.

AUTHOR CONTRIBUTIONS

S.J.C.G., A.M.G., E.F., and P.A.S. designed the study. S.D. performed ChIP-seq. C.K.J. performed over-expression experiments. A.B. performed the iCLIP. P.M., B.J.M., and Y.S.Y. prepared various samples for sequencing. Sequencing was performed at the BioMicroCenter at MIT. S.J.C.G. performed all the analyses. S.J.C.G., A.M.G., E.F. and P.A.S. wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank members of the E.F. and P.A.S. labs for useful discussions about this work. The work was supported by the NIH via grants U54-CA112967 (E.F.), R01-GM089903 (E.F.), U01-CA184898 (E.F.), R01-CA133404 (P.A.S.), PO1-CA042063 (P.A.S.) and RO1-GM34277 (P.A.S.) and partially by the Koch Institute Support (core) grant P30-CA14051 from the National Cancer Institute. Computing resources were funded by the National Science Foundation under Award No. DB1-0821391, and A.M.G. was supported by the Leukemia and Lymphoma Society grant 5198-09.

Received: February 20, 2015
Revised: October 16, 2015
Accepted: December 3, 2015
Published: December 31, 2015

REFERENCES

Afshar, A.S., Xu, J., and Goutsias, J. (2014). Integrative identification of deregulated miRNA/TF-mediated gene regulatory loops and networks in prostate cancer. PLoS ONE *9*, e100806. http://dx.doi.org/10.1371/journal.pone.0100806.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, e05005. http://dx.doi.org/10.7554/eLife.05005.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106. http://dx.doi.org/10.1186/gb-2010-11-10-r106.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64–71. http://dx.doi.org/10.1038/nature07242.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048. http://dx.doi.org/10.1038/nbt1010-1045.

Bisognin, A., Sales, G., Coppe, A., Bortoluzzi, S., and Romualdi, C. (2012). MAGIA^2: from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). Nucleic Acids Res. *40*, W13–W21. http://dx.doi.org/10.1093/nar/gks460.

Bosson, A.D., Zamudio, J.R., and Sharp, P.A. (2014). Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol. Cell *56*, 347–359. http://dx.doi.org/10.1016/j.molcel.2014.09.018.

Chen, C.-Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83–86. http://dx.doi.org/10.1126/science.1091903.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486. http://dx.doi.org/10.1038/nature08170.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. Nat. Struct. Mol. Biol. *19*, 321–327. http://dx.doi.org/10.1038/nsmb.2230.

Chiu, H.-S., Llobet-Navas, D., Yang, X., Chung, W.-J., Ambesi-Impiombato, A., Iyer, A., Kim, H.R., Seviour, E.G., Luo, Z., Sehgal, V., et al. (2015). Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. Genome Res. *25*, 257–267. http://dx.doi.org/10.1101/gr.178194.114.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936. http://dx.doi.org/10.1073/pnas.1016071107.

Cuellar-Partida, G., Buske, F.A., McLeay, R.C., Whitington, T., Noble, W.S., and Bailey, T.L. (2012). Epigenetic priors for identifying active transcription factor binding sites. Bioinformatics *28*, 56–62. http://dx.doi.org/10.1093/bioinformatics/btr614.

Du, N.-H., Arpat, A.B., De Matos, M., and Gatfield, D. (2014). MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. eLife *3*, e02510.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825. http://dx.doi.org/10.1038/nbt.1662.

Foat, B.C., Morozov, A.V., and Bussemaker, H.J. (2006). Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. Bioinformatics *22*, e141–e149. http://dx.doi.org/10.1093/bioinformatics/btl223.

Friard, O., Re, A., Taverna, D., De Bortoli, M., and Corá, D. (2010). CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. BMC Bioinformatics *11*, 435. http://dx.doi.org/10.1186/1471-2105-11-435.

Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. *33*, 722–729. http://dx.doi.org/10.1038/nbt.3269.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146. http://dx.doi.org/10.1038/nsmb.2115.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100. http://dx.doi.org/10.1038/nature11245.

Gosline, S.J., Spencer, S.J., Ursu, O., and Fraenkel, E. (2012). SAMNet: a network-based approach to integrate multi-dimensional high throughput datasets. Integr. Biol. (Camb.) *4*, 1415–1427. http://dx.doi.org/10.1039/c2ib20072d.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105. http://dx.doi.org/10.1016/j.molcel.2007.06.017.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840. http://dx.doi.org/10.1038/nature09267.

Guo, Y., Mahony, S., and Gifford, D.K. (2012). High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLoS Comput. Biol. *8*, e1002638. http://dx.doi.org/10.1371/journal.pcbi.1002638.

Gurtan, A.M., and Sharp, P.A. (2013). The role of miRNAs in regulating gene expression networks. J. Mol. Biol. *425*, 3582–3600. http://dx.doi.org/10.1016/j.jmb.2013.03.007.

Gurtan, A.M., Ravi, A., Rahl, P.B., Bosson, A.D., JnBaptiste, C.K., Bhutkar, A., Whittaker, C.A., Young, R.A., and Sharp, P.A. (2013). Let-7 represses Nr6a1 and a mid-gestation developmental program in adult fibroblasts. Genes Dev. *27*, 941–954. http://dx.doi.org/10.1101/gad.215376.113.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112. http://dx.doi.org/10.1038/nature07829.

Herranz, H., and Cohen, S.M. (2010). MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. Genes Dev. *24*, 1339–1344. http://dx.doi.org/10.1101/gad.1937010.

Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nat. Biotechnol. *27*, 549–555. http://dx.doi.org/10.1038/nbt.1543.

König, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2012). Protein-RNA interactions: new genomic technologies and perspectives. Nat. Rev. Genet. *13*, 77–83. http://dx.doi.org/10.1038/nrg3141.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. http://dx.doi.org/10.1038/nmeth.1923.

Lee, Y.S., and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev. *21*, 1025–1030. http://dx.doi.org/10.1101/gad.1540407.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. http://dx.doi.org/10.1186/PREACCEPT-8897612761307401.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834–838. http://dx.doi.org/10.1038/nature03702.

Macisaac, K.D., and Fraenkel, E. (2010). Sequence analysis of chromatin immunoprecipitation data for transcription factors. Methods Mol. Biol. *674*, 179–193. http://dx.doi.org/10.1007/978-1-60761-854-6_11.

Mayr, C., Hemann, M.T., and Bartel, D.P. (2007). Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science *315*, 1576–1579. http://dx.doi.org/10.1126/science.1137999.

Mendell, J.T., and Olson, E.N. (2012). MicroRNAs in stress signaling and human disease. Cell *148*, 1172–1187. http://dx.doi.org/10.1016/j.cell.2012.02.005.

Naeem, H., Küffner, R., and Zimmer, R. (2011). MIRTFnet: analysis of miRNA regulated transcription factors. PLoS ONE *6*, e22519. http://dx.doi.org/10.1371/journal.pone.0022519.

Pasquinelli, A.E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat. Rev. Genet. *13*, 271–282. http://dx.doi.org/10.1038/nrg3162.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455. http://dx.doi.org/10.1101/gr.112623.110.

Rani, S.B., Rathod, S.S., Karthik, S., Kaur, N., Muzumdar, D., and Shiras, A.S. (2013). MiR-145 functions as a tumor-suppressive RNA by targeting Sox9 and adducin 3 in human glioma cells. Neuro-oncol. *15*, 1302–1316. http://dx.doi.org/10.1093/neuonc/not090.

Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., et al. (2013). ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. *41* (Database issue), D56–D63. http://dx.doi.org/10.1093/nar/gks1172.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). Gene expression. MicroRNA control of protein expression noise. Science *348*, 128–132. http://dx.doi.org/10.1126/science.aaa1738.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63. http://dx.doi.org/10.1038/nature07228.

Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. (2007). Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comput. Biol. *3*, e131. http://dx.doi.org/10.1371/journal.pcbi.0030131.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. *32*, 171–178. http://dx.doi.org/10.1038/nbt.2798.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb. Protoc. *2010*, pdb.prot5384. http://dx.doi.org/10.1101/pdb.prot5384.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol. *13*, R67. http://dx.doi.org/10.1186/gb-2012-13-8-r67.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128. http://dx.doi.org/10.1038/nature07299.

Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol. Cell *26*, 753–767. http://dx.doi.org/10.1016/j.molcel.2007.05.018.

Tu, K., Yu, H., Hua, Y.-J., Li, Y.-Y., Liu, L., Xie, L., and Li, Y.-X. (2009). Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. Nucleic Acids Res. *37*, 5969–5980. http://dx.doi.org/10.1093/nar/gkp638.

Wen, J., Parker, B.J., Jacobsen, A., and Krogh, A. (2011). MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action. RNA *17*, 820–834. http://dx.doi.org/10.1261/rna.2387911.

Zadran, S., Remacle, F., and Levine, R.D. (2013). miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. Proc. Natl. Acad. Sci. *110*, 19160–19165. http://dx.doi.org/10.1073/pnas.1316991110.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614. http://dx.doi.org/10.1038/nbt.1873.