Identification of microRNA Biogenesis Regulators and Activity Modulators

Wei-Jen Chung

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

of the Graduate School of Arts and Sciences

Columbia University

2014

Abstract

Identification of microRNA Biogenesis Regulators and Activity Modulators

Wei-Jen Chung

MicroRNAs play a key role in post-transcriptional gene regulation. They regulate target gene expression with mRNA degradation or translation repression. Each miRNA is estimated to regulate dozens of genes in human, and dysregulation of miRNA leads to various diseases, such as cancer, heart disease and depression. Therefore, it is critical to understand the mechanism of miRNA biogenesis and targeting. This work integrated gene and miRNA expression profile from various cancer projects to screen for potential miRNA biogenesis regulators and activity modulators. In this analysis, we identified several genes that regulate miRNA biogenesis pathway and miRNA-mediated regulation. We also found the association between these genes and tumor progression.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Five years in Columbia is a memorable chapter in my life. I have received a lot of support from so many people, and I am extremely grateful for everybody's help.

First of all, I would like to thank Andrea Califano. You are the best mentor and advisor by giving me the enormous support and freedom, so that I can do the research I love. In the past five years, I learned so much in scientific research under your inspirational and positive guidance and also learned a lot in work attitude just by watching how you work tirelessly. You are a true role model, and I will do my best to emulate you. I also would like to thank my thesis committee members: Richard Friedman, Raul Rabadan, and Gustavo Stolovitzky. Thank you for your time to participate in my thesis proposal, progress report and defense. Thank you for all the useful advises on my projects, so I can keep my progress on the right track.

I also would like to thank the lab mates in Califano lab, especially Pavel Sumazin. I received great support from you. You are always very patient to me and very accessible for discussion. I also learned so much in conducting scientific research through our discussion. Thank you! I will not forget the help I obtained from Mukesh Bansal, Brygida Bisikirska, Jorida Coku, Federico Giorgi, Archana Iyer, Presha Rajbhandari, Beatrice Salvatori, Yishai Shimoni, Yao Shen, Will Shin, Xuerui Yang, and Sean Zhou. I will always cherish the friendship between all the lab mates and me.

Finally, I want to thank my family: my wife, parents-in-law and my parents. My lovely wife, Ting-Chun, gave me huge support in every way over these years. You also gave us a beautiful daughter, Sophie. I am really blessed to have you in my life. I love you and always will. I also want to thank my parents-in-law for giving us support while

visiting us, so I can focus on my research. And to my mom and dad, thank you for encouraging me chasing my dream and support my decision. Although you are in Taiwan, I can always feel your support and love.

Chapter I

Introduction

MicroRNAs (miRNAs) have emerged as key regulators of both normal and pathologic phenotypes, including cancer (Yi, Pasolli et al. 2009). Fine-grained regulation of their biogenesis; however, is still poorly understood and only a few of their key regulators have been characterized. In order to understand the extent and specificity of miRNA-biogenesis control, as well as the role of miRNA-biogenesis regulators in tumorigenesis and cancer progression (van Kouwenhove, Kedde et al. 2011), we set out to identify these regulators and profile their targets. We developed an algorithm for genome-wide inference of miRNA-biogenesis regulators. We identified biogenesis regulator candidates by assessing the mutual information between mature miRNA expression and the expression of its transcription unit after conditioning for the expression of candidate biogenesis regulator. Thus the algorithm identifies genes whose expression correlates with deviations between mature miRNAs and their precursors. In addition to biogenesis, the miRNA targeting pathway is also extensively controlled with different mechanisms. Several regulators modulate miRNA activity by interacting with miRISC or binding to target mRNA (Lu, Getz et al. 2005). We developed an algorithm to genome-wide screen for non-sponge modulators that affect miRNA activity from validated miRNA-target interactions. The predicted modulators were experimentally validated to regulate miRNA activity via post-transcriptional mechanism.

The intragenic miRNA and its host or neighboring mRNA or the miRNAs in the same transcription unit originate from the same nascent transcript (Kim and Kim 2007).

On the other hand, the miRNA-mediated regulation in animal is mainly through mRNA degradation (Bartel 2009). Therefore, the relationships between host and intragenic miRNA, miRNAs in the same cluster, or miRNA and target mRNA will be significantly correlated theoretically. However, the post-transcriptional biogenesis regulators and the miRNA modulators affect the level of mature miRNA (Krol, Loedige et al. 2010), and this suggests that the correlation between the host and intragenic miRNA, miRNAs in one cluster, or miRNA and target mRNA will be affected. The hypothesis is that we can identify genes whose expression profiles explain the discrepancy between the relationships, and the genes are involved in the pathway.

MiRNAs participate in various biological pathways by post-transcriptionally suppress target genes. Their maturation and targeting pathways are finely regulated by post-transcriptional regulators (Siomi and Siomi 2010). Currently, all of the known miRNA biogenesis regulators and function modulators were identified individually with experiments, such as immuno-precipitation for physical evidence and silencing for functional study. This project is the first genome-wide effort to screen for the factors that are involved in miRNA biogenesis and targeting. Furthermore, we showed these candidate miRNA regulators associate with tumorigenesis and progression of glioblastoma through regulating miRNA maturation and function.

We developed two algorithms, MIRAGE and Hermes, to identify miRNA biogenesis regulators and activity modulators, respectively. To identify miRNA biogenesis regulators, we first summarized intragenic miRNAs and host genes, or miRNA clusters with genomic coordinates, expression profiles and additional evidences. Secondly, MIRAGE predicted candidate biogenesis regulators by finding genes whose

expression is correlated with deviation in co-expression between miRNA and host using conditional mutual information. The candidate genes from all host-miRNA pairs were integrated by Fisher's method (Fisher 1925). Finally, we selected several candidate biogenesis regulators for siRNA silencing or over-expression. The miRNA expression level was then measured by real-time PCR or array-based technology, such as Wafergen and Fluidigm.

To identify miRNA function modulators, we collected miRNA-target interactions validated by low-throughput method, such as western blotting and luciferase assay. Then, Hermes screened for miRNA functional modulators by using conditional mutual information. We chose several candidates for silencing or over-expression; then measure the expression level of target mRNAs. In addition, we found EGFR, FYN and TRIM29 possessing the same of mode of action from two independent data sets.

Chapter 2 gives the introduction and review of current understanding about miRNA. It starts from the history of discovering miRNA followed by the mechanism of miRNA-mediated regulation and the control of miRNA maturation and activity from literatures. One main focus in miRNA field is to identify mRNAs targeted by miRNAs. This chapter also covers the current computational and experimental efforts to transcriptome-wide identify miRNA targets. Finally, we briefly reviewed the association of miRNAs and tumorigenesis, focusing on glioblastoma multiforme.

Chapter 3 described the identification of miRNA promoter by integrating chromHMM data from ENCODE and CAGE-seq data from FANTOM (Ernst and Kellis 2010; Ernst, Kheradpour et al. 2011). Chromatin profiling has emerged as an important

tool for genome annotation and detection of regulatory events. ENCODE carried out the profiling of histone modifications and CTCF binding sites, which all associate with different chromatin states. This chapter presents the list of intragenic miRNAs with independent promoters. The correlation between intragenic miRNAs and their hosts from different contexts were also summarized.

Chapter 4 demonstrated the computational screening of miRNA post-transcriptional regulators using gene and miRNA expression profile from TCGA glioblastoma data set (Lambertz, Nittner et al. 2010), then followed by functional and biochemical validation. Additionally, we examined the role of these candidates in cell proliferation and patient prognosis.

Chapter 5 described the analysis of miRNA activity modulators using validated miRNA-target interactions and gene/miRNA expression profile from TCGA breast cancer, glioblastoma and METABRIC (Melo, Ropero et al. 2009; Paroo, Ye et al. 2009). We also showed the functional study of selected candidate miRNA modulators and implicated their role in tumorigenesis.

Chapter II

Literature Review

In 1993, the first miRNA, lin-4, was identified in *C. elegans* by screening for mutants that cause temporal development deficiency (Lee, Feinbaum et al. 1993). However, lin-4 was unable to be identified in closely related species because of the lack of genomic information and unsuccessful screening. Therefore, the regulation of lin-4 was considered to be a *C. elegans* specific mechanism, until a highly conserved miRNA, let-7, and its orthologs in fly, mouse and human were identified in 2000 (Pasquinelli, Reinhart et al. 2000). In 2001, dozens of miRNAs were cloned from *C. elegans*, and miRNA started to be regarded as a new class of small RNA (Lagos-Quintana, Rauhut et al. 2001). Due to the special secondary structure of miRNA precursors, people predicted hundreds of miRNA from various model organisms. Later on, with the help of next generation sequencing, miRNAs were confirmed and identified with this technology. Currently, miRNAs have been identified in plants, animals and viruses, and there are more than 1400 miRNA genes have been identified in human (Kozomara and Griffiths-Jones 2011).

The first miRNA was discovered by screening mutant in *C. elegans* (Lee, Feinbaum et al. 1993). Eight years later, right after the first conserved miRNA, let-7, was identified, the first sets of miRNAs in worm, fly, and human were biochemically cloned (Lagos-Quintana, Rauhut et al. 2001; Newman, Thomson et al. 2008; Piskounova, Viswanathan et al. 2008). Generally, miRNAs possess several special features, including the hairpin-like secondary structure, evolutionary conservation, sequence biases in the

first five bases, the uniformed length of mature product and a tendency toward having symmetric internal loops and bulges in the miRNA region, and these features were used to predict miRNA in various organisms. Next generation sequencing was developed in the mid to late 1990s and was on market since 2004. Due to the read length (~50bp) and large quantities of data from high-throughput sequencing, scientists quickly applied this technology to identify various types of small RNAs, including miRNAs. Currently, miRBase records 24521 miRNA hairpin precursors, expressing 30424 mature miRNA products in 206 species, covering animals, plants and viruses, and human genome contains 1872 precursors, expressing 2578 mature miRNAs (Rybak, Fuchs et al. 2008).

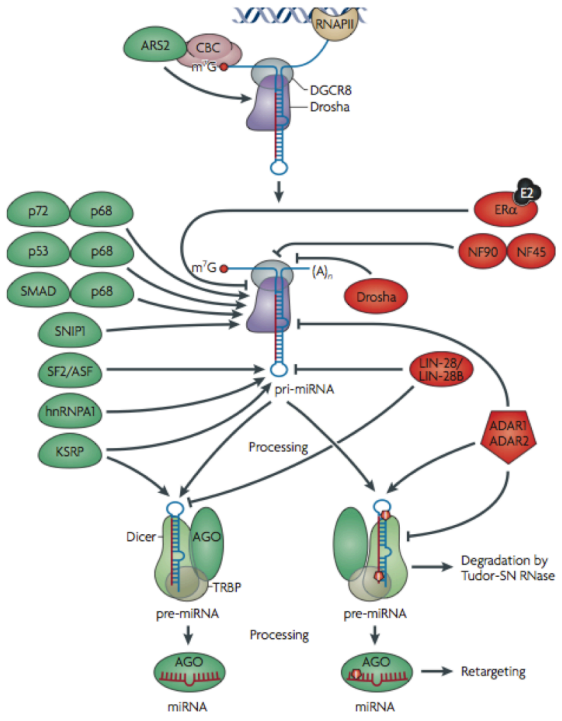
The miRNAs can be classified into intragenic and intergenic based on their genomic location: intragenic indicates the miRNA locates within the protein-coding gene, while intergenic means the miRNA is in the intergenic region. When the miRNA is expressed, the precursor will form a hairpin structure because of its sequence composition. At this step, the precursor is called “pri-miRNA”. DROSHA binds the pri-miRNA and removes the stem and the basal single strand RNA region of the hairpin. DROSHA processing creates a two-nucleotide overhang at the 3’ end, and this structure is important for binding of TARBP2, an important downstream factor (Kim, Han et al. 2009; Krol, Loedige et al. 2010; Siomi and Siomi 2010). Interestingly, DROSHA cleavage of intragenic miRNA is simultaneous with mRNA splicing (Kim and Kim 2007). DGCR8 binding at the junction of single strand RNA and hairpin facilitates DROSHA measuring the cleavage site to chop single strand RNA off (Han, Lee et al. 2006). There is another class of miRNA, called miRtron. This type of miRNAs is generated from small introns, normally short than 100 nucleotides. The entire intron can form a short hairpin

with two nucleotides overhang; therefore, it can bypass DROSHA cleavage and is transported to cytoplasm directly (Berezikov, Chung et al. 2007; Okamura, Hagen et al. 2007; Ruby, Jan et al. 2007). After splicing, instead of being degraded, certain small introns will form a hairpin structure with two-nucleotide overhang at the 3' end. MiRtron has a subtype called "tailed-miRtron", which comes from a longer intron (Flynt, Greimann et al. 2010). This type of pri-miRNA has a hairpin structure with a long tail, and the tail is trimmed off by 3'->5' exonuclease. This type of miRNA is able to bypass the DROSHA processing as well.

After DROSHA cleavage, the products are transported to cytoplasm by Exportin-5. The miRNA precursor at this step is called "pre-miRNA". TARBP2 binds the 3' overhang and recruit DICER to hairpin. DICER chops the terminal loop off and converts the pre-miRNA into miRNA duplex, which is subsequently bound by AGO2. Interestingly, not all pre-miRNAs require DICER to remove the hairpin region; pre-miR-451 use AGO2 to remove its hairpin loop region (Cheloufi, Dos Santos et al. 2010; Cifuentes, Xue et al. 2010; Yang, Maurin et al. 2010). When silencing DICER, miR-451 is the only miRNA normally processed. After DICER processing, miRNA duplex is loaded to AGO2. One strand of miRNA duplex will be kept in AGO2, while the other strand will be released. The detailed mechanism of strand selection is still not clear, but there is evidence showed the strand selection is determined by thermodynamics of pairing at both terminal of miRNA duplex. The strand kept in AGO2 is mature miRNA and has ability to regulate downstream targets. On the other hand, the released strand is called "miRNA star" (miRNA*). In general, miRNA* will be degraded, but deep-sequencing results showed many miRNA*s are not degraded, instead, they associate with AGO1 or

AGO2 and become a mature functional miRNA. These functional miRNA*s are more likely to be evolutionary-conserved (Okamura, Phillips et al. 2008). This finding showed some miRNAs are bi-functional in regulation.

In 2006, two groups showed the evidences of post-transcriptional regulation of miRNA processing. Obernosterer *et al.* first described the mature hsa-miR-138 only appears in certain cell types while its precursor is ubiquitously expressed (Obernosterer, Leuschner et al. 2006). Additionally, they found the pre-mir-138 accumulated in the cytoplasm and was unable to be processed by DICER. Secondly, Thomson *et al.* found out the mature hsa-let-7 is detectable only in the late stage in mouse development, whereas its precursor is expressed across the entire development (Thomson, Newman et al. 2006). They also found the pri-let-7 were accumulated in the nucleus and unable to be



(Krol *et al.* 2010)

Figure 2.1 Post-transcriptional regulation of miRNA pathway.

bound and processed by DROSHA while the DROSHA activity is normal. Both works suggested there is a post-transcriptional mechanism that controls miRNAs maturation.

Since then, remarkable progress has been made in identifying post-transcriptional biogenesis regulators, and DROSHA interacting proteins are the first targets. Two RNA helicases, DDX5 (p68) and DDX17 (p72), were found in DROSHA Microprocessor complex, and silencing DDX5 and DDX17 reduced one third of miRNAs' expression in mouse (Fukuda, Yamagata et al. 2007). They interact with DROSHA to stabilize the interaction between miRNA precursors and DROSHA. Heterodimer, DDX5/DDX17, also works as scaffold proteins and recruit other regulatory proteins. For example, SMAD1 and SMAD3 induce the maturation of a set of miRNAs by interacting with DDX5 (MH2 domain) and the miRNA precursors (DNA binding MH1 domain) when the BMP2/TGFα signaling pathway is activated (Davis, Hilyard et al. 2008; Davis, Hilyard et al. 2010). The miRNAs bound by SMAD1/3 (R-Smad) contain a conserved CAGAC motif, called RNA Smad binding element (R-SBE). SMAD4 (the co-Smad) does not bind to miRNA precursor directly but is able to facilitate the binding of other SMADs (Davis, Hilyard et al. 2010). SMADs directly bind to this motif and promote the miRNAs maturation (Figure 1b). KSRP is known to interact with single-strand AU-rich mRNA and plays a mediator of mRNA decay. Moreover, KSRP also serves as a component of DROSHA and DICER complex, and it promotes the maturation a subset of miRNA in nucleus and cytoplasm through binding to the terminal loop of miRNA precursors. Interestingly, the two types of miRNA biogenesis positive regulators can be phosphorylated and interact with each other. In mensenchymal C2C12 cell, phosphorylated SMAD proteins (SMAD4, SMAD5 and SMAD9) associate with

phosphorylated KSRP and block their binding to target primary miRNAs (Viswanathan, Daley et al. 2008) (Figure 2.1).

In MCF7 cell line, estrogen receptor alpha (ESR1) interacts with DDX5/DDX17 upon activated by estrogen. The interaction induces dissociation of pri-miRNAs from microprocessor or cause conformational change of microprocessor (Yamagata, Fujiyama et al. 2009). Therefore, it blocks DROSHA-mediated processing to certain pri-miRNAs, such as hsa-mir-125a and hsa-mir-145. Another example is tumor suppressor gene TP53, which stimulates the biogenesis of pri-miR-16-1, pri-miR-143 and pri-miR-145 through DDX5-mediated interaction with DROSHA complex (Suzuki, Yamagata et al. 2009). TP53 is known to regulate downstream genes through transcriptional regulation, but its modulation to pri-miRNA is transcription-independent.

In addition to DDX5 and DDX17, there are more proteins being pulled down from DROSHA-DGCR8 complex, including HNRNPH1 and HNRNPR. Immuno-precipitation showed both HNRNPH1 and HNRNPR interact with pri-miRNA. Knocking down of HNRNPR increases the level of several mature miRNAs, and silencing of HNRNPH1 has various effects on tested mature miRNAs (Volk and Shomron 2011). ILF3 is another interacting protein of DGCR8, and heterodimer ILF3/ILF2 (also known as NF90/NF45) competes against DROSHA/DGCR8 by binding to the basal region of miRNA precursors in sequence independent fashion. BRCA1 is known to form a BRCA1-associated genome surveillance complex in order to repair DNA damage and prevent cell with severe DNA damage from proliferation. DHX9, a putative RNA helicase, carries DEAH (Asp-Glu-Ala-His) box and is able to alter RNA secondary structure. It has been reported to interact with DDX5, DDX17 and BRCA1. In HeLa cell,

BRCA1 and DHX9 were found to interact with DROSHA, SMAD3 and modulate miRNA maturation. RNA immunoprecipitation showed the precursors of hsa-let-7a-1, hsa-mir-16-1, hsa-mir-145 and hsa-mir-34a associate with BRCA1 and DHX9, and silencing BRCA1 or DHX9 decreases the mature level of associated miRNAs. FUS is a RNA-binding protein and mostly located in nucleus. It was identified in DROSHA complex in 2004 and was shown to participate in miRNA biogenesis recently. FUS is recruited to the chromatin where the miRNA is transcribed and binds the corresponding pri-miRNA. FUS depletion also decreases DROSHA level at the same chromatin loci, indicating FUS also helps to bring DROSHA to the pri-miRNA.

Besides the proteins introduced above, there are still many proteins physically associate with DROSHA, such as DDX3X, HNRNPU, TDP43 and NCL. Some of them carry similar protein domain with known biogenesis regulators; therefore, their roles in miRNA biogenesis still need to be clearly investigated (Gregory, Yan et al. 2004; Shiohama, Sasaki et al. 2007).

MiRNA biogenesis can be regulated independently. Some miRNAs are located in a cluster and being expressed together from the same nascent transcript. Cluster miR-17-92 contains six miRNAs, and they all associate with several different cancers by targeting important tumor-suppressor genes. Interestingly, the miRNAs in this cluster have different expression profiles. The detailed biogenesis mechanism of this cluster is still being studied, but HNRNPA1, an alternative-splicing factor, has been identified to target the terminal loop region of pri-mir-18a specifically from the miR-17-92 miRNA cluster and promote its maturation by facilitating DROSHA-mediated processing (Michlewski, Guil et al. 2008).

The best-studied negative biogenesis regulator is LIN28, which inhibits the maturation of let-7 family by targeting the terminal loop of let-7 precursors and interfere DROSHA processing (Figure 1b) (Heo, Joo et al. 2008; Newman, Thomson et al. 2008; Piskounova, Viswanathan et al. 2008). Additionally, LIN28 works in the cytoplasm and prevents pre-let-7 from DICER cleavage. In human, there are LIN28A and LIN28B two homolog genes, and both of them are able to inhibit the maturation of let-7 miRNAs family; however, they use different mechanism to modulate maturation of let-7 (Piskounova, Polytarchou et al. 2011). LIN28A induces post-transcriptional modification to let-7 precursors and block Dicer processing by adding additional nucleotides in the cytoplasm. On the other hand, LIN28B functions in the nucleus by targeting the terminal loop of pre-let-7 family and inhibit DROSHA processing. Interestingly, LIN28 is one of the targets of mature let-7 family, and this makes the regulation between let-7 family and LIN28 a double-negative feedback loop. Another special negative regulator is DROSHA itself, which is essential to miRNA biogenesis. DROSHA recognizes the two stem loops formed within DGCR8 mRNA and degrade it. Thus, the ratio of DROSHA and DGCR8 is important for miRNA processing.

Maturation of certain miRNAs is regulated by RNA editing. ADAR is an adenosine deaminase, which converts adenosine to inosine (A-to-I). RNA editing is able to change the hairpin structure of miRNA precursors and decrease the binding affinity of miRNA precursors to microprocessor (Kawahara, Zinshteyn et al. 2007; Kawahara, Megraw et al. 2008; Heale, Keegan et al. 2009). The A-to-I RNA editing directs miRNAs to Tudor-SN and enhances miRNA degradation. In addition to controlling the miRNA

concentration, A-to-I RNA editing expands the miRNA-targeting repertoire if the seed sequence of mature miRNA was edited (Kawahara, Zinshteyn et al. 2007).

Deep-sequencing technology allows us to identify novel miRNAs and miRNAs with untemplated 3' end addition. LIN28 recruits a noncanonical poly-A polymerase, TUT4, and adds untemplated uridine at the 3' end of miRNA. With extended uridine at the 3' end, pre-miRNA is unable to bind TARBP2 and fail to be processed by DICER. This uridylyl group is able to recruit 3' to 5' exonuclease as well (Heo, Joo et al. 2008; Hagan, Piskounova et al. 2009; Heo, Joo et al. 2009; Jones, Quinton et al. 2009; Lehrbach, Armisen et al. 2009). Therefore, the modified and unprocessed pre-miRNAs are directed to degradation. The other 3' end addition is adenylation. The deep-sequencing data showed many mature miRNAs have additional adenine at 3' end. The cytoplasmic poly-A polymerase GLD2 was found to monoadenylates and stabilizes miR-122 in mammals (Katoh, Sakaguchi et al. 2009). This suggests 3' end adenylation is a mechanism to control the concentration of miRNAs. Like protein-coding genes, the concentration of miRNAs is determined by the rate of expression, processing and degradation. In *C. elegans*, miRNA degradation is mediated by 5' to 3' exonuclease XRN-2, a quality control protein that removes incorrectly modified tRNA in yeast. Mature miRNAs bound to miRISC, and both 5' and 3' end of miRNA are attached and occupied by AGO2, which makes miRNAs resistant to degradation by exonuclease (Chatterjee and Grosshans 2009). Interestingly, miRNAs bind to target mRNA are resistant to XRN-2 as well due to occupied 3' and 5' end. The observation suggests target mRNA and miRNA modulate the degradation rate of each other.

MiRNAs regulate gene expression post-transcriptionally by mRNA degradation or translational repression. Translational repression does not affect the mRNA level but inhibit translation by deadenylate mRNA and interfere mRNA circulation. In conventional translation process, PABP1 binds poly-A tail and interacts with the cap-initiation complex through eIF4G to make mRNA a circle. When the miRISC targets 3' UTR, PABP1 dissociates from poly-A tail; meanwhile, the cytoplasmic deadenylase complex, including CAF1, CCR5 and CNOT1, starts to degrade poly-A tail. This process is for impeding the mRNA circulation; therefore, the translation initiation complex will not be recruited to start codon. On the other hand, the mRNA degradation starts with the dissociation as well, then the DCP1-DCP2 decapping complex removes 5' cap of mRNA. Finally, the 5'-to-3' exonuclease, XRN1, degrades uncapped target mRNA. Currently, it is still unclear that which method is mainly adopted in miRNA regulation. MiRNA-mediated translational repression was first observed in *C. elegans*. However, more and more evidences suggested mRNA degradation is predominantly used in animal, while translational repression is mostly adopted in plants. When measure the protein and mRNA change simultaneously upon perturbing single miRNA, more than 80% of affected proteins have significant change in miRNA level. This implies that in animal, the negative correlation between miRNA and its target should be observed.

Hundreds of miRNAs are identified in human, and each miRNA was estimated to regulate more than a hundred target mRNA level. Therefore, the entire miRNA-mRNA network comprises millions of miRNA-target interactions, and the complexity brings a lot of interest in developing bioinformatics approach to predict miRNA targets. Many tools have been developed to predict miRNA-target interactions, and these tools use the

knowledge in miRNA biology to design the algorithms. The knowledge include: a) The perfect complimentary of miRNA seed region to mRNA 3' UTR b) The conservation of miRNA target site c) The secondary structure of mRNA d) The expression profile of mRNA and miRNA e) Number of miRNA target sites. Among all the tools, Targetscan is the most widely used miRNA target prediction tool. Besides the perfect complimentary of seed region and the conservation of predicted miRNA binding sites, Targetscan built context score for each predicted miRNA binding site. The context scores are determined by the location and the type of binding sites. There are four types of miRNA target sites: 8mer means the miRNA 1^{st} to 8^{th} nucleotides are perfectly match (Watson-Crick pairing) to the binding sites, or the 1^{st} nucleotide mismatch to an Adenosine on 3' UTR. 7mer-m8 means the 2^{nd} to 8^{th} nucleotides match to the binding site. 7mer-A1 means the 2^{nd} to 7^{th} nucleotides match to the binding site with 1^{st} nucleotide mismatching to Adenosine. These three types of sites are considered as "canonical" target sites. Six-mer means only the 2^{nd} to 7^{th} nucleotides matched to 3' UTR, and this type of sites are considered as "noncanonical". The site efficacy is determined by quantification from microarray data and the hierarchy is as follows: 8mer >> 7mer-m8 > 7mer-A1 >> 6mer > no site, with the 6mer differing only slightly from no site at all. The location of miRNA target sites also affects the miRNA efficiency: the target sites closer to both end of 3' UTR generally have higher efficiency than the sites in the middle. These two features comprise of the context scores of miRNA target sites in Targetscan. The other more common miRNA target prediction tools include PITA, miRanda, PicTar, RNA22. All of them follow similar guidelines to infer miRNA targets with different approaches. For example, PITA also considered the accessibility of miRNA target sites from the secondary structure of

3'UTR. However, all prediction tools still have room to improve their performance. The poor performance may not due to the design of algorithm but because our current understanding to miRNA biology is not thorough enough. Additional proteins may be involved in the miRNA-mediated regulation and need to be considered. Different contexts also have distinct miRNA targeting profiles, thus building context-specific miRNA target prediction is necessary to have a full understanding of miRNA regulation.

Several proteins are known to modulate miRNA-targeting mechanism described above. For example, FMRP and PUF are positive modulators that bind miRISC and stabilize the interaction between miRISC and miRNA (Nolde, Saka et al. 2007; Galgano, Forrer et al. 2008; Edbauer, Neilson et al. 2010). Negative modulators, such as DND1 and ELAVL1, interfere miRNA targeting by displacing miRISC from the miRNA binding site or competing against miRISC for the same binding site on the 3' UTR (Bhattacharyya, Habermacher et al. 2006; Mishima, Giraldez et al. 2006; Kedde, Strasser et al. 2007). Along with the invention of high-throughput technology, we can measure the RNA-protein interaction happened in the whole biological system. RIP-chip uses microarray to measure the endogenous mRNAs bound by protein and get the binding motif from the target mRNAs.

Currently, several techniques have been used to systematically identify endogenous miRNA-target interactions. Microarray and SILAC (stable isotope labeling by amino acids in cell culture) profile the entire transcriptome and proteome (Baek, Villen et al. 2008; Selbach, Schwanhäusser et al. 2008). Both methods were applied to detect miRNA targets after perturbing the expression of single miRNA. Complementary sequence of miRNA seed region is enriched in the 3'UTR of affected mRNA and protein,

suggesting the affected mRNAs and proteins are targeted by miRNA. The concern of using microarray and SILAC to validate miRNA targets is the differentially expressed mRNAs or proteins may not be directly targeted by tested miRNA, since the miRNA seed region is only seven-bases long and easy to find complementary sequence on 3'UTR.

HITS-CLIP (High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) and PAR-CLIP (Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation), both technologies use ultraviolet cross-linking to fix the interactions between RNA and protein and obtain the actual binding sites using next generation sequencing. In the first PAR-CLIP work, the binding sites of PUM1, QKI, AGO2, IGF2BP1, IGF2BP2, IGF2BP3, TNRC6A, TNRC6B, TNRC6C and ELAVL1 have been detected by this high-throughput method in HEK293 cell (Chi, Zang et al. 2009; Hafner, Landthaler et al. 2010; Lebedeva, Jens et al. 2011; Mukherjee, Corcoran et al. 2011). Many of the binding sites were found to overlap with potential miRNA binding sites, which indicates they also modulate miRNA activity and interfere the miRNA regulation by competing the same binding sites. PAR-CLIP of AGO2 and RNA revealed

In the HITS-CLIP work, Chi *et al.* performed cross-linking between AGO and RNA in mouse brain in order to detect endogenous AGO-RNA interactions. Two different molecular weights (~110 kD and ~130 kD) of radiolabeled protein-RNA complexes were detected using autoradiogram, suggesting AGO is crosslinked to two different types of RNA. RT-PCR revealed the 110 kD complex harbored short RNA around 22 bases and the 130 kD complex contained both longer and 22-base short RNA. The associated RNAs were then identified using next generation sequencing. In summary, over 60% AGO binding sites locate in 3'UTR and the most enriched miRNA and miRNA

binding site are mouse-specific miRNA, miR-124. The identified binding sites of miR-124 were further validated for evolutionary conservation and luciferase reporter assay. This high-throughput method was shown to successfully complement the bioinformatics approaches to identifying miRNA target sites only from AGO-crosslinked region (Chi, Zang et al. 2009).

Crosslinking, Ligation And Sequencing of Hybrids (CLASH) is the latest developed method to identify miRNA-target interaction (Helwak, Kudla et al. 2013). The miRNA-target duplexes associated with human AGO1 were cross-linked by ultraviolet and pulled down. After RNA fragmentation, ligation of miRNA and miRNA binding site, cDNA synthesis and sequencing of AGO1-associated RNAs, the miRNAs and their targets can be detected in one single chimeric read. This method found many overrepresented motifs on chimeric reads that are complementary to miRNA seed region as expected. However, some overrepresented motif associated with miRNAs is complementary to the middle or 3' end of miRNA instead of seed region. Moreover, the association between the results of HITS-CLIP, PAR-CLIP, CLASH and the predictions from TargetScan and Miranda is not significant. Only around 37% of CLASH results showed uninterrupted Watson-Crick pairing in seed region. CLASH also revealed many detected miRNA binding sites that are located in coding region. Over 70% of miR-149 binding sites fall in coding region while only 19% in 3'UTR. HITS-CLIP and PAR-CLIP also showed 37% and 50% of AGO binding sites are in coding region. All bioinformatics efforts to identify miRNA targets were built on the concepts of finding complementary sequence of seed region in 3'UTR, but the concept may not fully reflect the real mechanism of miRNA-mediated regulation. Such events may be due to non-specific

binding of AGO2 or noises generated from these new technologies, but it also suggests we still need to put many efforts to improve our understanding of miRNA targeting mechanism besides seed sequence recognition (Chi, Zang et al. 2009; Helwak, Kudla et al. 2013).

MiRNAs are known to participate in various biological pathways and diseases. Certain miRNAs are called oncomiRs or tumor suppressor miRNAs by targeting important tumor suppressor genes or oncogenes. Dysregulation of these miRNAs, consequently, were identified in most solid tumors and hematological malignancies. In GBM, miRNAs that participate in the tumorigenesis include miR-21, miR-26a, miR-221, miR-222 and miR-7. MiR-21 is one of the most highly expressed miRNA in GBM patient. It targets DAXX, TP53BP2, and HNRNPK to down-regulate TP53 (Papagiannakopoulos, Shapiro et al. 2008). MiR-21 also targets and suppresses tumor suppressor PDCD4 (Chen, Liu et al. 2008). This protein inhibits the initiation of translation and promotes cell cycle arrest via P21. PTEN is a crucial tumor suppressor gene and frequently found deleted in GBM patients. In addition, miR-26a has been identified to target PTEN in GBM context (Huse, Brennan et al. 2009). It is the only focally amplified miRNA at DNA level in TCGA GBM patients and mostly in association with monoallelic PTEN deletion. MiR-221 and miR-222 are both up-regulated in GBM, and they share the same targets due to the identical seed regions. They regulate various important targets, such as p27, PUMA, p57 and PTPμ (Lu, Zhao et al. 2009; Zhang, Zhang et al. 2010; Zhang, Han et al. 2010; Quintavalle, Garofalo et al. 2012). In glioma, miR-221 and miR-222 have cooperative effect to activate AKT pathway and promote malignant progression. They also target PUMA to induce glioma

cell proliferation. Overexpression of miR-221/222 induces glioma cell proliferation and invasion in vitro and glioma growth in subcutaneous mouse model. MiR-7 is a potential tumor suppressor miRNA, but it is often found down-regulated in GBM. EGFR/AKT pathway is important in developing primary GBM. The pathway downstream effects are inhibition of apoptosis, proliferation and growth. It is found that 60% of GBM patients carry EGFR amplification. MiR-7 targets EGFR and inhibits the downstream AKT pathway (Kefas, Godlewski et al. 2008). The expression level of pri-mir-7 is equal between normal brain and GBM tissue; therefore the down-regulation of miR-7 in GBM is the consequence of impaired maturation of miR-7.

Chapter III

Identification of miRNA promoters

3.1 Introduction

MiRNAs are key post-transcriptional regulators but their promoters and transcription start sites are poorly characterized. It is difficult to identified miRNAs promoter is mainly because of the short half-life of miRNA precursors and different miRNA biogenesis mechanisms. Currently, miRBase records 1872 human miRNA genes, but only the sequences of hairpin structure are documented. In our analysis, our strategy is to identify miRNA post-transcriptional regulators using intragenic miRNAs and their host mRNAs. Additionally, we included the miRNAs located in the same transcriptional unit and regard one of the miRNAs as host mRNA. To use intragenic miRNAs, the first work is to test if the intragenic miRNAs or clustered miRNAs are transcribed from the same transcription unit.

Transcriptionally active promoters and transcription start sites possess some special features, including nucleosome depletion around the transcription start sites, tri-methylation on Histone 3 Lysine 4 residue (H3K4me3) and acetylation on Histone 3 Lysine 9, Lysine 14 and Lysine 27 (H3K9ac, H3K14ac and H3K27ac) around active promoters and transcription start sites. Several methods have been developed based on these features to identify promoters, including nucleosome mapping and ChIP (chromatin immunoprecipitation) assay of histone modifications (Kodzius, Kojima et al. 2006; Ernst, Kheradpour et al. 2011). CAGE (Cap Analysis Gene Expression) sequencing is a new method derived on the basis of mRNA processing (Kodzius, Kojima et al. 2006). This method captures the capped 5’ end of processed full-length mRNAs and identifies

transcription start sites on a genome-wide scale. To carry out CAGE, biotin bead was firstly conjugated to capped 5' end of mRNA and perform reverse transcription using random primers. Then use RNAase to degrade single strand RNA and pull down biotin-modified RNA-cDNA duplex using streptavidin. Finally, denature the RNA-cDNA duplex and sequence the cDNA. This method detects the precise location of 5' end of mRNA and provides the direct evidence of transcription start sites.

Ozsolak *et al.* firstly applied nucleosome mapping and and ChIP-chip of chromatin signature (H3K4me3, H3K9ac, H3K14ac, RNA polymerase II/III) in MALME, UACC62 and MCF7 cell lines to identify miRNA promoters (Ozsolak, Poling et al. 2008). They identified the promoter region of 175 miRNAs expressed in these three cell lines. Furthermore, they observed one third of intronic miRNAs have promoters independent from the their host genes.

This chapter summarizes the work of using the annotation of ChIP-seq from ENCODE (Encyclopedia of DNA Elements) and CAGE-seq from FANTOM (Functional Annotation of the Mammalian Genome) to identify intragenic miRNA promoters (Kawaji, Severin et al. 2011; Consortium, Bernstein et al. 2012). The two data sets provide comprehensive coverage of different contexts and conditions, including various tissues and cell lines with different treatments and time points.

3.2 Material and Methods

Chromatin modifications have been used to annotate genome sequences and detect cis-regulatory elements, given the important role of chromatin in DNA accessibility and recruitment of regulatory elements. ENCODE provides ChIP-seq data for eight histone modifications and one insulator protein (CTCF) from nice human cell

lines. The cell lines used in this study consist of embryonic stem cells (H1 ES), erythrocytic leukemia cells (K562), B-lymphoblastoid cells (GM12878), hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVEC), skeletal muscle myoblasts (HSMM), normal lung fibroblasts (NHLF), normal epidermal keratinocytes (NHEK) and mammary epithelial cells (HMEC). The ChIP-seq provided by ENCODE include: H3K4me3, H3K4me2, H3K4me1, H3K9ac, H3K27ac, H3K36me3, H4K20me1, H3K27me3 and CTCF. CTCF recognizes CCGCGNGGNGGCAG motif and its binding blocks the interaction between promoter and enhancers; therefore, CTCF binding was used to identify insulators.

ENCODE consortium then applied multivariate Hidden Markov model that uses combinatorial of chromatin modifications to infer chromatin states, and the algorithm is called “chromHMM”. Fourteen distinct transcriptional states were summarized from ENCODE ChIP-seq data. The transcriptional states include active/weak/inactive promoter, strong/weak enhancer, insulator, transcriptional transition/elongation, weak transcribed, polycomb repressed, heterochromatin and repetitive region.

The other data set used in this analysis is CAGE-seq from FANTOM. The FANTOM consortium conducted CAGE-seq for 1065 samples from different tissues and human cell lines under different treatment and time points. Previous study showed intragenic miRNAs that carry independent promoters are able to co-expressed with host mRNA in certain contexts. Therefore, this comprehensive data set provides us a broad coverage of the transcriptional landscape of intragenic miRNAs. Additionally, CAGE-seq provides direct evidence of transcription start sites compared to other methods. Note the

CAGE-seq is not public available yet at the time when this thesis is being prepared. It will be available when the pilot paper of FANTOM5 is published.

3.3 Results

Based on miRBase v18 and RefSeq hg19 annotation, 697 out of 1523 human miRNAs locate in 621 host genes. Due to every project adopts different miRNA expression platform, the numbers of host-miRNA pairs in each dataset varies. For example, there are 470 miRNA expression profiles and 226 host-miRNA pairs in GBM data but 723 miRNA expression profiles and 350 host-miRNA pairs in OV data. To determine miRNA clusters, the criteria for miRNAs to be in the same transcription unit are: 1) the miRNAs on the same strand of chromosome. 2) The miRNAs are within a 10kb. 3) The miRNAs have significant correlated expression profiles. 4) No other annotated transcript located in between. 5) No active promoter region was detected by chromHMM between the miRNAs. We identified 89 miRNA clusters, ranging from two miRNAs to 46 miRNAs.

To identify the independent promoter of intragenic miRNA, the interval region between transcription start sites of host mRNA and 5' end of miRNA precursor was examined. We obtained the chromHMM annotation for nine human cell lines and

hsa-mir-103-2	hsa-mir-511-1
hsa-mir-1183	hsa-mir-511-2
hsa-mir-1203	hsa-mir-532
hsa-mir-1255a	hsa-mir-548c
hsa-mir-1278	hsa-mir-548d-1
hsa-mir-1280	hsa-mir-548h-3
hsa-mir-1293	hsa-mir-548p
hsa-mir-149	hsa-mir-550-1
hsa-mir-153-1	hsa-mir-554
hsa-mir-153-2	hsa-mir-555
hsa-mir-15b	hsa-mir-558
hsa-mir-16-2	hsa-mir-561
hsa-mir-188	hsa-mir-575
hsa-mir-198	hsa-mir-576
hsa-mir-21	hsa-mir-581
hsa-mir-218-1	hsa-mir-590
hsa-mir-23b	hsa-mir-593
hsa-mir-24-1	hsa-mir-598
hsa-mir-27b	hsa-mir-615
hsa-mir-28	hsa-mir-620
hsa-mir-30c-1	hsa-mir-626
hsa-mir-30e	hsa-mir-634
hsa-mir-326	hsa-mir-642
hsa-mir-335	hsa-mir-660
hsa-mir-340	hsa-mir-664
hsa-mir-361	hsa-mir-9-1
hsa-mir-362	hsa-mir-935
hsa-mir-500	hsa-mir-938
hsa-mir-501	hsa-mir-944
hsa-mir-502	

Table 3.1 the list of intragenic miRNAs with independent promoter from chromHMM annotation.

examine the existence of active promoters in the interval region. On the other hand, the sequence reads of CAGE-seq data from every sample was normalized into RPM (reads per million reads). Similarly, we examined the existence of sequencing reads from CAGE in the interval region to identify the promoters of intragenic miRNAs.

From the chromHMM annotation, 59 intragenic miRNAs have potential to be transcribed independently in the nine human cell lines (Table 3.1). From FANTOM CAGE-seq, 42 intragenic miRNAs have CAGE-seq between transcriptional start sites (Table 3.2). Notably, none of the intragenic miRNAs with independent promoter have

miRNA	host	Max Read	Total Read
hsa-let-7f-2	HUWE1	10	826
hsa-mir-211	TRPM1	10	42
hsa-mir-579	ZFR	10	754
hsa-mir-598	XKR6	10	112
hsa-mir-623	UBAC2	10	59
hsa-mir-625	FUT8	10	348
hsa-mir-652	TMEM164	11	385
hsa-mir-106b	MCM7	12	104
hsa-mir-1250	AATK	12	56
hsa-mir-147b	C15orf48	12	148
hsa-mir-218-2	SLIT3	12	49
hsa-mir-1228	LRP1	15	960
hsa-mir-1248	EIF4A2	15	76
hsa-mir-1307	USMG5	15	37
hsa-mir-934	VGLL1	15	42
hsa-mir-186	ZRANB2	16	781
hsa-mir-224	GABRE	16	614
hsa-mir-335	MEST	17	220
hsa-mir-1291	C12orf41	18	494
hsa-mir-24-1	C9orf3	18	611
hsa-mir-338	AATK	19	123
hsa-mir-16-2	SMC4	20	1395
hsa-mir-7-1	HNRNPK	21	783
hsa-mir-101-2	RCL1	22	819
hsa-mir-548c	RASSF3	22	946
hsa-mir-22	C17orf91	28	1011
hsa-mir-708	ODZ4	28	161
hsa-mir-1238	ATG4D	29	151
hsa-mir-126	EGFL7	36	607
hsa-mir-128-2	ARPP21	37	254
hsa-mir-647	UCKL1	38	945
hsa-mir-1914	UCKL1	43	226
hsa-mir-9-1	C1orf61	43	750
hsa-mir-1234	CPSF1	53	1065
hsa-mir-330	EML2	60	668
hsa-mir-204	TRPM3	82	238
hsa-mir-15b	SMC4	88	5564
hsa-mir-1469	NR2F2	94	2900
hsa-mir-1908	FADS1	102	6930
hsa-mir-1178	CIT	110	367
hsa-mir-636	SRSF2	157	12219
hsa-mir-198	FSTL1	243	9510

Table 3.2 Intragenic miRNAs with CAGE-seq reads in FANTOM. The listed reads are normalized RPM (reads per million). The "Max Reads" represents the number of max RPM from single sample, and the "Total Reads" represents the sum of RPM from all 1065 samples in FANTOM.

CAGE-seq reads in all samples. This circumstance may be due to the intragenic miRNAs are not ubiquitously expressed in all contexts. Another explanation is that they can adopt the promoter of host genes in certain conditions.

An example of intragenic miRNA possessing independent promoter: miR-21 is the most well studied oncomiR, and it is only around 700 bases away from its

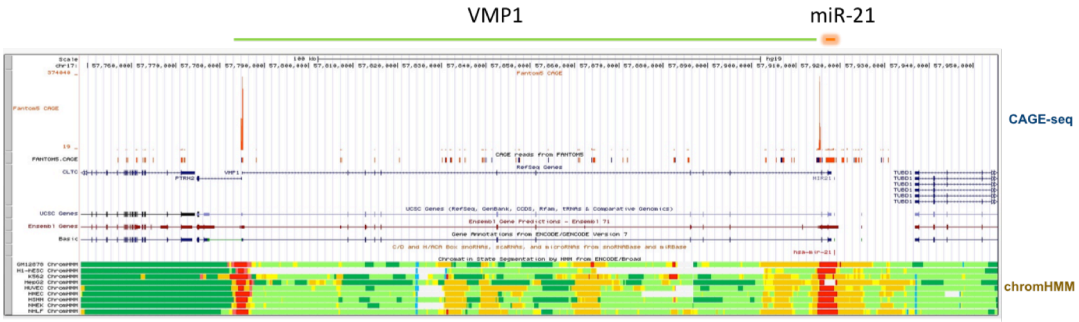


Figure 3.1 miR-21 possesses its own promoter. CAGE-seq and chromHMM showed VMP1 and miR-21 are transcribed independently

neighboring gene, VMP1. In some annotation, miR-21 was annotated in the same transcription unit as VMP1. Here, CAGE-seq data and chromHMM both indicates miR-21 possesses its promoter (Figure 3.1). Another example is miR-9. Mir-9-1 resides in the second intron of C1orf61. The CAGE-seq and ChIP-seq of H3K4me3 clearly showed miR-9-1 possesses its own promoter, and it can be expressed independently. The data also showed even if the miRNA possesses its own promoter, it can still be co-transcribed with its host gene in certain contexts (Figure 3.2).

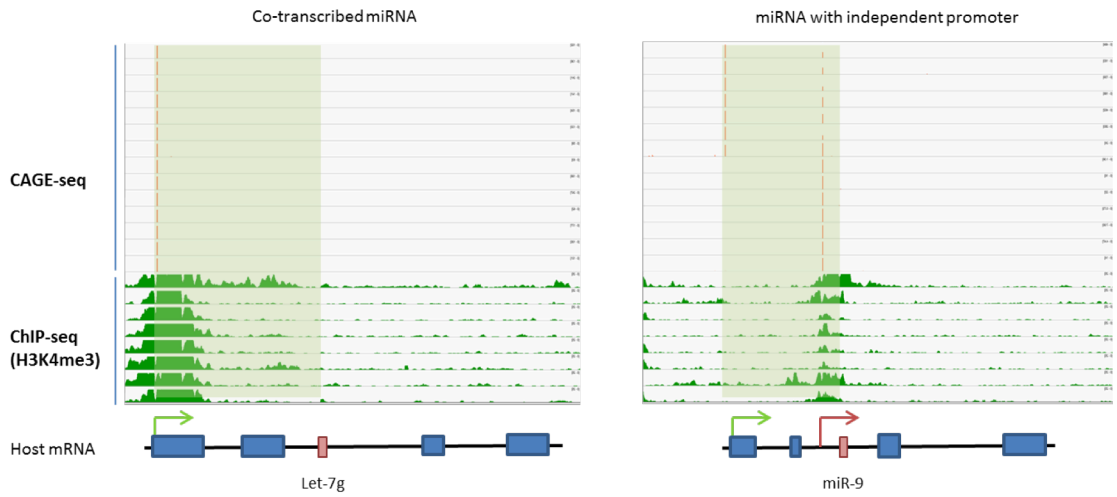


Figure 3.2 let-7g is co-transcribed with host and miR-9 expresses independently. CAGE-seq and ChIP-seq of H3K4me3 showed let-7g is co-transcribed with its host gene, while miR-9-1 possesses an independent promoter. Notably, CAGE-seq showed miR-9-1 is co-transcribed with its host, C1orf61 in certain conditions.

3.4 Discussion

In this chapter, we summarized the results from chromHMM and CAGE-seq to identify the intragenic miRNAs with independent promoters. After selecting miRNAs that are co-transcribed with host genes, we examine the relationship between the host mRNA and the intragenic miRNA in glioblastoma multiforme (GBM), ovarian serous cystadenocarcinoma (OV), acute myeloid leukemia (AML), breast invasive carcinoma (BRCA) from TCGA, and prostate adenocarcinoma (PRAD) from GSE21032 using spearman correlation (Figure 3.3). Although these intragenic miRNAs co-transcribe with their host genes, the relationships between some host-miRNA pairs are still not significant or even anti-correlated, indicating miRNA biogenesis pathway is indeed extensively controlled by post-transcriptional regulators.

Oh the other hand, the distribution of spearman correlation between miRNAs in the same cluster also showed that clustered miRNAs might undergo separate regulation in the maturation process (Figure 3.4). Indeed, we observed the same scenario from small RNA sequencing for RNA between 50 and 150 bases and shorter than 30 bases in SNB19 cells. MiR-17, miR-18a and miR-19a are in the same transcription unit, miR-17-92 cluster, and they are 376 bases apart. However, only miR-17 and miR-19a were successfully being processed into mature miRNAs. The maturation of miR-18a was stalled by post-transcriptional regulation; therefore, pri-miR-18a was accumulated in SNB19 cell (Figure 3.5).

In summary, we have collected around 250 host and miRNA pairs by using chromHMM in ENCODE and CAGE-seq in FANTOM. The next step is to take advantage of these host-miRNA pairs and perform genome-wide scanning for regulators that affect miRNA level post-transcriptionally.

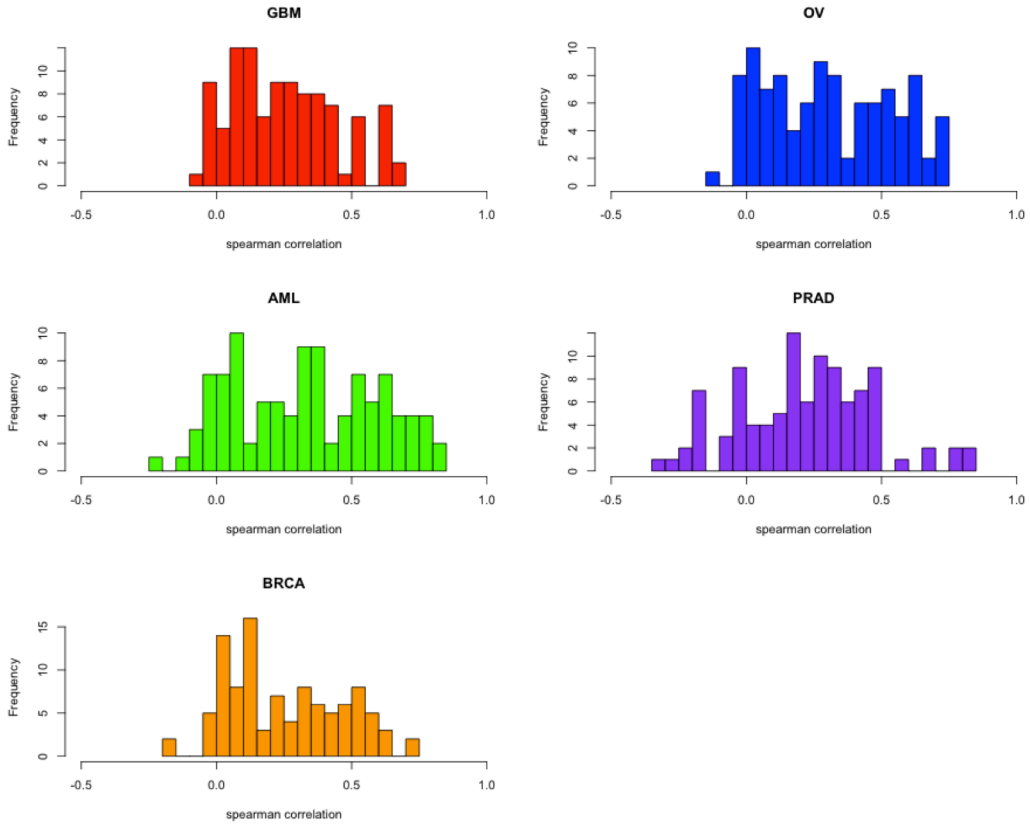


Figure 3.3 Histograms of spearman correlation between host mRNAs and intragenic miRNAs from five different tumor types.

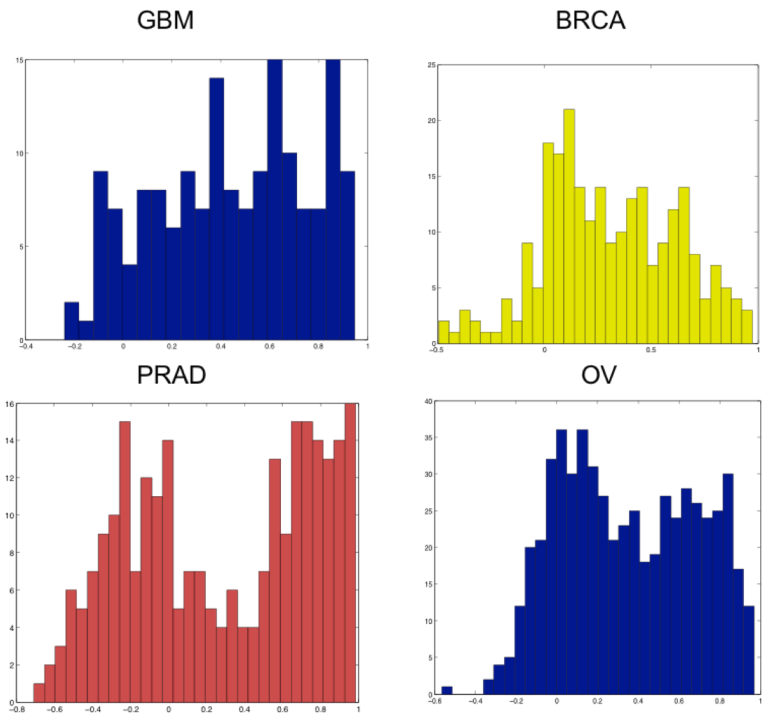


Figure 3.4 Histograms of spearman correlation between two miRNAs in the same cluster from four different tumor types.

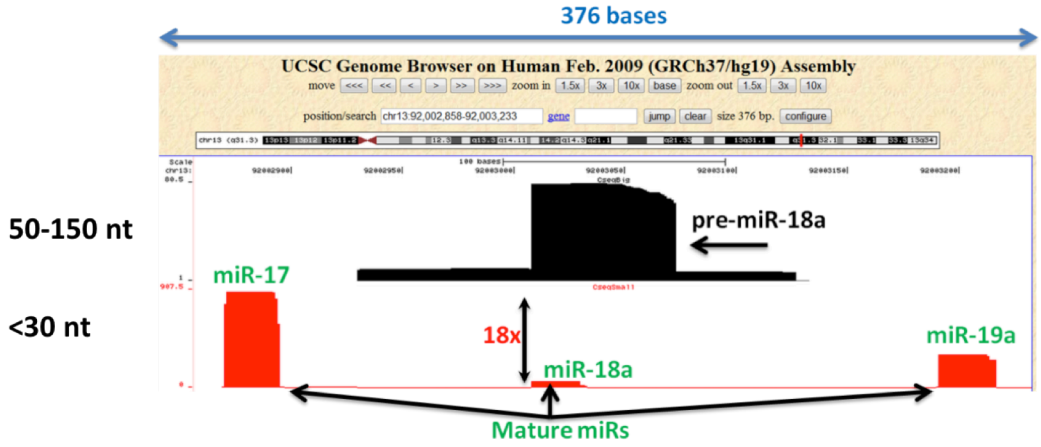


Figure 3.5 Small RNA deep-sequencing showed miR-18a level was regulated by distinct post-transcriptional mechanism in SNB19 cell.

Chapter IV

Identification of miRNA biogenesis regulators

4.1 Introduction

Tumor-specific aberrant expression of microRNAs (miRNAs) contributes to many cancer types, and the dysregulation of their post-transcriptional biogenesis regulators have been repeatedly seen in cancer (van Kouwenhove, Kedde et al. 2011) and is known to affect tumor growth. Particularly, the inactivation of DROSHA, DICER and other members of the microprocessing complexes, the canonical miRNA biogenesis regulators, is known to affect both development (Han, Lee et al. 2006; Suzuki, Yamagata et al. 2009; Melo, Moutinho et al. 2010; Franceschini, Szklarczyk et al. 2013) and carcinogenesis (Viswanathan, Daley et al. 2008), including gliomagenesis (Yamagata, Fujiyama et al. 2009; Ernst, Campos et al. 2010), breast cancer survival (Dews, Fox et al. 2010), and ovarian cancer progression and prognosis (Malzkorn, Wolter et al. 2010; Rao, Santosh et al. 2010; Lages, Guttin et al. 2011) through miRNA-dependent and miRNA-independent pathways (Wuchty, Arjona et al. 2011). For example, TARBP2 is a vital component of the DICER processing complex (Glesne and Huberman 2006), and Exportin 5 (XPO5) transports pre-miRNAs to the cytoplasm. Down regulation of either protein depresses the expression of mature miRNAs and promotes tumorigenesis (Friedman, Farh et al. 2009; Melo, Ropero et al. 2009; Melo, Moutinho et al. 2010). While perturbations to canonical miRNA biogenesis regulators affect pathophysiologically relevant processes in the cell, it does not account for all tumor specific expression of miRNAs (Lu, Getz et al. 2005).

Non-canonical regulators of miRNA biogenesis target mature miRNAs and/or their precursors with sequence and structure bias (Grimson, Farh et al. 2007; Yi, Pasolli et al. 2009; Volk and Shomron 2011). Among all identified non-canonical miRNA biogenesis regulators, LIN28 is one of the best-studied regulators (Zhang, Zhang et al. 2010; Piskounova, Polytarchou et al. 2011). It recognizes the motif on the hairpin loop of let-7 family miRNAs and represses their processing. Over-expression of the oncogene LIN28B is known to increase tumor proliferation by suppressing the biogenesis of the tumor suppressor let-7. Moreover, deleterious SNPs at its locus influence susceptibility to epithelial ovarian cancer (Zhang, Han et al. 2010), leukemia (Medina, Zaidi et al. 2008; Lu, Zhao et al. 2009), breast cancer (Quintavalle, Garofalo et al. 2012) and other malignancies (Thornton and Gregory 2012). Other well-known examples are the dead-box RNA helicases DDX5 and DDX17 (Fukuda, Yamagata et al. 2007; Paris, Ferraro et al. 2012). DDX5 and DDX17 form a complex with DROSHA and DGCR8, and interact with DHX9 (Kawai and Amano 2012) to promote miRNA maturation. The oncogenic receptor-regulated SMAD proteins, SMAD1, SMAD3 and SMAD5, have been shown to regulate miRNA-biogenesis regulation by controlling DROSHA-mediated miRNA maturation through interaction with DDX5 and miRNA precursors in response to TGFβ and BMP signaling (Davis, Hilyard et al. 2008). Recently, another study showed SMAD4, SMAD5 and SMAD9 can associate with phosphorylated KSRP and block its interaction with primary myogenic miRNAs (Pasero, Giovarelli et al. 2012). The interaction abrogates the KSRP-dependent miRNA maturation on let-7 and, therefore, affects cell differentiation (Viswanathan, Daley et al. 2008; Trabucchi, Briata et al. 2009; Pasero, Giovarelli et al. 2012). HNRNP complex modifies pre-mRNA processing, stability and

transport post-transcriptionally, and some subunits in HNRNP complex are directly involved in miRNA processing as well. HNRNPA1, HNRNPH1 and HNRNPR are able to interact with pri-miRNA and DGCR8 complex. HNRNAPA1 and HNRNPR have inhibitory effects on several mature miRNAs, while HNRNPH1 promotes miRNA processing. In total, over a dozen non-canonical miRNA biogenesis regulators have been identified (Grimson, Farh et al. 2007), and many of these have already been associated with pathogenesis and tumorigenesis. Their full repertoire, including their tissue and tumor specific activities remains unknown and the focus of extensive research for nearly

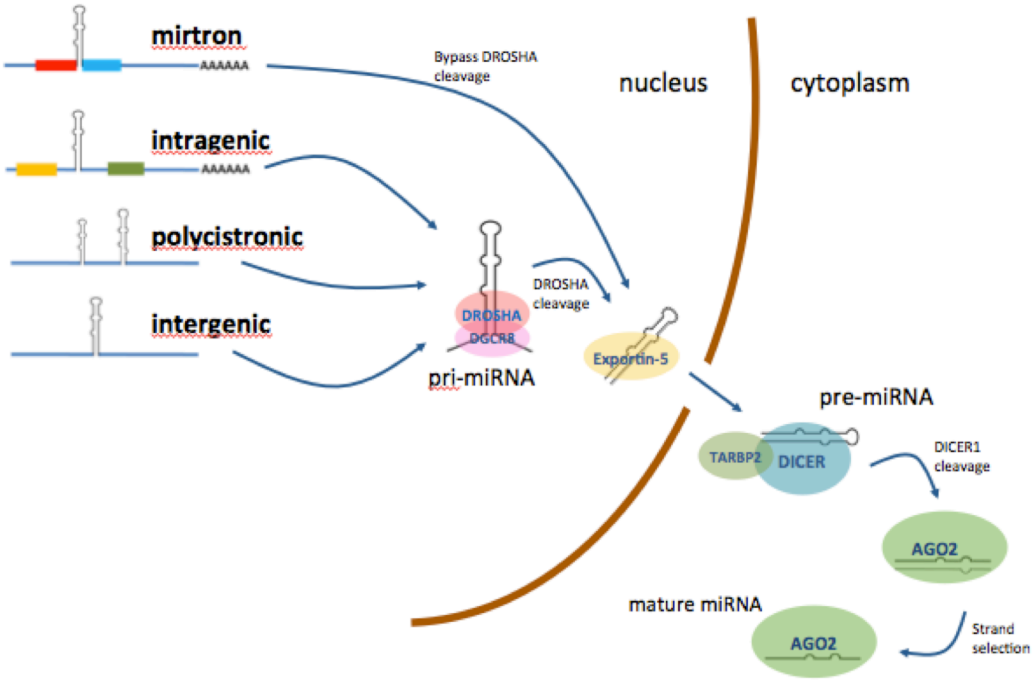


Figure 4.1 The canonical miRNA biogenesis pathway.

The miRNAs are transcribed and then being processed into pri-miRNA, pre-miRNA and miRNA-miRNA* duplex and mature miRNA. The core components in this pathway are DROSHA, XPO5, DICER1 and AGO2

a decade (Bartel 2004).

The regulatory role of miRNAs in gliomagenesis has been extensively studied, and numerous miRNA species have been associated with glioma and its subtypes (Singh, Soon et al. 2012; Moller, Rasmussen et al. 2013). These glioma-related miRNAs include the viability and proliferation regulators miR-7 (Kefas, Godlewski et al. 2008) and miR-34a (Li, Guessous et al. 2009), the oncogenic cluster miR-17-9 (Dews, Fox et al. 2010; Ernst, Campos et al. 2010) the PTEN inhibitor miR-26a (Huse, Brennan et al. 2009), and the neuromesenchymal-specific regulators miR-27b (Chen, Li et al. 2011) and miR-9 (Kim, Huang et al. 2011; Schraivogel, Weinmann et al. 2011). Recent systems biology approaches have identified numerous miRNAs that are differentially expressed in glioma subtypes and are predicted to regulate gene expression in glioma (Sumazin, Yang et al. 2011; Setty, Helmy et al. 2012) and its precursor-associated subclasses (Kim, Huang et al. 2011). In total, over 100 miRNA species are differentially ($p<1e-03$ by t-test) regulated across glioblastoma subclasses (Lambertz, Nittner et al. 2010; Kim, Huang et al. 2011). While the aberrant expression of some miRNAs, including miR-26a (Huse, Brennan et al. 2009), may be linked to the copy number variation at their loci, in 80% of the instances where miRNAs are aberrantly expressed their loci have no significant (aCGH ratio below -0.25 or above 0.25) copy number variations (Lambertz, Nittner et al. 2010). In these instances, aberrant expression of miRNAs is either due to transcriptional dysregulation or to the dysregulation of their biogenesis and decay. We set out to identify RNA binding proteins (RBPs) that post-transcriptionally influence the abundance of mature miRNAs in glioblastoma, thus accounting for some of their tumor specific expression. We report on a

genome-wide screen for such RBPs using TCGA expression of genes and miRNAs in glioblastoma tumors (Lambertz, Nittner et al. 2010).

Our results suggest that dozens of RBPs post-transcriptionally regulate miRNA biogenesis. Specifically, our computational screen identified thirty-three miRNA regulators in glioblastoma that target both intragenic and intragenic miRNAs, of which eight have been previously shown to regulate miRNA biogenesis. Selecting uniformly across the list we validated seven predictions, including regulators that are expressed in tumor specific manner. For these seven RBPs, we biochemically identified miRNA targets in a glioblastoma cell line using qRT-PCR and Wafergen technologies. DDX10 depletion in glioblastoma tumors downregulated the biogenesis of miRNAs, including miR-218, miR-18a and miR-25, while SMAD6 siRNA-mediated silencing lead to down regulation of mature miRNAs, including miR-196a and miR-25. Finally, we showed that over expression of IGF2BP3, whose locus is amplified in 40% of TCGA glioblastoma patients, up regulates the abundance of oncogenic miRNAs including miR-7, miR-9, miR-26a, miR-19a, miR-17 and miR-16 and is predictive of patient survival rates. MiRNAs that were downregulated as a result of siRNA-mediated silencing of IGF2BP3 are enriched for PTEN-regulators, and IGF2BP3 silencing lead to (1) upregulation of PTEN 3'-UTR activity, (2) DICER-dependent upregulation of PTEN mRNA expression, and (3) reduction in glioma cell line growth rate.

The Cancer Genome Atlas (TCGA) is initiated by National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) in 2006. TCGA build a national network of research and technology teams to generate and pool data from multiple cancer types. The multi-dimensional data of TCGA include gene, exon and

miRNA expression profile, copy number, DNA methylation, somatic mutation, single nucleotide polymorphism and clinical information about patients. Multiple data types enable scientists and clinicians to approach cancer genomics from various perspectives. Currently, the project collects data from 20 different types of cancer, including glioblastoma multiforme (GBM). We utilized the TCGA GBM dataset, which consists gene and miRNA expression profiles from 493 patients.

Glioblastoma is a highly aggressive and essentially incurable tumor. By integrating various types of high-throughput data, the TCGA dataset can be classified into 4 subtypes: proneural, neural, classical and mesenchymal (Verhaak, Hoadley et al. 2010). Expression and genetic aberrations of EGFR, NF1 and PDGFRA/IDH1 define the classical, mesenchymal, and proneural subtypes, respectively. The gene expression profile in proneural is significantly different from mesenchymal. Additionally, the responses to chemotherapy are different between subtypes: classical and mesenchymal patients are sensitive to intensive chemotherapy and radiation. The mesenchymal subtype is associated with a state of uncontrolled ability to invade and stimulate angiogenesis to metastasize. The patients in mesenchymal subtype have the shortest survival span among all subtypes. The mesenchymal gene expression signature (MGES) is the gene set that is up-regulated in the mesenchymal subtype. The mechanism that controls MGES is still unclear. Previously, Califano lab developed ARACNe to identify the relationship between transcription factor and target genes using mutual information (Carro, Lim et al. 2010). Carro *et al.* used ARACNe to identify the master regulators that control the MGES, including RUNX1, STAT3, CEBPB/D, FOSL2 and BHLHB2 as positive master modulators and ZNF238 as negative master modulator for MGES.

In 2010, Noushmehr *et al.* utilized DNA methylation profile in the promoter region to identify another distinct subgroup in the TCGA context (Noushmehr, Weisenberger et al. 2010). A large number of loci were identified to be hypermethylated in a subgroup of proneural patients, and the hypermethylation in the promoter regions leads to down-regulation of gene expression. This subtype is called glioma-CpG island methylator phenotype (G-CIMP). The characteristics of G-CIMP patients are younger at diagnosis and significantly improved survival span compared to the rest of proneural patients. Generally, G-CIMP subtype is associated with low-grade glioma and indicates good prognosis after intensive treatment. At molecular level, G-CIMP subtype is closely associated with IDH1 mutation or copy number alteration.

MiRNA expression profile has been utilized to classify different tumors and trace the cell origins the tissue successfully. In TCGA GBM context, people used (i) miRNAs with higher variability (median absolute deviation or MAD > 0.1) (ii) miRNAs that are patient survival-related (significant in univariate Cox model) (iii) miRNAs that are neurodevelopment-related obtained from literatures to identify subtypes in glioblastoma. MiRNA expression profile classified TCGA GBM patients into five subtypes: oligoneural, neural, radial glial, neuralmesenchymal and astrocytic. Compared to the classification obtained from gene expression profile, oligoneural approximately corresponds to proneural and astrocytic corresponds to mesenchymal. The major difference between miRNA- and gene-defined glioblastoma subtypes is that the survival spans of patients after diagnosed are significantly different. Patients in oligoneural subtype has significantly longer survival span than patients in astrocytic subtype, while there is no difference between patients in proneural and mesenchymal subtypes. The

responses to radiation (at least 54 Gy) and 2 or more cycles of temozolomide treatment are clearly different as well. Astrocytic subtype is the only subtype sensitive to intensive treatment, while it is the most aggressive subtype as well (Kim, Huang et al. 2011).

In this study, we applied information theory to infer the genes that are involved in miRNA processing network, including the post-transcriptional regulators and interacting proteins. Information theory is developed by Claude Shannon in 1940s' and has widely been applied to a variety of field, such as statistical inference, natural language processing, quantum computing and systems biology. Mutual information is one of the important concepts in information theory. Mutual information is a measurement of the mutual dependence between two random variables. It showed the amount of information we can obtain for one random variable given the other variable. The mutual information between two variable X and Y can expressed as:

$$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y)$$

Where H(X) represents the entropy of discrete random variable X. Shannon entropy is a measurement in information theory, measuring the amount of uncertainty associated with the variable X. The entropy of X can be written as:

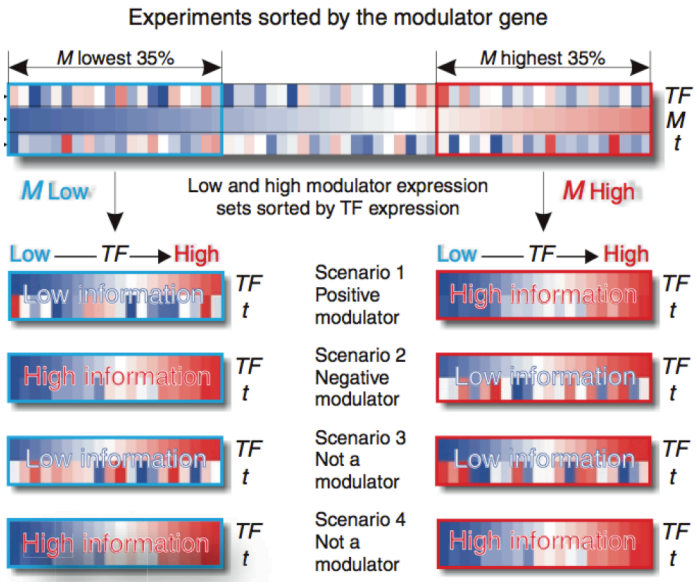
$$\begin{cases} H(X) = -\int P(X_i) log_2\big(P(X_i)\big) \;\; (continuous\ variable) \\ H(X) = -\sum_i P(X_i)\, log_2\big(P(X_i)\big) \; (discrete\ variable) \end{cases}$$

Conditional mutual information is another measurement in information theory to capture three-way relationship. It showed the amount of mutual information we can

obtain between two random variables given the third variable. The conditional mutual information of variable X and Y given Z can be written as:

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z)$$

Since 1953, information theory has been regarded as a tool to study biology (Quastler 1953). It was used to finding splicing sites in mRNA precursors, locating protein-binding sites and building phylogenetic trees. Information theory has been successfully applied in systems biology as well. ARACNe (Algorithm for the reconstruction of accurate cellular networks) predicts the targets of transcription factors and build transcriptional network using information theoretic approach (Margolin, Nemenman et al. 2006; Margolin, Wang et al. 2006). It identifies the transcription factors and their targets with significant mutual information, and it also implements another concept in information theory, Data Processing inequality (DPI) to remove inferred



(Wang 2009)

Fig 4.2 Concept of MINDy algorithm.

Delta mutual information was used to identify candidate modulator genes

relationships that may be mediated by intermediate TF. The first work of ARACNe successfully builds the transcriptional network in B cell context.

MINDy (Modulator inference by network dynamics) is designed to infer post-translational regulatory network. Signaling proteins, such as kinases, phosphatases, can modulate the activity of transcription factors by post-translational modification. The three-way relationship between TF, target and signaling protein can be inferred by MINDy using delta mutual information. The idea is to measure the difference of mutual information between TF and target when candidate modulator is differentially expressed (Wang, Saito et al. 2009). In the first MINDy work, several novel modulators of MYC

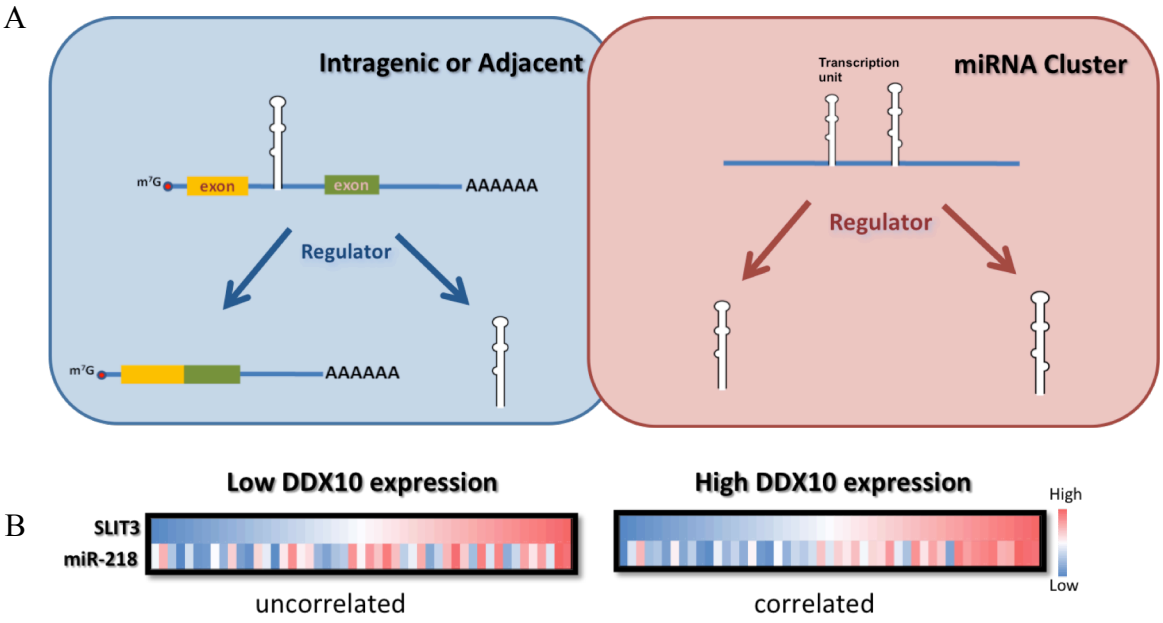


Figure 4.3 Concept of MIRAGE.

(A) Candidate miRNA biogenesis regulators are predicted using intragenic and clustered miRNAs.

(B) SLIT3 and miR-218 are correlated when DDX10 is highly expressed; suggesting DDX10 is involved in the maturation of miR-218

activity were inferred, such as STK38, and successfully validated in human B lymphocytes.

In this analysis, we developed an algorithm, MIRAGE, to identify miRNA biogenesis regulators. To identify miRNA biogenesis regulators, we first summarized intragenic miRNAs and host genes, or miRNA clusters with genomic coordinates, expression profiles and additional evidences (described in previous chapter). Secondly, MIRAGE predicted candidate biogenesis regulators by finding genes whose expression is correlated with deviation in co-expression between miRNA and host using conditional mutual information. The candidate genes from all host-miRNA pairs were integrated by Fisher's method (Fisher 1925). Finally, we validate several candidate biogenesis regulators using siRNA silencing or over-expression. The miRNA expression level was then measured by real-time PCR or array-based technology, such qRT-PCR, WaferGen and Fluidigm.

4.2 Material and Methods

Estimation of mutual information and conditional mutual information

Mutual information in this study was estimated by Gaussian kernel estimator for computational efficiency. The pairwise mutual information is estimated as:

$$I(X;Y) = \int\int f(x,y)\log\left(\frac{f(x,y)}{f(x)f(y)}\right) dx\, dy$$

And the 1-dimensional and 2-dimensional Gaussian kernel density estimator, f(x) and f(x,y) are defined as:

$$f(x) = \frac{1}{N}\frac{1}{h\sqrt{2\pi}}\sum_{i=1}^{N}\exp\left(-\frac{(x-x_i)^2}{2h^2}\right)$$

$$f(x,y) = \frac{1}{Nh^2}\frac{1}{2\pi}\sum_{i=1}^{N}\exp\left(-\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{2h^2}\right)$$

The parameter N in the equations above is the number of sample used in the estimation, and the parameter h is smoothing parameter or bandwidth.

On the other hand, conditional mutual information can be defined as:

$$I(X;Y|Z) = \iint p(x,y,z)\log\left(\frac{p(x,y,z)}{p(x,z)p(y,z)}\right)dx\,dy$$

The statistical significance of mutual information and conditional mutual information were determined by null distribution built using shuffling the expression profile of candidate genes. The expression profile of host mRNA and intragenic miRNA were hold constantly. The cutoff of conditional mutual information is less than 0.01 after Bonferroni correction.

Data source

Gene and miRNA expression was obtained from TCGA glioblastoma multiforme data portal. Gene and miRNA expression from glioblastoma patients were profiled on Agilent 244K whole genome expression array and miRNA_8x15K human specific microarray, respectively. Gene expression profile was Lowess normalized and the ratio of Cy5 (sample) to Cy3 (reference) was log2 transformed to create gene expression values for 18624 genes. MiRNA expression profile was quantile normalized on the probe level.

Signals form the probes designed to measure the same miRNA were summed together and log2 transformed in order to obtain miRNA-centered value (Lambertz, Nittner et al. 2010).

List of RNA binding proteins were obtained from Gene ontology. The protein-protein interactions were acquired from StringDB and PrePPI (Franceschini, Szklarczyk et al. 2013; Zhang, Petrey et al. 2013).

Screening for miRNA regulator

A post-transcriptional miRNA regulator (R) is able to explain the disagreement between host mRNA and the miRNA in the same transcription unit. Similar with MINDY and Hermes, we use conditional mutual information (CMI), I(miRNA;Host|R), to identify post-transcriptional miRNA regulators. The statistical significance (p-value) of CMI is evaluated by permutation testing. All selected triplets have significant CMI (p-value < 1e-5) and I(miRNA;Host|R) > I(miRNA;Host). The mutual information and CMI are estimated using adaptive partitioning algorithm. To evaluate the significance of each predicted miRNA regulators, the p-value of triplets from the same miRNA regulators are integrated using Fisher's method, where the $-2\sum_{k=1}^{N}\ln(p_k)$ follows the chi-square distribution and the degree of freedom equals to the number of integrated p-values.

Cell and Culture Condition

The glioma-derived cell lines SNB19 were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS (GIBCO/BRL). Freshly trypsinized cells were suspended at $3x10^5$ cells/ml in standard culture medium and seeded at a density of

$3x10^5$ cells per well in standard six-well tissue culture plates. After seeding, the cells were incubated at 37°C in a 95% air/5% CO_2 humidified atmosphere, and one milliliter of fresh medium was supplied every other day to the cultures after removal of the supernatant.

Transformation with Heat Shock

Take competent *E. coli*, DH5α, cells from –80°C freezer. Turn on water bath to 42°C. Put competent cells in a 1.5 ml eppendorf. (For transforming a DNA construct, use 50 μl of competent cells. For transforming a ligation, use 100 μl of competent cells.) Keep tubes on ice; then, add 50 ng of circular DNA into *E. coli* cells. Incubate the mixture on ice for 10 minutes. Next is to thaw competent cells: put the tube with DNA and *E. coli* into water bath at 42°C for 45 seconds. Then, put the tube back on ice for 2 minutes to reduce damage to *E. coli*. Add 1 ml of LB media without antibiotics and incubate the tube at 37°C for 1 hour. Finally, spread about 100 μl of the resulting culture on LB plates with appropriate antibiotic added (usually Carbenicillin or Kanamycin). Grow overnight at 37°C and then pick colonies after 12-16 hours for further experiments.

RNA Extraction and Quantitative Reverse Transcription Polymerize Chain Reaction

To perform cell2cDNA and reverse transcription, firstly, remove the medium from plate, and wash the plate with cold PBS. Aspirate the PBS and then add trypsin. After cells start to detach from plate, inactivate the trypsin with serum, and spin at 4°C and remove the supernatant to collect cells. Resuspend the pelleted cells in ice cold PBS, spin again at 4°C < 1200g for 5 minutes.

Add 100 μl of ice-cold cDNA lysis buffer. Mix the cell and lysis buffer by vortex and pipetting, and then transfer the mixture into PCR machine immediately and incubate at 75°C for 10 minutes. Cool the sample on ice, and add 2 μl DNAase per 100 μl lysis buffer. Mix the sample thoroughly and vortex gently. Put the sample back to PCR machine, and incubate the sample at 37°C for 15 minutes, then 75°C for 5 minutes to inactivate the DNAase. The RNA sample can be stored at -20°C for one week or at -80°C for 2 months.

To reverse transcription (RT), first of all, mix 5 μl of RNA sample with 2 μl of random primers, 8 μl of nuclease free water and vortex the mixture for 10 seconds. Incubate the mixture at 65°C for 5 minutes and immediately transfer the sample on ice. Add 4 μl of qScript Flex Reaction Mix (5X) and 1 μl of qScript Reverse transcriptase to the sample. Mix the sample by vortex for 10 seconds and incubate the sample 10 minutes at 25°C, followed by 45 minutes at 42°C and 5 minutes at 85°C. Finally, hold the temperature at 4°C. After completing reaction, cDNA sample can be stored at -20°C.

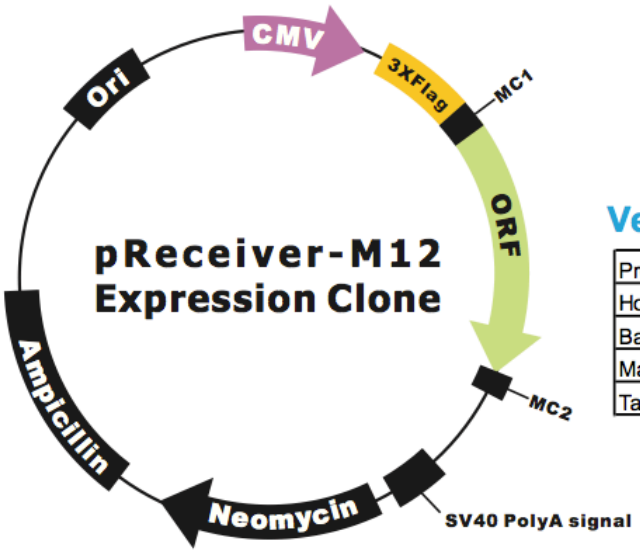
To perform quantitative RT-PCR for gene, prepare the PCR mixture and samples and aliquot into a 96-well plate for ease with multichannel pipet. Every reaction contains 12.5 μl ABI 2X sybergold buffer with 2 μl primer (3.75mM), 8.5 μl of water and 2 μl of sample cDNA. Seal the plate and corners well with plastic film. Spin the plate at 1500 rpm for 5 minutes. The qRT-PCR was performed using ABI 7300. The thermal cycles of reaction include two steps: step 1 (1 cycle) starts with 2 minutes at 50°C followed by 10 minutes at 95°C. Step 2 starts with 15 seconds at 95°C, followed by 60 seconds at 60°C, repeating for 40 cycles. The Ct value of the gene was firstly normalized with internal

control, such as GAPDH or RNU44. The value is called delta Ct (dCt). Subsequently, normalize the dCt of the gene with the dCt of the same gene from control experiment to obtain delta delta Ct (ddCt), which represents the fold change of gene expression.

Procedure of qRT-PCR for pri-miRNA is the same as qRT-PCR for gene. Every reaction of qRT-PCR for pri-miRNA contains 10 μl TaqMan 2X Gene Expression Master Mix with 1 μl 20X TaqMan Pri-miRNA Assays, 5 μl of water and 4 μl of sample cDNA.

SMAD6 primers: forward 5'-ACGGTGACCTGCTGTCTCTT-3', reverse 5'-ACGTGACGGTTTTGAGTTCC-3'. DICER1 primers: forward 5'-AGGATGAGGAGGAGGAGAGC-3', reverse 5'-TTTGGGCATTTTCCATTCAT-3'. DROSHA primers: forward 5'-CATGCGGAAGAAAGGGATTA-3', reverse 5'-AATGGCAGTCCGATAGGTTG-3'. SLIT3 primers: forward 5'-CCTGCCCCTACAGCTACAAG-3', reverse 5'-TTGTTTTCGCAGTCGTTGTC-3'. DDX10 primers: forward 5'-AAATCTGCCATCAAGGATGC-3', reverse 5'-

TCTTTGGCCTTTGCTTGTCT-3'.



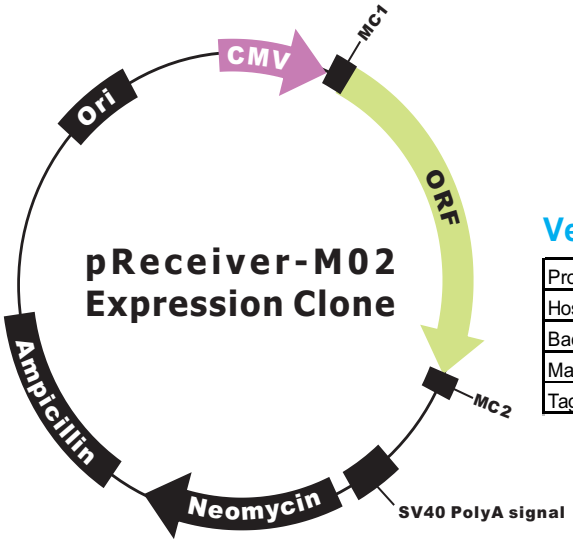
Vector Features

Promoter	CMV
Host Cell	Mammalian
Bacterial selection antibiotic	Ampicillin
Mammalian selection marker	Yes
Tag	3XFlag

Transfection

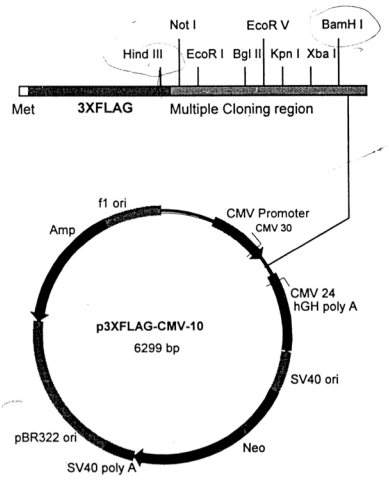
Reverse transfection of siRNA was performed as described previously. Silencer Select siRNA of DDX10, DDX11, DDX17, DDX42, RBM39, SMAD6, IGF2BP3 was purchased from Life Technologies. DDX10 and triple-flagged tagged SMAD6 vector were purchased from GeneCopoeia. The triplet-flag tagged empty vector was provided from Dr. Ricardo Dalla-Favera lab. The anti-flag antibody with beads was purchased from Sigma-Aldrich.

To perform reverse transfection in 6-well plate, firstly, mix 2 µl siRNA (10 mM) and 38 µl optiMEM and incubate for 5 minutes. On the other hand, mix 2 µl RNAiMAX and 38 µl optiMEM and incubate for 5 minutes. Add two mixtures together and incubate 20 minutes. After incub



Vector Features

Promoter	CMV
Host Cell	Mammalian
Bacterial selection antibiotic	Ampicillin
Mammalian selection marker	Yes
Tag	N/A



p3XFLAG-CMV-10 Features

Feature	Map Position
CMV promoter	166-916
CMV 30 sequencing primer	825-854
Translational initiation	928-930
3XFLAG sequence	931-996
Multiple cloning region	994–1056
hGH poly A	1061-1680
CMV 24 sequencing primer	1118-1141
SV40 ori	1699-2037
Neo	2073-2864
SV40 poly A	3511-3609
pBR322 ori	4528-4647
Ampicillin resistance	4824-5684
f_1 ori	5847-6299

Figure 4.4 Vector Map 3 constructs (DDX10, 3xFlag-SMAD6, 3xFlag)

ation, add 80 μl mixture with 420 μl media (without antibiotics) and 500 μl diluted cells (~10^5 cells) to each well.

To perform forward transfection, seed cells on the previous day. Mix 500 μl optiMEM and 15 μg of plasmid. On the other hand, mix 500 μl optiMEM and 10 μl lipofectamine 200 and incubate at room temperature for 5 minutes. Add two mixtures together and incubate for 30 minutes. Remove the media from the plate, and add 5 ml media without antibiotics and mixture for lipofection to each well. Due to the toxicity of lipofectamine, replace or add media after 6 or 8 hours after performing transfection.

RNA-immunoprecipitation

Prepare around 10^7 or two 15-centimeter plates with full confluency of cells after 24 hours of transfection. Wash the cells with PBS; then harvest the cells by trypsinization. Spin down cell by centrifugation at 1200 rpm for 5 minutes. Resuspend cell in 2 ml PBS, 2 ml freshly prepared nuclear isolation buffer and 6 ml water, and keep the cells on ice for 20 minutes with frequent mixing. Pellet nuclei by centrifugation at 2500g for 15 minutes. Then, discard the supernatant and resuspend the nuclear pellet with 1ml freshly prepared buffer C. Shear the nuclear membrane and chromatin by sonication at high level for 15 min. Then, pellet the nuclear membrane and debris by centrifugation at 13000 rpm for 10 minutes. Take 20 μl supernatant for loading control. Add beads conjugated with antibody to protein of interest (40 μl) to supernatant and incubate at 4°C overnight with gentle rotation. Pellet beads at 2500 rpm for 30 seconds, then remove supernatant, and resuspend beads with 1000 μl RIP buffer. Repeat washing three to five times with RIP buffer to reduce non-specific binding, followed by one time wash with PBS. Then, take

20 μl of beads for western blotting. Finally, extract RNA from beads using cell2cDNA in order to avoid loss of RNA yield. TaqMan primers of primary miRNAs for qRT-PCR were obtained from Life Technologies.

Statistical Analysis

All experiments were performed in triplicate. All data were shown in mean ± standard error. All statistical analyses were performed by MATLAB or R. The Kaplan-Meier plot was generated from NCI Repository for Molecular Brain Neoplasia Data (REMBRANT) using all glioma samples.

Cell Proliferation Assay

Cell proliferation assay for SMAD6: cells were seed in 12-well plate in triplicate (200 cells per well). Prepare 7 plates in total for one week. After 24 hours, remove the media and fix cells with 4% paraformaldehyde in PBS for 15 min and wash 1x with water. Stain with 0.1% crystal violet for 20 min (1 ml per 12-well). Aspirate and wash 3x with water and then aspirate and allow to air dry. Add 1 ml 10% acetic acid to each well and incubate 20 min with shaking in order to extract proteins. Swirl and pipet up and down the lysate. Finally, measure the absorbance of protein lysate at 590nm.

Cell proliferation assay for IGF2BP2 and IGF2BP3: cells were seeded in 16 well plates (E-plate 16, Roche) (4,000 cells in 150 μl medium/well), following the xCELLigence Real Time Cell Analyzer (RTCA) DP manual provided by the manufacturer (Roche). Cell index, which is proportional to the number of cells attached to the culturing surface, is recorded real time every 1-2 hours for up to 3-4 days. For each well, the cell index recorded 4 hours after seeding was used as the baseline to get fold

changes of cell indices afterwards. The time point of 4 hours after seeding is therefore used as time point zero in Figure 4.13 and Figure 4.16. Average fold changes of cell index were calculated based on at least two measurements from 4 replicate experiments ± SE.

Preparation of lentivirus

Plate around $6x10^6$ HEK293T cells in 15 cm dish day before transfection (alternatively, split one confluent 15 cm dish into three 15 cm dishes). Ninety to ninety-five percent of cell confluence is expected when performing transfection. After 24 hours, aspirate supernatant and replace with 15 ml of fresh medium (DMEM, 10% FBS, no antibiotics). Prepare the mixture of transfection reagent, including 840 ng of pVSV-G, 7500 ng of pdelta8.9, 8400 ng of pLKO-shSTK38, 1500 μl of jetPrime buffer and 60 μl of jetPrime reagent. Vortex the mixture for 10 sec, spin down and incubate at room temperature for 10 minutes. Then, add the transfection mixture to the cells, and add 20 ml of fresh medium in the evening, so that the total volume of medium for 15 cm plate will be 35 ml. On day 3, replace medium with fresh 35 ml of medium. On day 4, collect the supernatants with lentivirus into 50 ml conical tubes and add 35 ml of fresh medium to plate. Filter supernatant with Steriflip filtration tubes and place the tubes on ice and store in the cold room until next day. On day 5, collect 35 ml supernatant for the 2^{nd} time. Filter and pool the supernatants together. Combine 3 volume of supernatant with 1 volume of Lenti-X (around 10 to 12 ml) and then incubate mixture at 4°C overnight. On the last day, centrifuge the sample at 1,500xg for 45 min at 4°C. Remove supernatant carefully and resuspend off-white pellet in 200 μl of PBS. Finally, titrate the virus with Lenti-X stick, split 50 μl of virus in each tube and store in -80°C for future experiments.

Infection of lentivirus

Seed 10^5 cells in 6-well plate. Next day, mix 5μg/ml polybrene, 100 μl virus and add medium to make total volume 600 μl. Then, add the mixture into the well directly. On day 3, replace medium with fresh 2 ml of medium. On day 4, replace medium with 2 μg/ml puromycin for selecting successfully transfected cells (note that the concentration of puromycin depends on cell line). After 24 or 48 hours, check cell survival rate. The survived cells can be collected for measuring the efficiency of knocking down using qRT-PCR.

4.3 Results

Identification of post-transcriptional miRNA regulators

Two thirds of miRBase miRNAs (Rybak, Fuchs et al. 2008) are located within transcriptional units that include other miRNAs, non-coding RNAs or mRNAs. To identify the post-transcriptional miRNA regulators, we restricted the study of miRNAs and their hosts to instances where deviations are associated with lower mature miRNA abundance, we exclude deviations due to independent miRNA transcription (Ozsolak, Poling et al. 2008). Histone modification of H3K4me3 and H3K27ac are known histone markers for transcription start sites and active regulatory elements. Therefore, we also used chromHMM data from ENCODE to examine the transcription start sites of the intragenic miRNAs (Ernst, Kheradpour et al. 2011). We found 234 out of 290 intragenic miRNAs do not have the sign of independent promoters.

We screened for post-transcriptional regulators of intergenic miRNAs that are co-transcribed with other miRNAs, and intragenic miRNAs that are co-transcribed with their host genes (Rodriguez, Griffiths-Jones et al. 2004; Gennarino, Sardiello et al. 2009). Considering the co-transcribed RNAs that are significantly correlated in glioblastoma tumors (Lambertz, Nittner et al. 2010), samples with relatively improved or lost correlation harbor the footprints of tumor-specific post-transcriptional regulation. More precisely, deviations in correlations between expression profiles of intragenic miRNAs and their hosts that results in lower mature miRNA abundance, and deviations in correlation between expression profiles of co-transcribed miRNAs must be due to post-transcriptional events. Our main premise is that RBPs whose expression is significantly predicative of such deviations may contribute to the regulation of miRNAs. Consequently, we expect perturbations that target these RBPs to affect the abundance of one miRNA species but not its co-transcribed counterpart, and perturbation to regulators that enhance or inhibit mature miRNA biogenesis are expected to result in matching depletion or accumulation, respectively, of the corresponding intermediary miRNA products.

Gene	pFisher	Pairs			
PIWIL2	1.89E-24	GABRE:hsa-miR-452_s	MIRH1:hsa-miR-19a	MIRH1:hsa-miR-20a	MIRH1:hsa-miR-19b
IGF2BP	9.27E-19	HOXC10:hsa-miR-196a	MIRH1:hsa-miR-17	GRM8:hsa-miR-592	
LIN28	2.28E-15	MIRH1:hsa-miR-92a	MIRH1:hsa-miR-18a_s		
NXF5	1.80E-14	hsa-miR-183:hsa-miR-96	GRM8:hsa-miR-592		
RBMS1	9.39E-12	hsa-miR-23b:hsa-miR-24	GPC1:hsa-miR-149		
PTRF	1.36E-10	GPC1:hsa-miR-149	EFEMP1:hsa-miR-217		
NXF3	2.01E-10	MIRH1:hsa-miR-17	HOXC10:hsa-miR-196a		
ILF3	1.17E-09	ZNRF2:hsa-miR-550a_s	RTCD1:hsa-miR-553		
DDX10	5.61E-07	SLIT3:hsa-miR-218			
C14orf21	5.61E-07	TRPM3:hsa-miR-204			
EIF4G1	5.61E-07	SREBF2:hsa-miR-33a			
NOL4	5.61E-07	TRPM3:hsa-miR-204			
PCBP3	5.61E-07	GABRE:hsa-miR-224			
DDX11	5.61E-07	GPR157:hsa-miR-34a			
HNRNPA1	5.61E-07	FAM114A1:hsa-miR-574-3p			
HNRNPC	5.61E-07	MIRH1:hsa-miR-17_s			
MEX3C	5.61E-07	CHRM2:hsa-miR-490-3p			
ERN1	1.68E-06	GRM8:hsa-miR-592			
SNRPG	1.68E-06	CTDSP2:hsa-miR-26a			
XPO1	1.68E-06	EGFL7:hsa-miR-126_s			
SMAD6	2.25E-06	ARRB1:hsa-miR-326			
DHX9	2.25E-06	EML2:hsa-miR-330-3p			
ELAVL3	3.93E-06	MIRH1:hsa-miR-18a			
DDX17	3.93E-06	ZNRF2:hsa-miR-550a_s			
EIF1AY	4.49E-06	PDE2A:hsa-miR-139-5p			
DUSP11	5.05E-06	EVL:hsa-miR-342-3p			
LIN28B	5.61E-06	GRID1:hsa-miR-346			
DNAJC17	6.74E-06	PDE4D:hsa-miR-582-5p			
DDX42	7.30E-06	WDR82:hsa-let-7g			
PAPD4	8.42E-06	miR-524-3p:miR-517_s			
PNPT1	8.42E-06	CTDSP1:hsa-miR-26b			
RBM24	8.42E-06	CLCN5:hsa-miR-500a_s			
ESRP2	9.54E-06	MIRH1:hsa-miR-92a			

Table 4.1 Predicted miRNA biogenesis regulator from TCGA GBM data set

The pairs are the host-miRNA or miRNA cluster pairs regulated by the candidate in the prediction (conditional mutual information, p-value < 1e-5). The highlighted genes are validated miRNA biogenesis regulators

Information theory has been successfully applied to infer transcriptional and posttranslational network previously. Here, we use conditional mutual information identify the relationship of host-miRNA-regulator (p-value < 1e-5), meaning a miRNA regulator should be able to significantly improve the relationship between intragenic miRNA and its host mRNA. Then, we use Fisher's method to determine the significance of each predicted miRNA RBP regulator. In total, our screen identified 33 candidate RBP regulators that target miRNAs, including regulators that are expressed in tumor specific manner (Table 4.1). Among candidate RBP regulators, seven are known miRNA biogenesis regulators. DDX17, HNRNPA1 increase miRNA level through binding to primary miRNAs and facilitate DROSHA processing. On the other hand, LIN28A and ILF3 repress the biogenesis of targeted miRNAs by competing against DROSHA binding to miRNA precursors. In our prediction, LIN28A is able to modulate the correlation between miR-18, miR-92 and their host transcript, MIR17HG. The correlation between MIR17HG and miR-18 or miR-92 both increase under lowly-expressed LIN28A (spearman correlation, p-value < 1e-4), suggesting LIN28A plays a minor inhibitory role in the miRNAs other than the widely-known let-7 family.

ERN1 is able to terminate miRNA biogenesis through cleaving miRNAs during ER stress (Stefani and Slack 2008). Several miRNAs, such as miR-17, miR-34a, miR-96 and miR-125b, are rapidly decayed by ERN1 in response to ER stress. We predicted ERN1 regulates the level of miR-592. MiR-592 and its host, GRM8 is significantly correlated when ERN1 is lowly expressed (spearman correlation, p-value < 1e-4), suggesting miR-592 is another new target of ERN1 cleavage (Figure 4.5).

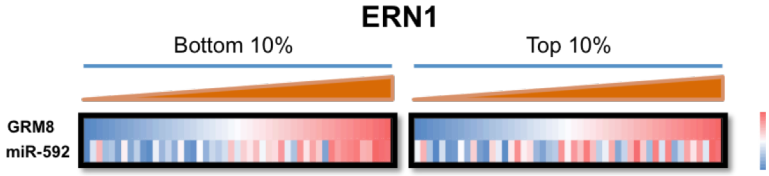


Figure 4.5 The relationship between miR-592 and its host, GRM8 is associated with ERN1 expression.

ERN1 is a validated miRNA post-transcriptional regulator.

ILF3 is known to bind small double stranded RNA, such as adenovirus viral-associated RNAs and miRNA precursors. Heterodimer ILF2-ILF3 associates with over-expressed DROSHA and interferes miRNA biogenesis by competing against DGCR8 for binding to the basal region of certain pri-miRNAs (Sakamoto, Aoki et al. 2009). Previous study showed overexpressing ILF3 causes the accumulation of precursors of miR-30a, let-7a, miR-15a, miR-16 and miR-21, and the heterodimer ILF2-ILF3 associates with these precursors. Here, we predicted ILF3 participates in maturation of miR-550a-5p and miR-553. DHX9 contains a DEAH domain, which function as RNA helicase and alters RNA secondary structure. It interacts with BRCA1-DROSHA duplex and several pri-miRNAs. Silencing DHX9 in HeLa cell decreases the level of mature let-7a, miR-16, miR-145 and miR-34a. These pri-miRNAs have physical interactions with DHX9 as well. Here, we identified DHX9 as biogenesis regulator by affecting the relationship between miR-330 and its host gene, EML2.

Biochemical target identification

To test the validity of our method to identify miRNA biogenesis regulators, we use qRT-PCR to measure the pri-miRNA and mature miRNA. First, we observe silencing DDX17, a validated miRNA posttranscriptional regulator, causes the downregulation of several mature miRNAs in SNB19 cell. While depleting another RNA-binding protein,

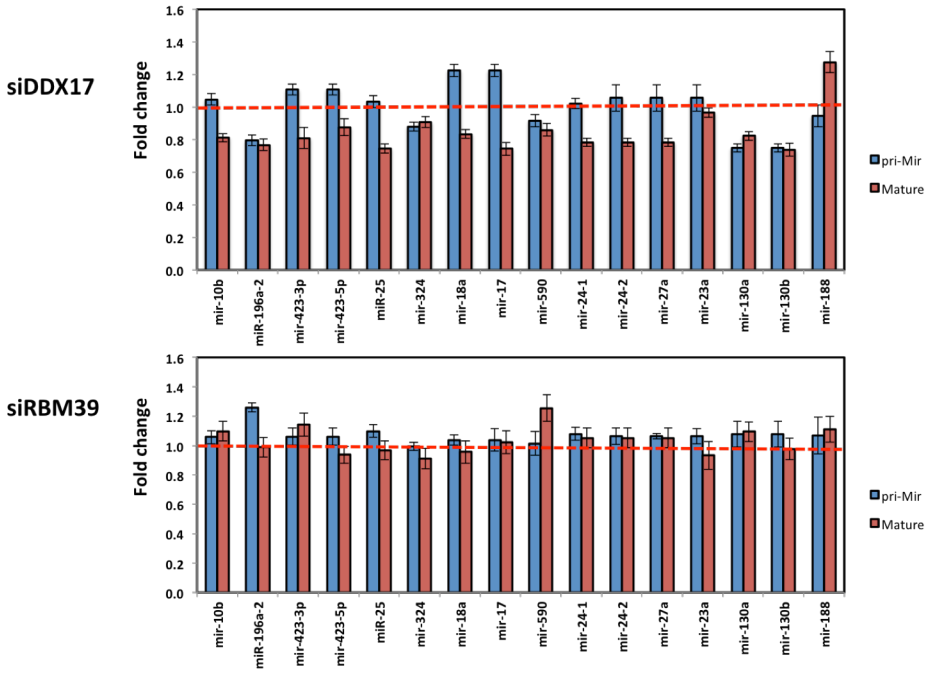


Figure 4.6 DDX17 and RBM39 as positive and negative control in qRT-PCR validation.

DDX17 is a validated miRNA biogenesis regulator, and silencing DDX17 affects the level of mature miRNA and miRNA precursors. On the other hand, silencing RBM39 has no effect on selected miRNAs.

RBM39, which is not identified in our prediction, has no effect on the tested mature miRNAs or precursors (Figure 4.6).

Since three genes containing DEAD or DEAH domain (DDX5, DDX17 and DHX9) have been shown to modulate miRNA at posttranscriptional stage, we firstly

selected three candidates in DDX family throughout the prediction for further validation. DDX10 regulates the maturation of miR-218, which is co-transcribed with SLIT3 in our prediction. Overexpression of DDX10 only increases mature miR-218 level by 30% but not affecting the host gene, SLIT3. Depletion of DDX10 and DDX42 both downregulate the level of mature miRNAs, including miR-423-3p, miR-423-5p, miR-24, miR-590, miR-324 and miR-25 (Figure 4.8). Meanwhile, the levels of miRNA precursors remain

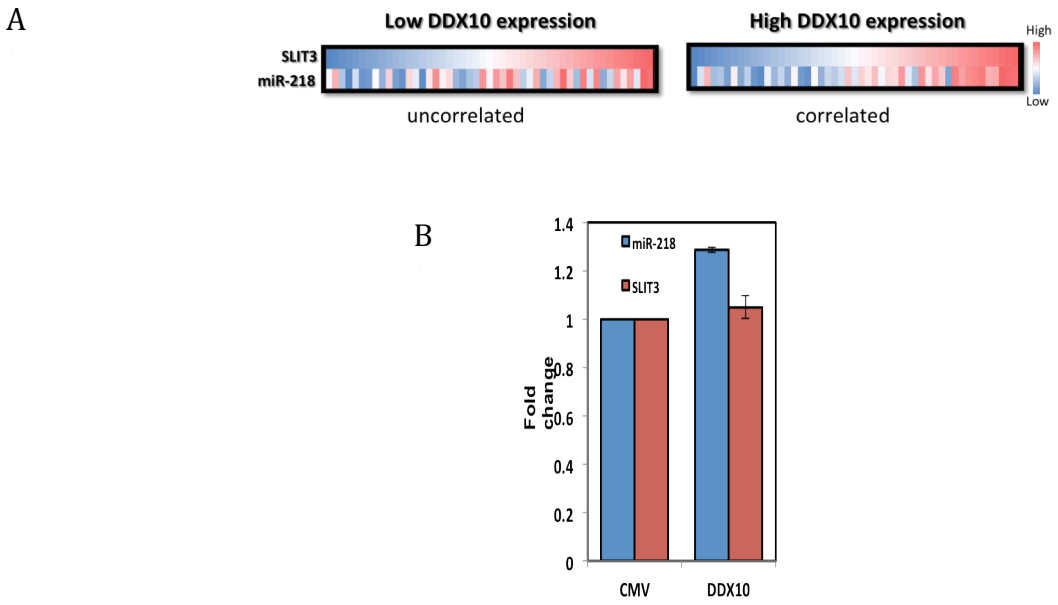


Figure 4.7 DDX10 regulate the maturation of miR-218.

(A) RNA-binding protein, DDX10 affects the relationship between host mRNA, SLIT3 and intragenic miRNA, miR-218. The host mRNA and intragenic miRNA are only correlated when DDX10 is highly expressed. (B) RT-PCR showed overexpression of DDX10 upregulates miR-218 by 30% but not SLIT3 level.

unaffected or upregulated indicating both DDX10 and DDX42 modulate mature miRNA level at posttranscriptional stage. On the other hand, silencing DDX11 has similar effect as DDX10 on miR-423-3p: both downregulate mature miR-423-3p and causes accumulation of pri-miR-423.

However, DDX11 seems to interfere the processing of miR-10b, miR-196a-2 and miR-24. All these mature miRNAs are significantly increased upon silencing DDX11. We selected these candidates throughout the prediction and validated with qRT-PCR. The results showed, in addition to DDX5, DDX17 and DHX9, more proteins in DDX family are involved in miRNA maturation (Figure 4.8).

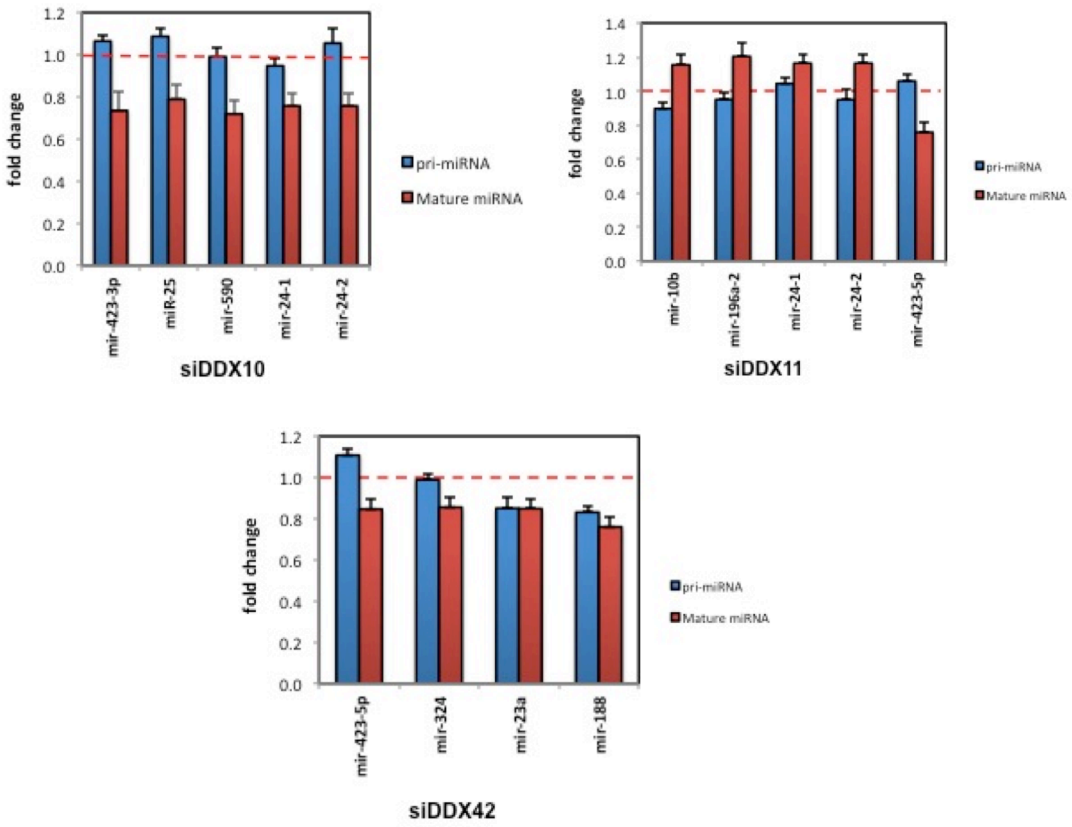


Figure 4.8 More DDX family members participate in miRNA maturation.

Various mature miRNA and miRNA precursors differentially expressed under siRNA mediated silencing DDX10, DDX11 and DDX42

SMAD6 regulate miRNA biogenesis

SMAD proteins contain three classes: the receptor-regulated SMADs (R-SMAD), the common-mediator SMADs (co-SMAD), and the antagonistic SMADs (I-SMAD). The R-SMADs, including SMAD1, SMAD3 and SMAD5, promote the maturation of a subset of miRNA by binding to SMAD-binding elements on miRNAs and facilitating DROSHA cleavage. SMAD6 is a member of I-SMAD, a class that counteracts against R-SMAD in transcriptional regulation. However, SMAD6 possess the same protein domains as R-SMAD proteins, and we predicted SMAD6 a candidate miRNA biogenesis regulator like other R-SMAD. To validate if SMAD6 regulates miRNA maturation, we silenced SMAD6 with SMAD6-specific siRNA and performed miRNA profiling using Wafergen. We found 9 out of 16 predicted miRNAs are significantly affected when silencing SMAD6 (Fisher's exact test, p-value < 0.01), suggesting SMAD6 has similar function with other R-SMADs in miRNA biogenesis (Figure 4.9A). From the downregulated miRNA precursor sequences, we identified a motif similar with the consensus sequences of SMAD binding element. (Figure 4.9 B)

To examine whether SMAD6 affects miRNA biogenesis post-transcriptionally, we selected miRNAs that are highly expressed in SNB19 then measured the level of pri-miRNA and mature miRNA with qRT-PCR simultaneously. When silencing SMAD6, 10 selected mature miRNAs are significantly downregulated, while the levels of their primary transcripts are all increased. It indicates SMAD6 affects the miRNA maturation

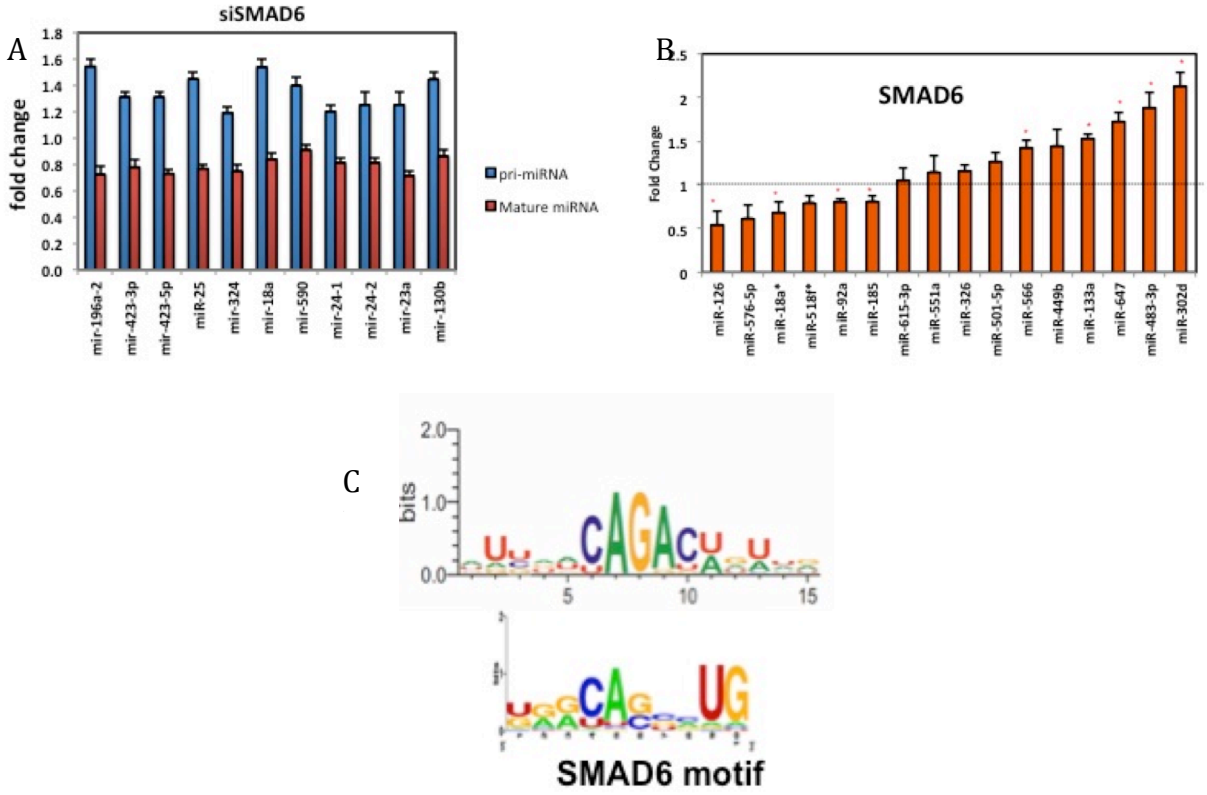


Figure 4.9 SMAD6 regulate miRNA post-transcriptionally.

(A) Ten mature miRNAs are downregulated while their eleven precursors are accumulated when knocking down SMAD6. Noted that miR-24 can be expressed from two different loci (B) SMAD6 has two modes of regulations of miRNAs. The levels of mature miRNAs were measured by Wafergen. MiRNAs with Asterisk represent the candidate miRNAs regulated by SMAD6 in the prediction (p-value <1e-2) (C) The sequence motif identified from SMAD6 affected miRNA precursors is similar with the motif recognized by SMAD1, SMAD3 and SMAD5 in Davis *et al.*

at post-transcription stage (Figure 4.9 A). Furthermore, we biochemically tested if SMAD6 physically associates with miRNA precursors using RNA-immunoprecipitation (RNA-IP). RNA-IP indicated that SMAD6 interacts with the primary transcripts of tested miRNAs (mir-196-2, mir-324, mir-590, mir-18a and mir-130b), but not with mir-10b where the mature miRNAs and primary transcripts are not differentially expressed by knocking down SMAD6 (Figure 4.10).

A B

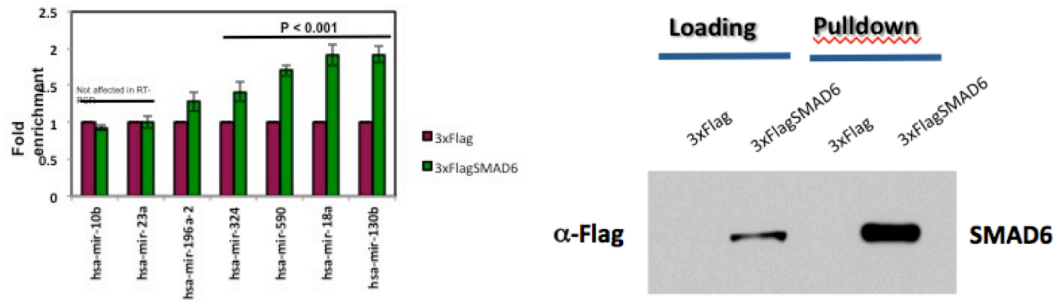


Figure 4.10 SMAD6 interacts with pri-miRNA.

(A) RIP showed SMAD6 interacts with five pri-miRNAs, mir-196-2, mir-324, mir-590, mir-18a and mir-130b, which are significantly accumulated when knocking down SMAD6. The fold changes are the enrichment status of pri-miRNA after 24hr ectopic expression of 3X-flag-tagged SMAD6 compared to 3X-flag-tagged empty vector. (B) Western blotting showed SMAD6 is enriched after pulling down with anti-flag antibody.

IGF2BP2 and IGF2BP3 regulate miRNA biogenesis

IGF2BP1 is another predicted miRNA biogenesis regulator. IGF2BPs family includes IGF2BP1, IGF2BP2 and IGF2BP3, and they all carry two RNA recognition motifs, six hnRNP K homology domains and have identical binding profiles. Due to IGF2BP1 is poorly expressed in SNB19 cell; we validated the role of the highly expressed IGF2BP2 and IGF2BP3 in miRNA regulation. We profiled miRNA expressions with Wafergen, after siRNA-mediated silencing of IGF2BP2 and IGF2BP3.

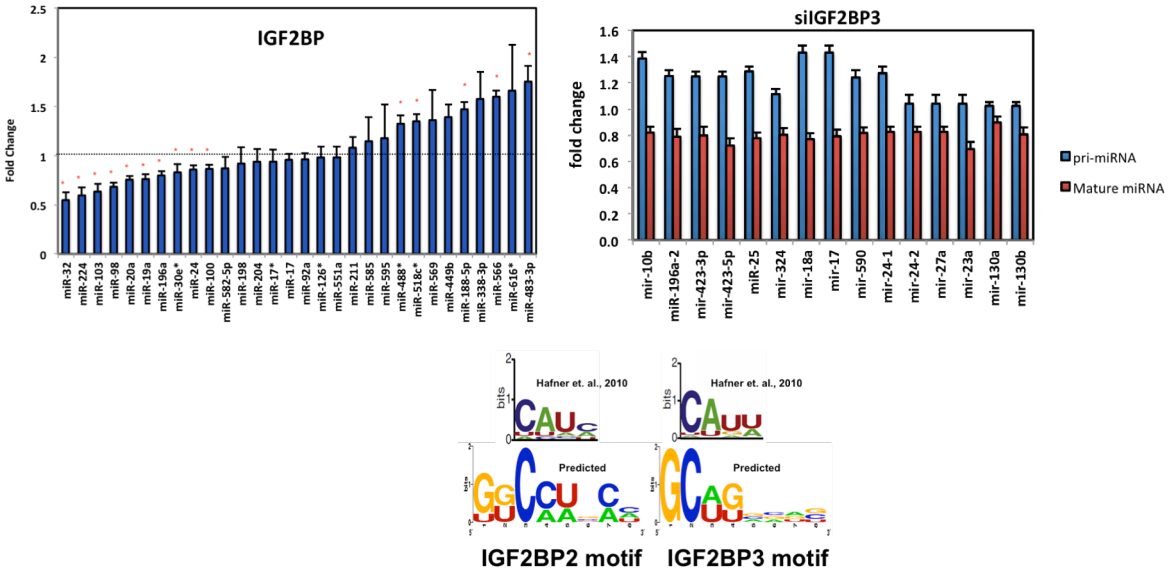


Figure 4.11 IGF2BP3 regulates miRNA level post-transcriptionally.

(A) Fifteen mature miRNAs are significantly repressed when knocking down IGF2BP3 while eight of their precursors are accumulated. Noted that the precursors of miR-423-5p and miR-423-3p, miR-18a and miR-17 were measured by same primers. (B) The levels of mature miRNAs were measured by Wafergen, and the candidate IGF2BP3-regulated miRNAs tend to be downregulated when silencing IGF2BP3. MiRNAs with Asterisk represent the candidate miRNAs regulated by IGF2BP3 in the prediction (Fisher's exact test, p-value <0.004) (C) The sequence motif found from IGF2BP3 affected miRNA precursors compared to IGF2BP3 binding motif from PAR-CLIP

Fifteen out of thirty predicted miRNAs are significantly affected upon silencing IGF2BPs (Fisher's exact test, p-value < 0.004). Similar with SMAD6, qRT-PCR also showed the fifteen highly expressed miRNAs are downregulated by silencing IGF2BP3, and their primary transcripts are all accumulated (Figure 4.11 A). This indicates IGF2BP3 is a posttranscriptional regulator of miRNA biogenesis, and knocking down IGF2BP3 interferes the processing from pri-miRNA to mature miRNA. Notably, miR-17 is a predicted IGF2BP target, but it is not differentially expressed in the Wafergen system (Figure 4.11 B). We validated miR-17 with RT-PCR again and found out the mature and primary transcripts of miR-17 were both affected upon IGF2BP3 perturbation. The previous work of PAR-CLIP showed IGF2BP2 and IGF2BP3 recognize a (CAUH) motif on bound mRNA. We identified a similar motif from the validated miRNA sequences (Figure 4.11 C).

4.4 Discussion

The maturation of miRNA is finely controlled transcriptionally and post-transcriptionally, and the miRNA level is important to miRNA-mediated regulation. Study has shown lowly expressed miRNA may not have the ability to regulate target mRNA expression (Mullokandov, Baccarini et al. 2012). Moreover, a large number of miRNAs have significantly lower expression levels in tumors and cancer cell lines than in normal tissues, indicating the miRNA-mediated regulation is dysregulated in various types of tumors (Lu, Getz et al. 2005). Currently, a dozen of posttranscriptional regulators have been identified modulating miRNA level in various contexts by interacting with miRNA precursors, DROSHA and DICER1. Here, we performed a computational screening to identify candidate miRNA biogenesis regulators from RNA

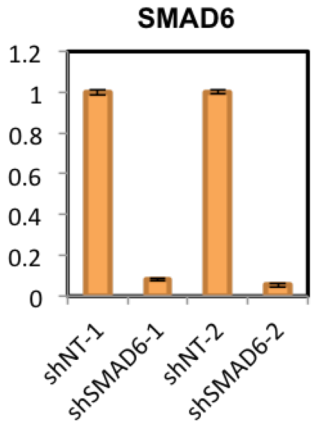
binding proteins. We predicted 33 RBPs may be involved in miRNA regulation, and seven of them are validated miRNA regulators. We carried out qRT-PCR to measure the pri-miRNA and mature miRNA level upon perturbing the predicted miRNA regulators. Silencing DDX10, DDX11, DDX42, SMAD6 and IGF2BP3 cause a subset of miRNAs and their precursors significantly differentially expressed. Furthermore, we showed the physical interactions between SMAD6, IGF2BP3 and miRNAs.

SMAD6 is an inhibitory SMAD, and its major function is to by interfering the interaction between SMAD4 and R-SMADs. Current understanding of SMAD6 binding to nucleic acid is not clear. However, like other SMAD family member, SMAD6 contains a Mad Homology 1 (MH1) domain for nucleic acid binding. Previous study has shown SMAD6 is able to regulate Id2 and Hex by binding to their promoter regions and is required cell differentiation in HL-60 cell (Lu, Getz et al. 2005; Glesne and Huberman 2006). In this study, we further identify SMAD6 possess the ability to bind miRNAs and modulate miRNA maturation.

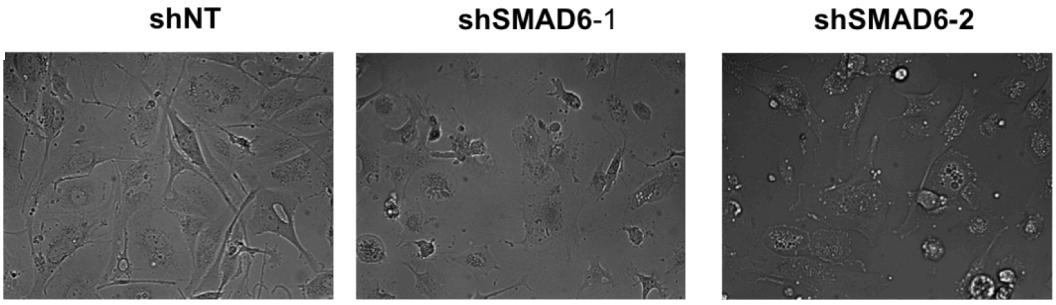
To study SMAD6 function in glioblastoma, we silenced SMAD6 in SNB19 using lentivirus carrying SMAD6-specific shRNA. After puromycin selection for 48 hours, only cells infected by lentivirus survived, and the survived cells are stable cell lines. SMAD6 level in this stable cell line was knocked down by around 90% and their cell morphology turned into irregular while the morphology of stable cell line with non-targeting control is still similar with original SNB19. This suggests SMAD6 could block epithelial-mesenchymal transition (EMT) in glioblastoma context (Figure 4.12 B). Previous studies have shown SMAD6 inhibits EMT during cardiac valve formation and in the proepicardium (Desgrosellier, Mundell et al. 2005; Olivey, Mundell et al. 2006).

However, the cell proliferation assay showed there is no significant difference between shSMAD6 and shNT cell lines in proliferation (Figure 4.13). Interestingly, knocking down SMAD6 in SNB19 cell using SMAD6-specific shRNA or siRNA both induced significant reduction of DICER1 mRNA level while DROSHA is not affected. This suggests the transcription of DICER1 may be regulated by conventional SMAD pathway (Figure 4.12 C).

A



B



C

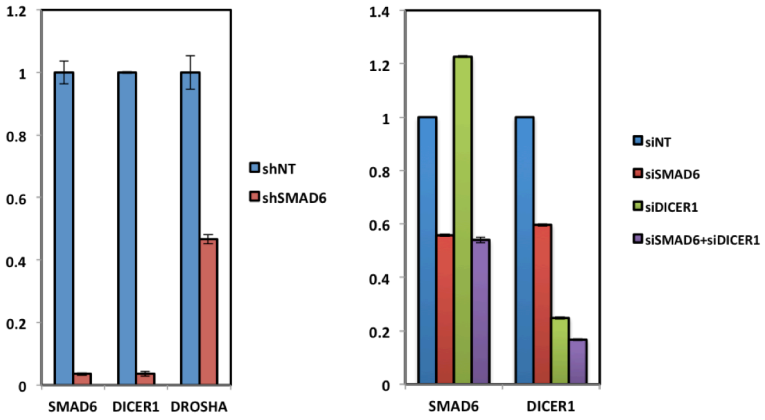


Figure 4.12 Silencing SMAD6 using shRNA in SNB19 cell.

(A) SMAD6 mRNA level were repressed by 90% in the two shSMAD6 stable cell lines (B) Two shSMAD6 cell lines have significant morphological change while shNT cell line is still like original SNB19 cell (C) Silencing SMAD6 with shRNA also brings down DICER1 expression. The association between SMAD6 and DICER1 was also validated using SMAD6 siRNA.

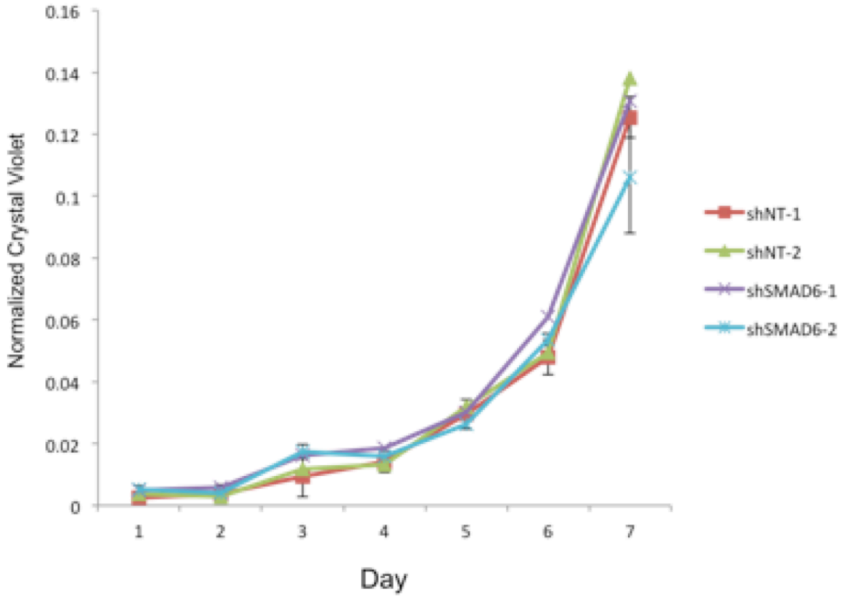


Figure 4.13 Silencing SMAD6 has little effect on cell proliferation.

IGF2BP family (IGF2BP1, IGF2BP2 and IGF2BP3) carry six RNA binding domain: four KH domains and two RRM domains, and their protein domain structures and binding motif are highly similar. Based on PAR-CLIP data, they tend to bind exonic

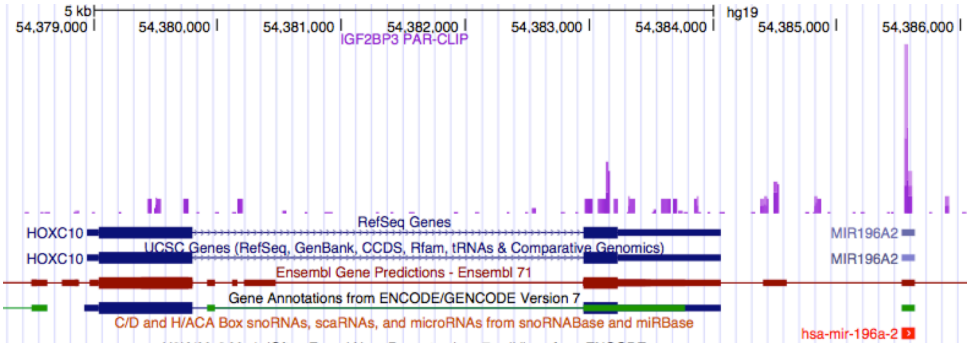


Figure 4.14 IGF2BP3 interacts with miR-196-2.

region and have slight preference to 3' UTR. The binding of IGF2BP protein can prevent mRNA from degradation. From PAR-CLIP of IGF2BP family from HEK293 cell, we found these proteins not only interact with mRNA but with miRNA as well. Each IGF2BP protein interacts with roughly 400 miRNAs in HEK293 cell with more than 10 RPM. This observation indicates IGF2BP family has the ability to interact with miRNA and slow the miRNA turnover rate. Figure 4.10 showed IGF2BP3 binds to miR-196a-2, which is predicted to be target of IGF2BP1 in our prediction.

IGF2BP3 is upregulated in various types of tumor and modulate miRNA level in our prediction. To examine the biological significance of IGF2BP3 regulating miRNA level, we screened for the common targets of IGF2BP3 modulated miRNAs, and one common candidate is the tumor suppressor gene, PTEN. The miRNA profile from qRT-PCR showed IGF2BP3 regulates the maturation of multiple miRNAs (miR-26a, miR-19a, miR-19b, miR-106a, let-7b, let-7c) known to target PTEN. MiR-26a was also bound by IGF2BP family (Figure 4.15). We measured PTEN 3' UTR luciferase activity in SNB19

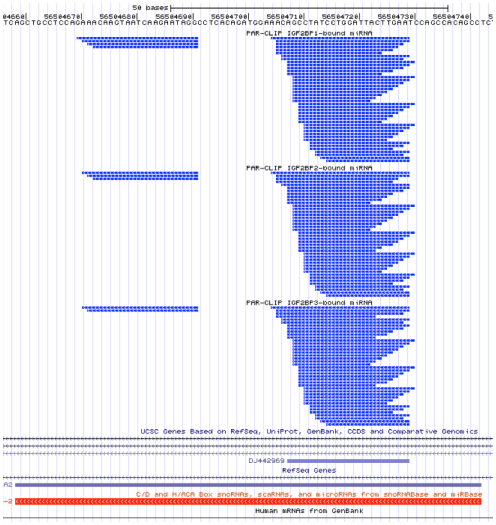
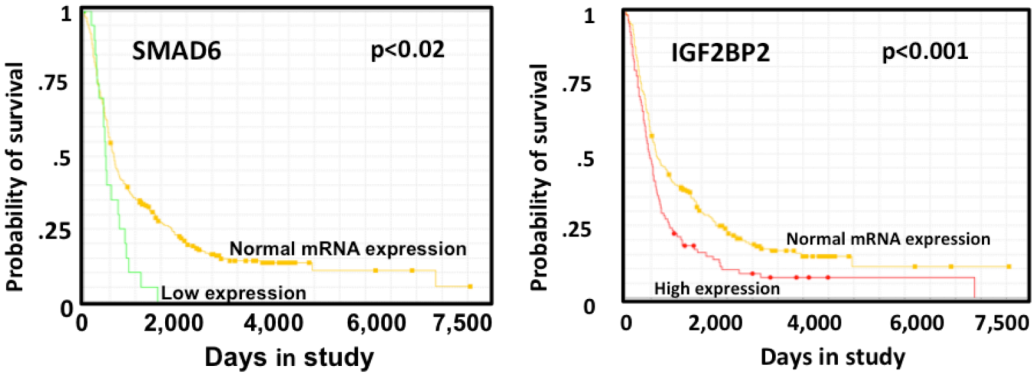
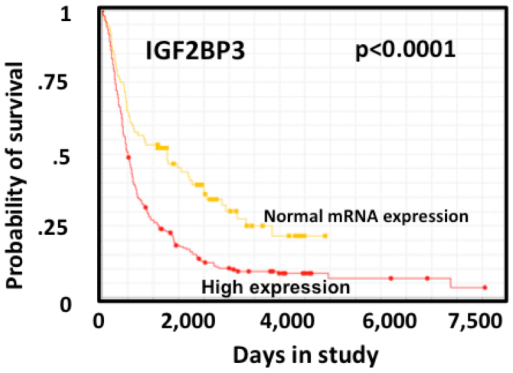


Figure 4.15 IGF2BPs binds to the oncomiR, miR-26a.

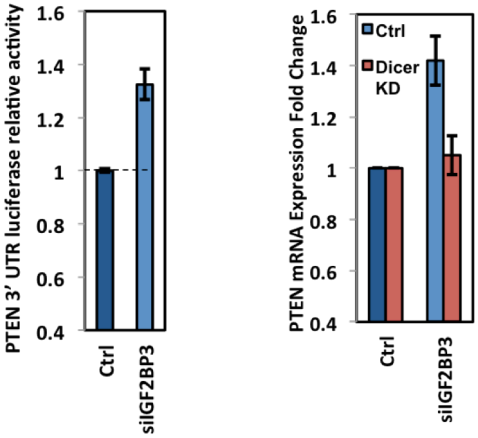
cells transfected with mock and IGF2BP3-specific siRNA. The result showed PTEN 3' UTR luciferase activity is increased by nearly 40% with IGF2BP3 knockdown, indicating IGF2BP3 regulates PTEN through PTEN 3'UTR. To test the regulation of IGF2BP3 on PTEN is miRNA mediated, we co-silenced DICER1 and IGF2BP3 and measured PTEN mRNA level (Figure 4.16 B). The expression of PTEN is significantly lower when IGF2BP3 and DICER1 were knocked down simultaneously than IGF2BP3 alone, suggesting that silencing DICER1 abrogated the regulation of IGF2BP3 on PTEN. Therefore, the regulation of IGF2BP3 on PTEN is miRNA mediated. PTEN is known to induce apoptosis and cell cycle arrest; therefore, we asked whether silencing IGF2BP3 have significant physiological effect through PTEN. IGF2BP3 is amplified in 40% of samples in TCGA GBM data set (aCGH ratio above 0.3). The REMBRADT database showed glioma patients with high IGF2BP3 expression (>= 2.0X) have significantly shorter survival rate than patients with normal IGF2BP3 expression (p-value < 1e-4 by log-rank test, Figure 4.16 A). The same clinical outcome also happens on the glioma patients with high IGF2BP2 level (>= 2.0X). We measured the cell proliferation rate of SNB19 cells transfected with IGF2BP3-specific shRNA. Knocking down IGF2BP3 decreases PTEN targeting miRNAs, so the increased PTEN is able to slow down the cell proliferation (Figure 4.16 C).

A


SMAD6
p<0.02
Probability of survival
Normal mRNA expression
Low expression
Days in study
IGF2BP2
p<0.001
Normal mRNA expression
High expression
Days in study


IGF2BP3
p<0.0001
Probability of survival
Normal mRNA expression
High expression
Days in study

B


PTEN 3' UTR luciferase relative activity
Ctrl
siIGF2BP3
PTEN mRNA Expression Fold Change
Ctrl
Dicer KD
Ctrl
siIGF2BP3

C

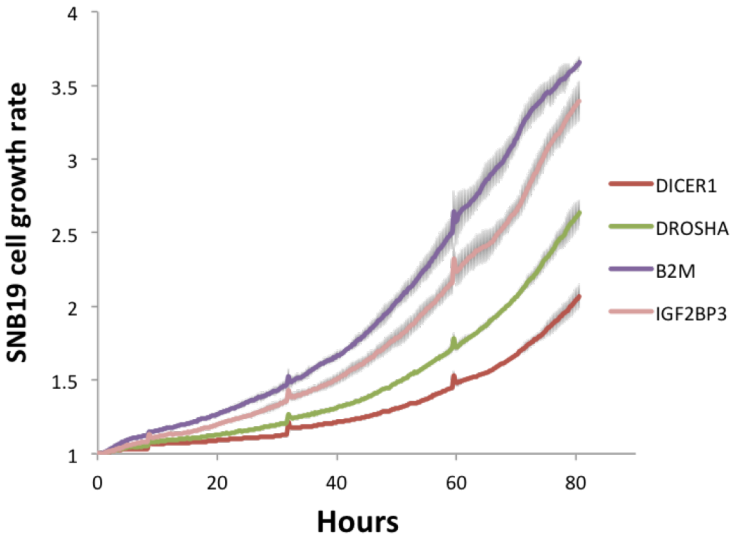

SNB19 cell growth rate
Hours
DICER1
DROSHA
B2M
IGF2BP3

Figure 4.16 IGF2BP3 associates with patient prognosis and cell growth.

(A) Kaplan-Meier plot showed the prognosis status of glioma patients in REMBRANT given SMAD6, IGF2BP2 and IGF2BP3 expression. Patients with high IGF2BP3 expression (upregulated more than 2 folds) have worse survival curve compared to patients with intermediate IGF2BP3 expression (upregulated and downregulated less than 2 folds). Statistical significance was determined by log-rank test. (B) IGF2BP3 regulates PTEN level via miRNA regulation (C) Cell proliferation assay of SNB19 cells was performed at 5 minutes interval and up to 70 hours, following shRNA mediated IGF2BP3, DICER1 and DROSHA silencing. Silencing B2M was used as negative control.

Chapter V

Identification for miRNA activity modulators

5.1 Introduction

MiRNA-mediated post-transcriptional regulation has emerged as a key gene regulation mechanism in diverse biological pathways. These small RNAs recognize target mRNAs and repress their expression through two distinct pathway, mRNA degradation or translational repression. It is estimated around 30% to 60% genes are under regulation of miRNA, and each miRNA can target up to 200 different mRNAs.

The initial observation of miRNA-mediated regulation was observed in *C. elegans*. Lee et al. found lin-4 small RNA is partial complementary to the 3' UTR of its target mRNA, lin-14 (Wightman, Ha et al. 1993). Subsequently, lin-4 and let-7 also found complementarity in 3' UTR of lin-28 and lin-41, respectively (Seggerson, Tang et al. 2002). When hundreds of miRNAs identified using computational prediction or biochemical cloning from various model organisms, people realize this type of small RNA brings huge impact to our understanding of gene regulation. In miRNA field, the most important problem is how miRNAs recognize their target mRNAs. The first observation is that the pairing between miRNAs and their target mRNA followed Watson-Crick sequence complementarity. Plant miRNAs usually form nearly perfect RNA duplex with their target mRNAs and induce subsequent post-transcriptional regulation. This phenomenon indicates it is simpler to identify miRNA targets in plants merely based on sequence complementarity. In animal, partial sequence complementarity between miRNA and mRNA is sufficient to induce miRNA-mediated regulation. Animal

miRNAs' 2^{nd}-7^{th} nucleotides are called "seed" region. Perfect pairing between seed region and target 3' UTR is essential for miRNA-mediated regulation. Multiple validations in different organisms all showed that substituting nucleotides in the seed region dramatically reduce the ability of miRNA targeting its target. This observation became the major principle of miRNA target prediction algorithms. Notably, more and more non-canonical miRNA-target interactions, where no perfect match between seed region and 3' UTR, were identified using crosslinking and high-throughput sequencing technology. The miRNA seed region is usually the most evolutionary conserved region in a mature miRNA, meaning miRNA seed region tend to be preserved under selection pressure. On the other hand, miRNA target sites are more conserved as well compared to flanking 3'UTR region. This feature was adopted as additional principle by many target prediction algorithm as well (Lu, Getz et al. 2005; Bartel 2009).

In addition to the 5' seed region of miRNA, 3' end of miRNA participates in pairing target mRNA in certain situations. Although seed region pairing is sufficient to initiate miRNA regulation, thirteenth-to-sixteenth nucleotides of mature miRNA pairing, 3'-supplementary sites, may serve as a supplement to seed pairing (Shin, Nam et al. 2010). In this case, the 3'-supplementary sites on miRNA target also tend to be more evolutionally conserved than flanking region. Pairing between 3' end of miRNA and target mRNA sometimes provide more than supplementary role. It can compensate the minor imperfect pairing, such as one nucleotide bulge or internal loop, between seed region and miRNA target site although this situation is atypical (Shin, Nam et al. 2010).

Computational predicting miRNA target always results in dozens of highly significant candidate targets. To obtain true candidates from the prediction, we can

validate the target using genetic or biochemical methods. In principle, genetic methods validate the miRNA-target interaction *in vivo* by performing gain or loss of function experiments and observing the phenotypes. The advantage of genetic method is the identified targets will be important physiologically in the tested model organism and disrupting the regulation of miRNA to the target causes the observable mutant phenotype. The caveat is the identified genetic interaction could be argued that it is an indirect interaction between miRNA and target; meaning the tested miRNA recognizes other target mRNA that mediates the expression of hypothesized target. Also, if the phenotype is controlled by multiple genes, which is common in higher organisms, then the phenotype may be too subtle to be detected. On the other hand, biochemical methods provide direct evidence to illustrate the interaction between miRNA and target mRNA. Earlier approaches include knocking down or overexpressing miRNA and measure the expression of predicted targets using luciferase assay, qRT-PCR, western blotting or ELISA. Furthermore, the 3'UTR sequence can be altered by site-directed mutagenesis in order to find out the precise miRNA binding sites. Recently, two methods, HITS-CLIP and PAR-CLIP, were developed utilizing crosslinking of protein-RNA interactions and immunoprecipitation (CLIP) to detect the precise RNA-binding protein binding sites on the bound RNAs (Chi, Zang et al. 2009; Hafner, Landthaler et al. 2010). Both methods accompanied with next generation sequencing are called CLIP-seq. This approach with Argonaute2 pull-down enables the identification of miRNAs and their endogenous targets in a transcriptome-wide scale and potentially the actual miRNA-binding sites on the 3'UTRs. Surprisingly, many noncanonical miRNA binding sites were identified in this study, which do not carry the perfect seed-matching motif. This observation suggests the

miRNA-targeting mechanism is much more complicated than what we currently understood and/or mRNA bound by miRISC alone does not guarantee that it is regulated by miRNA-mediated regulation.

MiRNAs regulate target mRNA expression through mRNA degradation and translational repression. The mRNA degradation can be done via two pathways: 1) Perfect pairing between miRNA and mRNA induces endonucleolytic cleavage by Argonaute2 in bound miRISC. This is the fastest mechanism to regulate target mRNA level. This pathway occurs commonly in plants but rarely in animal due to the imperfect pairing between animal miRNAs and their target mRNAs. 2) Partial miRNA-target pairing induces deadenylation of target mRNA through recruiting CCR4-NOT by miRISC component, GW182. Loss of the poly-A tail induces dissociation of poly-A binding protein, PABPC and leads to degradation of target mRNA. Two different pathways comprised of translational repression by miRNA: the first pathway is to recruit CCR4-NOT complex to block translational initiation while the other pathway is to block translation after the process has been initiated. The second pathway of translational repression occurs after translational initiation. In this pathway, miRISC complex can induce dissociation of ribosome from target mRNAs or proteolysis of nascent polypeptides. Interestingly, there are also studies showing miRNA binding can stimulate protein expression under certain conditions. Upon cell cycle arrests, TNFα mRNA is targeted by miR-369-3, and FXR1 is recruited by Argonaute in miRISC simultaneously in order to activate the translation of TNFα. Another study also showed FXR1 upregulates the translation along with Argonaute in *Xenopus laevis* oocyte (Mortensen, Serra et al. 2011). The other example is miR-10a binds 5'UTR of ribosomal mRNA and

upregulates the expression of ribosomal protein, and this finding suggests miR-10a can potentially induce protein synthesis globally (Vasudevan, Tong et al. 2007; Orom, Nielsen et al. 2008). Despite of three examples above, it is still very rare to find miRNAs positively regulate the expression of target mRNAs.

Four distinct types of miRNA-mediated regulation were introduced above, and several attempts have been made to understand which miRNA-mediated pathway is predominantly adopted. Selbach *et al.* and Baek *et al.* both utilized SILAC to monitor the global change in protein level caused by miRNA regulation. When comparing the fold change of proteins and mRNAs, they concluded the majority of target mRNAs were regulated via mRNA degradation. Another work utilized ribosomal profiling to measure the overall effect of miRNA regulation on translation efficiency and compare with the simultaneously performed mRNA profiling. They found more than 84% of protein downregulation were accounted for lowered mRNA level. All these results indicate mRNA degradation was the predominant pathway in miRNA-mediated regulation in mammalian system (Guo, Ingolia et al. 2010). Due to the high frequency of imperfect pairing between mammalian miRNAs and target mRNAs, we can further claim that miRISC regulate mRNA level mainly through CCR4-NOT induced deadenylation. Deadenylated mRNAs are then brought to P-body (processing body) by GW182 and being degraded.

In addition to miRISC, the efficiency of miRNA-mediated regulation is also determined by various factors. The first factor is the location of miRNA binding site. MiRNA not only binds to 3'UTR but also coding region or 5'UTR. However, the binding of miRISC to coding region may be disrupted by ribosome. Post-translational

modification of Argonaute is another factor in miRNA targeting efficiency. Prolyl-4-hydroxylation of Argonaute stabilizes the Argonaute protein itself and enhances miRISC activity, while phosphorylation, ubiquitylation and poly-ADP-ribosylation on Argonaute protein decreases its affinity to small RNA, destabilizes the protein and interferes the target accessibility, respectively. RNA-binding protein is another major factor affecting miRNA activity, and the locations of binding sites determine the role of RNA-binding protein in miRNA pathway. In zebrafish, the binding sites of DND1 on several target mRNAs are in the proximity of miR-430 target sites; hence DND1 competes against miRISC and serve as negative modulator of miRNA regulation. ELAVL1 possesses dual functions in the modulation of miRNA efficiency. ELAVL1 has preference to bind to 3'UTR of target mRNA. Over 75% of 3' UTRs with Argonaute-binding also contains ELAVL1 binding sites. Notably, only ELAVL1 binding sites overlapped with miRISC binding sites have adverse effect to miRNA activity; otherwise, ELAVL1 can act as a positive modulator of miRNA pathway.

In addition to RNA binding proteins, recent study showed RNA editing plays a role in modulating miRNA regulation as well. ADAR1 mediates RNA editing by converting adenosine to inosine in the double strand RNA region. The RNA editing alters the secondary structure of miRNA precursor and affects miRNA biogenesis. Besides modifying miRNA precursors, ADAR1 also edits the 3'UTR of the Rho GTPase activating protein 26 (ARHGAP26) and disrupts the binding sites of miR-30-3p and miR-573 causing the upregulation of ARHGAP26 mRNA and protein level. The last factor of miRNA activity modulation is Argonaute-interacting protein. Studies have shown Argonaute binding ability is modulated by its cofactors, including NHL2, LIN41 and

Mei-P26 (Neumuller, Betschinger et al. 2008; Hammell, Lubin et al. 2009; Rybak, Fuchs et al. 2009). Some negative modulators affect the stability of Argonaute or its binding affinity to small RNA and target mRNA; the detailed mechanism of positive modulators is not clear though. Currently, only a dozen genes were identified to affect miRNA activity in different contexts and model organisms. One miRNA activity modulator, Mei-P26, was identified to decrease AGO2 affinity to target mRNAs in *D. melanogaster*; however, there is no homologous gene of Mei-P26 in mammals. This shows our understanding about the miRNA-mediated regulation is still poor, and we need to have a better method to identify the remaining miRNA activity modulators more efficiently.

The goal of this study is to identify genes affect miRNA activity, and we focused on RNA-binding proteins and miRISC-interacting proteins. Due to miRISC induces mRNA degradation to the majority of target mRNAs; we assumed there is observable relationship between miRNA and target mRNA from miRNA and mRNA expression profiles. We can therefore identify miRNA modulators that are able to disrupt or improve the miRNA-target relationship. Figure 5.1 showed the three-way relationship of miRNA-target-modulator we plan to identify in this study.

5.2 Material and Methods

Data source

The list of validated miRNA-target interactions in human was collected from miRecords and miRTarBase (Xiao, Zuo et al. 2009; Hsu, Lin et al. 2011). The collected miRNA-target interactions only validated by low-throughput experiments, such RT-PCR, western blot, ELISA or luciferase assay. Some validated miRNA-target interactions were

further validated by site-directed mutagenesis. On the contrary, high-throughput experiments, such as microarray and SILAC (stable isotope labeling with amino acids in cell culture), identify affected mRNA or protein upon disrupting endogenous miRNA level, but those affected mRNA or protein could be indirectly regulated by miRNA.

Gene and miRNA expression profile was obtained from TCGA Breast invasive carcinoma (BRCA) data portal (Paroo, Ye et al. 2009). RNA-seq from breast cancer patients were performed by Illumina HiSeq, and miRNA-seq was performed by Illumina HiSeq and Genome Analyzer. The libraries undergo QC assessment and then trimmed adaptor. The reads are mapped to reference genome using BWA alignment. Only reads with perfect match to genome are kept for further analysis. The number of reads for each gene and miRNA represented RPKM (reads per kilobase per million reads) and RPM (reads per million reads). The RNA-seq and miRNA-seq are paired using sample ID, and the processed data set consists of 820 samples.

Another gene and miRNA expression profiles were obtained from METABRIC (The Molecular Taxonomy of Breast Cancer International Consortium) (Melo, Ropero et al. 2009; Dvinge, Git et al. 2013). Gene expression was profiled by Illumina HT-12 v3 array, and miRNA expression on customized Agilent array. The gene and miRNA expression profile are paired by sample ID and the processed data comprised of 1302 samples.

The list of RNA binding proteins was obtained from Gene ontology. The protein-protein interactions were downloaded from StringDB and PrePPI (Franceschini, Szklarczyk et al. 2013; Zhang, Petrey et al. 2013). Predicted miRNA target interactions

were obtained from Targetscan (Lewis, Burge et al. 2005; Grimson, Farh et al. 2007; Friedman, Farh et al. 2009; Garcia, Baek et al. 2011) with context score higher than 80, a cutoff suggested by system.

Screening for miRNA activity modulators

The strategy to identify miRNA activity modulators (M) is similar with miRNA biogenesis regulators described in Chapter 4. We calculated the conditional mutual information between miRNA and target given any candidate gene, I(miRNA;target|M). The statistical significance of CMI is evaluated using null distribution of CMI built from shuffling of candidate miRNA modulator's sample identity while maintaining the expression of miRNA and targets constantly. All significant triplets have significant CMI with p-value less than 1e-5, and I(miRNA;target|M) > I(miRNA;target). The computation of mutual information and conditional mutual information was performed as described previously. To determine the significance of all predicted modulators, the p-value of all significant triplets were integrated by candidate modulators using Fisher's method.

Determining the mode of action

For each predicted miRNA modulator, the whole data set was sorted according to the expression of tested modulator. Then, we selected the samples with highest and lowest (for example, top and bottom 30%) expression of tested modulator. The relationship between miRNA and target in two tails of the sample set was calculated using mutual information or spearman correlation. Then, check all predicted miRNA-target pairs to see if they have better relationship when the predicted modulator is highly

or lowly expressed. Compare the ratio to unpredicted miRNA-target pairs using Fisher's exact test to determine the mode of action of predicted modulators.

Cell and culture condition

The breast cancer cell line, HCC1395, was obtained from Columbia Genome Center. It was grown in RPMI-1640 medium supplement with 10% fetal bovine serum (GIBCO/BRL). HCC1395 cells were incubated at 37°C in a 95% air/5% CO2 humidified atmosphere. Fresh RPMI medium was supplied every other day to the cultures after removal of the supernatant.

RNA Extraction and Quantitative Reverse Transcription Polymerize Chain Reaction

The RNA extraction, quantitative PCR with reverse transcription to measure the expression levels were performed as described previously. The primers of AGO2 and EGFR were obtained from Sigma-Aldrich. EGFR primers: forward 5'-CAGCGCTACCTTGTCATTCA-3', reverse 5'- TGCACTCAGAGAGCTCAGGA-3'. AGO2 primers: forward 5'-TCGCACTATCACGTCCTCTG-3', reverse 5'-ATGGCTTCCTTCAGCACTGT-3'. EGFR primers: forward 5'-CAGCGCTACCTTGTCATTCA-3', reverse 5'- TGCACTCAGAGAGCTCAGGA-3'. WNT7A primers: forward 5'-CCCGGGCGGGCTATGTTGATT-3', reverse 5'-GCTTGCGCCCAGAGCTACCA-3'. WIPF2 primers: forward 5'-CAGCCCGAGACCCTCCCAGT-3', reverse 5'-GCCCAGCTGGCGTCCTTGA-3'. PALB2 primers: forward 5'-TCTGTCGCCTGCCCGATGGA-3', reverse 5'-CGCTGAAGGCGGGCTAGTGT-3'.

RNA interference and Reverse transfection

EGFR specific siRNA was obtained from Columbia Genome Center. Reverse transfection was performed as described previously in Chapter 4.

5.3 Results

Collected miRNA-target interactions were validated with mutual information and spearman correlation to ensure the relationship existing in two breast cancer data sets. From various databases, I collected 2374 experimentally validated human miRNA-target interactions from different contexts. All collected interactions were validated by low-throughput biochemical methods. Due to the majority of miRNA-target interaction leads to mRNA degradation in mammalian system, we hypothesize there is measurable relationship between miRNAs and their targets. In total, 672 miRNA-targets interactions are found to have significant mutual information and anti-correlation in TCGA breast cancer (BRCA) data set. MiR-21 and miR-17 are known as oncomiRs for targeting tumor suppressor genes, such as PTEN, PDCD4 and RB1, and it is significantly highly expressed in various types of tumor. Among the 672 selected miR-target pairs, miR-21 and miR-17 comprises of 29 and 28 pairs, respectively, showing that they have been widely studied and they both actively regulate the target mRNAs in breast cancer context.

We use conditional mutual information to identify the triplets of miRNA-target-modulator with significant p-value (conditional mutual information, p-value < 1e-5). Similar with the idea of finding miRNA biogenesis regulator, a miRNA activity modulator improves the relationship (anti-correlation) between the miRNA and its targets. To find out which gene is significant, we use Fisher's method to integrate p-values from all significant triplets by activity modulators. In total, 4732 miRNA-target-modulator triplets have significant conditional mutual information from TCGA BRCA data set. The

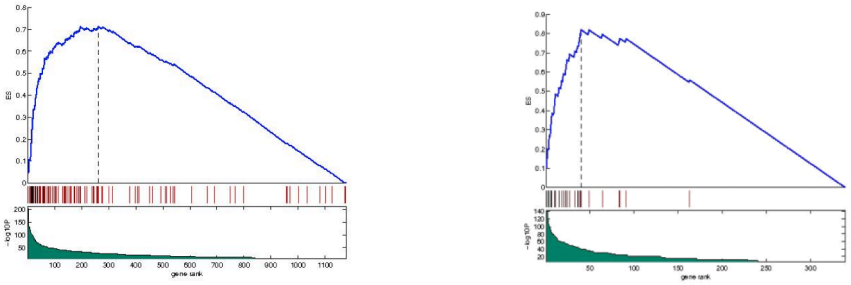
TCGA		
Modulator	No. Pairs	pFisher
EGFR	44	6.153E-212
STAC	38	1.744E-185
QKI	29	8.88E-141
NDN	28	6.629E-133
RNF150	27	1.759E-130
RBMS1	26	7.143E-130
CSDA	27	8.667E-130
PABPC4L	25	3.862E-123
PTRF	24	2.489E-118
IGF2BP2	24	9.235E-118
FYN	23	5.636E-113
ETS1	22	1.362E-108
TNFRSF1B	22	1.293E-105
SERPING1	22	2.794E-104
ESR1	21	3.081E-103
SGOL1	21	2.846E-102
TRIM29	21	3.151E-102
FBP1	21	1.608E-99
CCR4	20	3.0861E-97
WWTR1	20	3.9979E-95

METABRIC		
Modulator	No. Pairs	pFisher
NDN	30	4.11E-143
EGFR	29	8.51E-140
GAS7	23	3.27E-110
FYN	21	4.23E-101
CDCA5	21	5.15E-100
UBE2C	18	1.86E-87
CENPA	17	4.43E-82
SPRY1	17	8.94E-81
MYB	16	3.22E-79
SERPING1	16	6.21E-78
ESR1	15	1.26E-72
TCF4	15	5.73E-72
PDGFRB	14	4.63E-69
CCL5	13	5.14E-64
RARA	13	1.64E-62
NCF4	12	2.92E-60
GZMA	12	3.48E-60
PRKCA	12	1.65E-59
LMO4	12	1.04E-58
TNFRSF1B	12	2.19E-58

Table 5.1 Candidate miRNA activity modulators from TCGA BRCA and METABRIC.

table showed top 20 candidates miRNA modulators. Interestingly, EGFR is the best candidate in this analysis, modulating 44 miRNA-target interactions. We repeated the same analysis using another independent breast cancer data set, METABRIC. We identified 1364 significant triplets, which consist of 339 candidate miRNA modulators.

NDN is the top candidate miRNA modulator in METABRIC analysis while EGFR is the second candidate regulating 31 miRNA-target pairs. We compared the results from METABRIC and TCGA using GSEA (Gene Set Enrichment Analysis) and found out the results from these two data sets are highly similar (GSEA, p-value < 1e-5), suggesting this method is robust to identify consistent output from independent sources. It also showed the candidate modulators identified in both predictions are highly likely to play a role in miRNA-mediated regulation in breast cancer.



METABRIC candidates in TCGA BRCA TCGA BRCA candidates in METABRIC

Figure 5.2 Significant enrichment between the predictions from TCGA BRCA and METABRIC.

The top 100 candidate miRNA activity modulators from METABRIC also have high ranking in TCGA, and vice versa. The enrichment analysis was performed by GSEA, and both analyses showed significant enrichment with p-value less than 1e-5

Most of validated miRNA activity modulators have obvious mode of action, facilitating or interfering, in regulating miRNA targeting given present studies. For example, DND1 acts as a negative miRNA modulator by competing against miRNA binding to target motif. FMRP or PUF were found to bind 3'UTR and stabilize the

neighboring miRISC binding to target mRNA. ELAVL1 is the only case that play dual role in regulating miRNA activity. To determine the mode of action of candidate miRNA modulators, we sorted the samples according to the expression of the testes modulator. The samples with highest and lowest expression of modulators were then selected. In

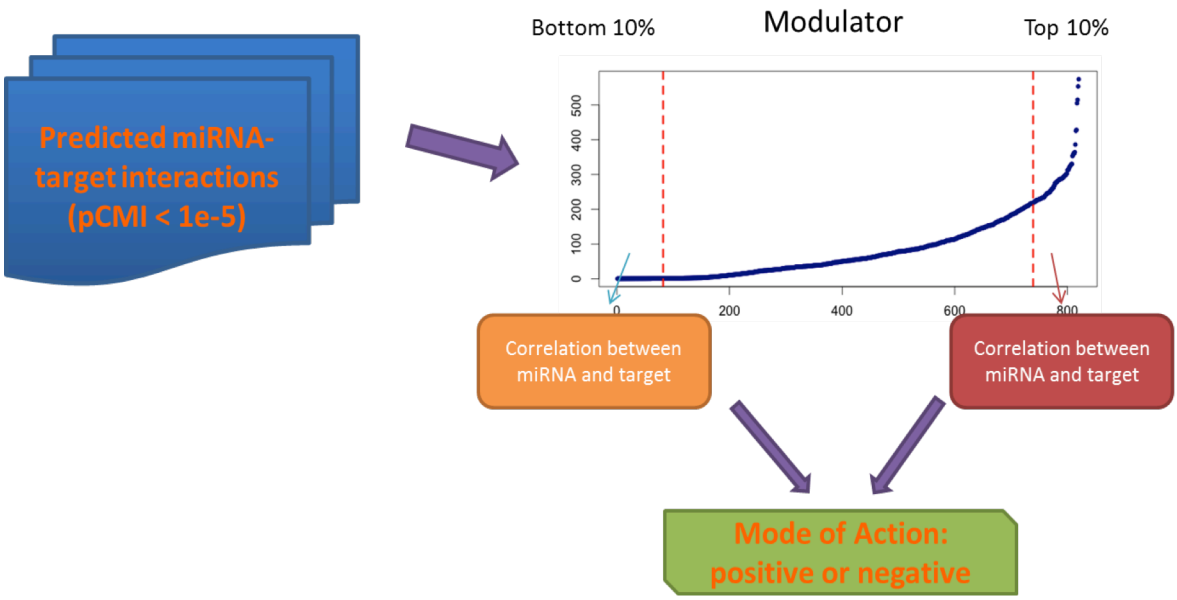


Figure 5.3 Flowchart to determine mode of action of miRNA modulator.

				Predicted		Not Predicted		
	Modulator	pFisher	Type	ΔMI < 0	ΔMI > 0	ΔMI < 0	ΔMI > 0	pFET
TCGA	EGFR	6.15E-212	Positive	7	37	1076	1108	9.72E-06
	FYN	5.64E-113	Positive	1	22	1099	1099	5.30E-06
	TRIM29	3.15E-102	Positive	0	21	987	1217	5.13E-06
METABRIC	EGFR	8.51E-140	Positive	2	27	609	825	6.98E-04
	FYN	4.23E-101	Positive	1	20	555	887	3.50E-03
	TRIM29	1.66E-50	Positive	0	10	520	934	1.74E-02

Table 5.2 EGFR, FYN and TRIM29 are all positive miRNA modulator in TCGA BRCA and METABRIC.

Most of miRNA-target pairs modulated the miRNA modulators have better MI, while the miRNA-target pairs

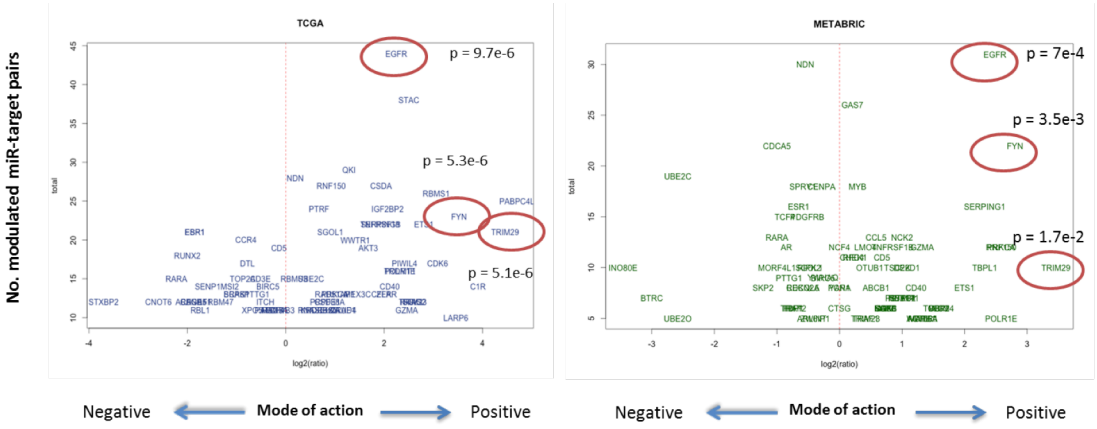
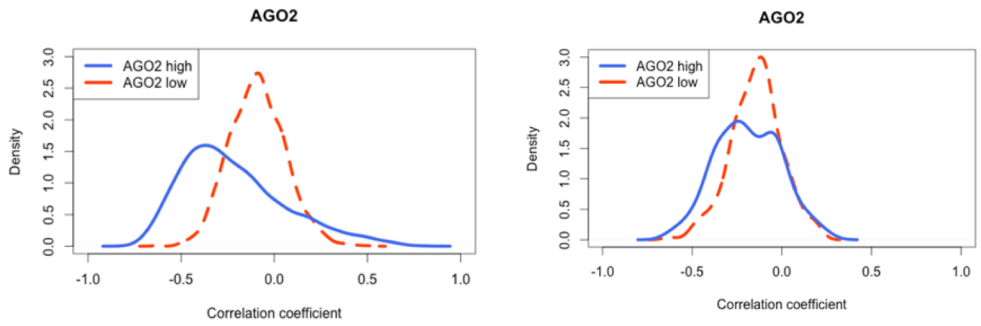


Figure 5.4 Mode of action of candidate miRNA activity.

Thirty and seven candidate miRNA activity modulators from TCGA BRCA and METABRIC have significant mode of action, respectively. Three genes, EGFR, FYN and TRIM29, are significant in both data sets. P-values were estimated using Fisher's exact test.

each sample set, the relationship between miRNA and target was measured using mutual information and spearman correlation. By summarizing the response of all regulated miRNA-target interactions, we can determine the mode of action of miRNA modulator (Figure 5.3). This analysis showed EGFR is a positive miRNA activity modulator: when

EGFR is highly expressed, it improves the relationships of 37 out of 44 modulated miRNA-target pairs, while the bias is not obvious among non-EGFR modulated pairs (Fisher's exact test, p-value = 9.7e-6). PABPC4L is also identified as a positive miRNA modulator (Fisher's exact test, p-value = 1.0076e-7); all (25/25) modulated miRNA-target interactions are better correlated when PABPC4L is highly expressed. PABPC4L is a poly-A tail binding protein, and it may be interacting with another poly-A tail binding protein, PABP1, and involved in recruiting CCR4-NOT complex and induce subsequent mRNA degradation. The only known PABPC4L interacting protein is PAIP1, and PAIP1



TCGA **METABRIC**

Figure 5.5 miRNA-target pairs have better relationship when AGO2 is highly expressed.

In TCGA BRCA and METABRIC, miRNAs and targets are negatively correlated when AGO2 is highly expressed

is highly similar with EIF4G1, a eukaryotic translation initiation factor. This implies PABPC4L alone can interact with the translation initiation complex and modulate miRNA activity.

In METABRIC, there are seven predicted miRNA activity modulators have significant mode of action: UBE2C, EGFR, INO80E, FYN, BTRC, TRIM29 and RNF150 (Fisher's exact test, $p < 5e-2$). Notably, only EGFR, FYN and TRIM29 are also predicted as miRNA modulators using TCGA data set. Moreover, all three modulators increase the miRNA activity in TCGA BRCA and METABRIC (Figure 5.4), suggesting EGFR, FYN and TRIM29 are high confidence candidates. To test if candidate miRNA modulators have widespread effects on more miRNA-target interactions, we took over 60000 predicted miRNA-target interactions from Targetscan with context score higher than 80 and examine the consistence of our prediction. In total, 24126 and 23173 predicted miRNA-target pairs could exist (mutual information, p-value < 1e-2 and spearman correlation < 0) in TCGA BRCA and METABRIC, respectively. First of all, I apply the mode of action analysis on AGO2, the core component in miRNA-mediated regulation. More than 2000 predicted miRNA-target interactions can be modulated by AGO2 in our prediction (conditional mutual information, $p < 1e-5$). The mode of action analysis clearly showed the anti-correlation between miRNA and target when AGO2 is highly expressed, and the anti-correlation between miRNAs and targets disappear when AGO2 is low (Wilcoxon signed-rank test, p-value < 2.2e-16) (Figure 5.5). The result from AGO2 agrees with our hypothesis; the essentiality of AGO2 was clearly depicted in the mode of action analysis. In our prediction, EGFR, FYN and TRIM29 regulate 2981, 672 and 1590 predicted miRNA-target interactions in TCGA BRCA data set, and they modulate 643, 386 and 203 predicted pairs in METABRIC. The mode of action analysis shows they all still act as positive miRNA modulators in the much bigger collection of predicted miRNA-target pairs using two independent breast cancer data sets. The

distribution of spearman correlation between the predicted miRNA-target pairs is significantly shifted when EGFR, FYN and TRIM29 are highly expressed, while EGFR, FYN and TRIM29 are lowly expressed, the distribution of correlation is centered on 0 (Wilcoxon signed-rank test, p-value < 2.2e-16). It suggests they all have wide spread effect on miRNA-mediated post-transcriptional regulation. EGFR is a membrane-associated tyrosine kinase, and it phosphorylates downstream targets to initiate several signaling pathway, including MAPK and AKT pathway. AGO2 is the core protein of miRISC complex. Not only binds to mature miRNAs and protect them from turnover, AGO2 also induces endo-nucleic cleavage when miRNAs perfectly match to their target mRNA. The relationship between EGFR and AGO2 was established by Adams *et al.* in

2009. They showed the expression of AGO2 is upregulated in estrogen receptor (ER) α-negative breast cancer cell lines compared to (ER) α-positive breast cancer cell lines. In the ERα-negative cell line, MDA-MB-231, AGO2 protein level was severely abrogated upon silencing EGFR. We also observed the significant reduction of AGO2 mRNA level while knocking down EGFR in another ERα-negative cell line, HCC1395, using EGFR-specific siRNA. Due to the centrality and essentiality of AGO2 in miRNA-mediated

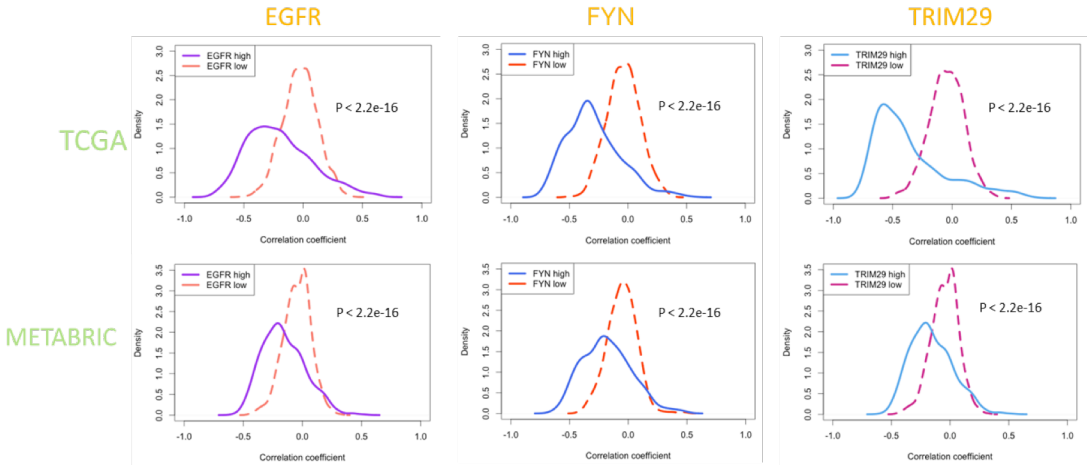


Figure 5.6 EGFR, FYN and TRIM29 affect the relationship of predicted miRNA-target pairs.

EGFR, FYN and TRIM29 are candidate positive miRNA modulators for predicted miRNA-target interactions in TCGA BRCA and METABRIC

regulation, EGFR ranked as the top candidate in our analysis.

Besides breast cancer, we also used gene and miRNA expression profile from TCGA glioblastoma data set to identify miRNA activity modulators. As discussed before, to identify miRNA activity modulators, we restricted our analysis to candidate

modulators of literature-validated miRNA-target interactions. Because prediction of individual miRNA targets is still inaccurate, we could not rely on the robust statistics afforded by large miRNA programs, as used when predicting sponge modulators. Even

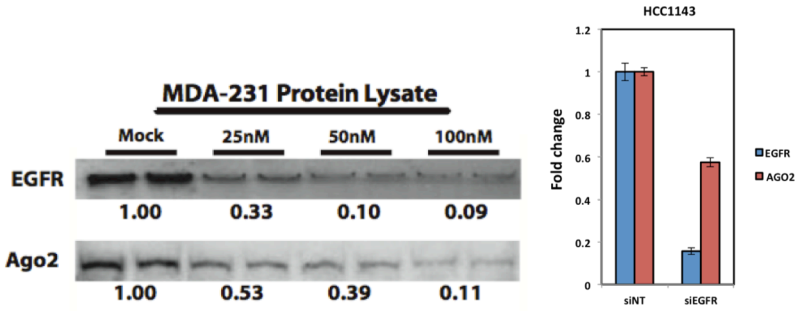


Figure 5.7 EGFR level affects AGO2 in breast cancer basal cell lines.

Silencing EGFR in MDA-MB-231 and HCC1143 both repress the AGO2 level.

with this substantial limitation, Hermes could identify 148 miRNA activity modulators, suggesting that this number may increase substantially once a more comprehensive, high-accuracy miRNA target network is available. To experimentally confirm nonsponge modulator candidates inferred by our analysis, we selected three interactions, two affecting PTEN and one affecting RUNX1. These include WIPF2, as a miR-mediated regulator of RUNX1, and PALB2 and WNT7A, as miRNA-mediated regulators of PTEN (Figure 5.8 A). Both PTEN and RUNX1 are known drivers of gliomagenesis (Carro et al., 2010; Verhaak et al., 2010), and genes that regulate their expression may play a role in this disease.

Upregulation of WNT7A by transfection of its cDNA, which lacks both 5' and 3'UTRs, led to 1.5-fold upregulation of PTEN or PTEN 3' UTR luciferase activity

(Figures 5.8 B and 5.8 E), suggesting that, as predicted, WNT7A regulation of PTEN is miRNA dependent, but not sponge mediated. Silencing PALB2 led to a 1.5-fold upregulation of PTEN or PTEN 3'UTR luciferase activity (Figures 5.8 C and 5.8 E), and silencing WIPF2 led to a 3-fold upregulation of RUNX1 expression and 2-fold upregulation of RUNX1 3'UTR luciferase activity (Figures 5.8 D and 5.8 F), both consistent with computational predictions of miRNA-mediated downregulation. Consistent with miRNA-dependent regulation, DROSHA and DICER silencing (Figure 5.8 G) abrogated PTEN and RUNX1 regulation by WNT7a/PALB2 and WIPF2, respectively (Figures 5.7 E and 5.7 F). Expression of validated miRNAs targeting these genes, such as miR-21 (WNT7A), miR-106a (PALB2), and miR-17-5p (WIPF2), were relatively unchanged (Figure 5.9).

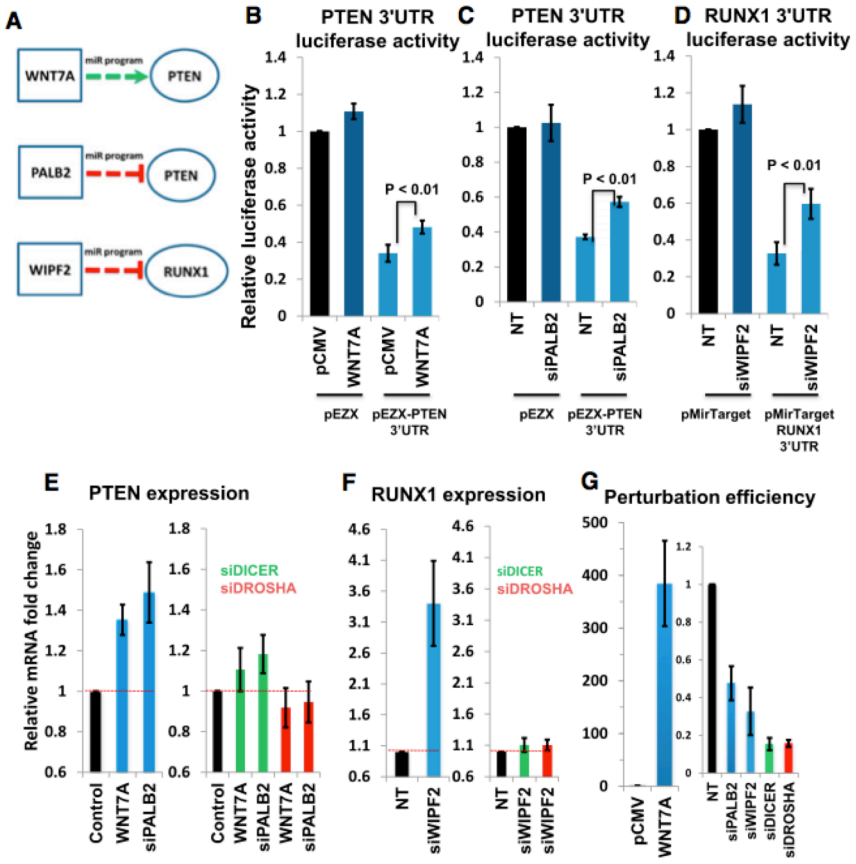


Figure 5.8 validation of miRNA modulator and PTEN in SNB19.

(A) Validated miRNA modulators include WNT7A and PALB2 (predicted to induce miR- dependent upregulation and downregulation of PTEN, respectively) and WIPF2 (predicted to induce miR-dependent downregulation of RUNX1).

(B) PTEN 3'UTR luciferase activity and activity of the empty luciferase vector were measured at 24 hr following ectopic expression of pCMV-WNT7A or empty vector pCMV.

(C) PTEN 3'UTR luciferase activity and activity of the empty luciferase vector were measured at 24 hr following siRNA-mediated silencing of PALB2 versus nontarget siRNA.

(D) RUNX1 3'UTR luciferase activity) and activity of the empty luciferase vector at 24 hr following siRNA- mediated silencing of WIPF2 and nontarget (NT5) siRNA.

(E) qRT-PCR analysis of PTEN gene expression fold change at 24 hr following ectopic expression of WNT7A and siRNA mediated silencing of PALB2 without (left) and with (right) siRNA-mediated silencing of DROSHA and DICER.

(F) qRT-PCR analysis of RUNX1 gene expression fold change at 24 hr following siRNA-mediated silencing of WIPF2 without (left) and with (right) siRNA-mediated silencing of DROSHA and DICER.

(G) Efficiency of WNT7A ectopic expression and of siRNA-mediated silencing of PALB2, WIPF2, DICER, and DROSHA, measured by qRT-PCR analysis.

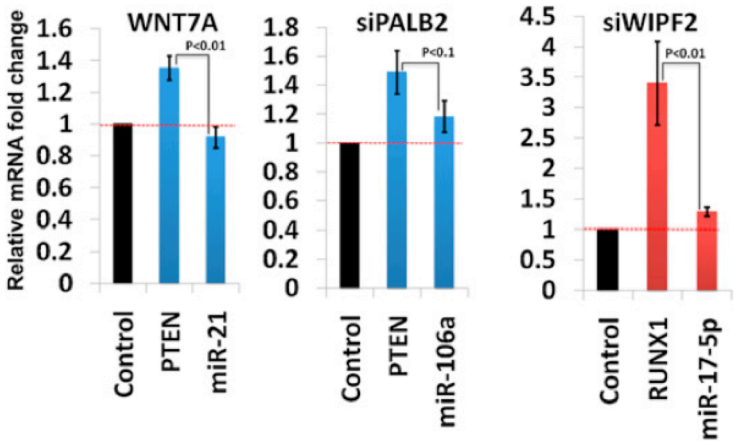


Figure 5.9. Nonsponge modulators have little effect on the expression of regulating miRNAs.

The qPCR expression of validated targeting miRs found to be significantly modulated are unchanged relative to their targets by pCMV-WNT7A transfection or PALB2 and WIPF2 silencing. Data are represented as mean ± SEM.

5.4 Discussion

MiRNA-mediated regulation represses target gene expression by mRNA degradation and translational repression. In addition to core miRISC component, such as AGO2 or GW182, dozens of other proteins also contribute to the regulation of miRNA activity. These proteins modulate miRNA efficiency by interacting with miRISC components or binding to target mRNA in order to compete against or facilitate miRISC. Since there are many proteins bind to mRNA at 3'UTR or interact with miRISC components, we hypothesize that there are many miRNA modulators to be identified in different contexts. This is the first attempt to computationally identify miRNA

modulators. Notably, we obtained very similar predictions from two independent data sets, implicating the candidate genes are conserved and active in breast cancer context.

In our analysis, EGFR, FYN and TRIM29 all predicted as positive miRNA modulator using TCGA BRCA and METABRIC data. EGFR, a membrane-associated tyrosine kinase, phosphorylate downstream targets and stimulates various cell signaling pathways that controls cell proliferation, apoptosis and survival. Amplifications or somatic mutations of EGFR are often identified in multiple tumors, such as glioblastoma and non-small cell lung cancer, and dysregulation of EGFR is strongly implicated in the biology of human epithelial malignancies. In this analysis, we suggest there could be a novel role of EGFR to influence the downstream targets using different mechanism, miRNA-mediated regulation. We found the upregulation of EGFR associates with better relationship (negative correlation) between miRNAs and their target mRNAs. The phenomenon not only presents in validated miRNA-target interactions; it applies to whole miRNAome as well. Over 80% of predicted miRNA-target interactions modulated by EGFR are better correlated (negatively) under high EGFR context. We found the expression of the core component of miRISC, AGO2, is regulated by EGFR. Therefore, it may explain EGFR have a widespread influence on miRNA regulation. Besides regulating AGO2 expression, EGFR also interact and phosphorylate AGO2 at Try393 in response to hypoxia. Phosphorylation at Tyr393 on AGO2 interferes the interaction between DICER1 and AGO2; therefore, it interferes miRNA biogenesis by preventing several miRNA-miRNA* duplexes being transferred from DICER1 to AGO2. The miRNAs with long hairpin loop are particularly affected by Tyr393 phosphorylation, such as miR-31, miR-192 and miR-193a-5p. The p-Y393-AGO2 is significantly

upregulated during tumor progression and enriched in hypoxic tumor area. More importantly, breast cancer patients with higher expression of p-Y393-AGO2 have significantly poor survival rate.

Tyrosine kinases can be classified into two categories: receptor and nonreceptor tyrosine kinase. Receptor kinase, such as EGFR and VEGFR (vascular endothelial growth factor receptor), can be stimulated by extracellular ligands and activate downstream targets. FYN is a membrane-associated nonreceptor tyrosine kinase and belongs to Src family kinases. FYN can serve as mediator of growth factor-induced AKT activity: knocking down FYN together with Src abolished the activation of AKT by EGF. FYN is found to be upregulated in various types of cancer, including melanoma, glioblastoma multiforme and squamous cell carcinoma of the head and neck. More intriguingly, overexpression the dominant negative FYN reduced primary tumor weights in a mouse squamous cancer model. Similar with EGFR, we proposed that FYN also plays an important role in miRNA-mediated regulation by improving miRNA efficiency. Over 80% of FYN-modulated miRNA-target interactions have better relationship when FYN is highly expressed, which implicated the core component of miRISC could be downstream substrate of FYN.

At present, over a dozen of proteins are known to be able to modulate miRNA activity. Several validated miRNA modulators belong to TRIM-NHL family, possessing E3 ubiquitin ligase activity, including Mei-P26, TRIM32, NHL-2 and TRIM71. They can act as positive or negative modulators of miRNA. Mei-P26 in *Drosophila melanogaster* is a negative miRNA modulator, and its homolog in *C. elegans* is NHL-2. Mei-P26 mutant showed aberrant cell proliferation and differentiation in female germline of *D.*

melanogaster, and the phenotype was partially caused by hyperactivity of miRNA regulation. Mei-P26 interacts with AGO1, which is the Argonaute protein mainly responsible for miRNA activity in fly, but the detailed mechanism of Mei-P26 remains to be clarified. Interestingly, while Mei-P26 interferes miRNA pathway, its homolog in *C. elegans*, NHL-2, promotes miRNA efficiency. NHL-2 interacts with AGO and GW182 physically and genetically, and it also interacts with another P-body component, CGH-1. In *C. elegans*, loss of nhl-2 weakens the repression of let-60 and cog-1 while the levels of let-7 and lsy-6 remain unaffected. Trim71, also known as Lin41, is an evolutionary conserved E3 ubiquitin ligase. It is found to induce ubiquitylation and proteasome-mediated degradation of Ago2 in mice. Overexpression and knocking down of Trim71 leads to decrease and upregulation in the level of Ago2, respectively, indicating Trim71 represses endogenous Ago2 level. Another validated miRNA modulator, Trim32, was also found to interact with AGO1 in mice. The interaction enhances the miRNP activity and increases the miRNA efficiency. TRIM32 carries NHL domain, RING domain (E3 ubiquitin ligase) and BBOX domain; however, it only ubiquitinates the transcription factor, c-myc, via RING domain to induce protein degradation. TRIM32 interacts with AGO1 via NHL domain, like Mei-P26, but TRIM32 and Mei-P26 have opposite effects upon interacting with AGO1, suggesting other domains are involved in determining the mode of action of AGO1 interacting protein. We predicted TRIM29 as a positive regulator; almost 90% of TRIM29-modulated predicted miRNA-target interactions have better relationship when TRIM29 is highly expressed. Currently, our understanding about TRIM29 is still very limited. TRIM29 is evolutionary conserved in vertebrates and carries a BBOX domain that can be used as a mediator of protein-protein interaction.

Here we proposed TRIM29 may also interacts with miRISC core components and regulate miRNAome activity due to the significant bias of mode of action of TRIM29.

Chapter 6

Conclusions

This thesis demonstrates the first attempt to genome-wide infer miRNA biogenesis regulators and activity modulators. The first two chapters described the motivation of this analysis and reviewed the current knowledge regarding to the regulation of miRNA biogenesis and miRNA efficiency. Chapter 3 described how we utilized high-throughput data (ENCODE and FANTOM) to obtain highly confident intragenic miRNAs that are co-transcribed with host mRNAs. Furthermore, we showed that despite miRNA and mRNA are in the same transcription unit, the relationship between many host-miRNA pairs are still very subtle, implicating many intragenic miRNAs' level are under extensive regulation. Chapter 4 demonstrated the effort to computationally and experimentally screen for miRNA biogenesis regulators. The algorithm, MIRAGE, identified several validated miRNA biogenesis regulators, including LIN28, ILF3 and ERN1. We validated several candidates throughout the prediction and showed their ability to affect miRNA processing at the post-transcriptional stage. The pri-miRNA binding activity of SMAD6 was also biochemically demonstrated in this chapter. Additionally, I use PAR-CLIP data from public domain to show IGF2BP3 interacts with miRNA in HEK293 cell. From the IGF2BP3-interacting miRNAs, we also identify the regulation of IGF2BP3 on PTEN through miRNA program. Chapter 5 showed the result of screening for miRNA activity modulators. We developed an algorithm to infer miRNA activity modulators using the expression profile of miRNA and targets. The results showed EGFR, FYN and TRIM29 are potential positive miRNA

modulators in breast cancer context. Over 80% of predicted miRNA-target pairs are better correlated negatively when these three genes are highly expressed. The detailed biochemical mechanism still needs to be studied. We also demonstrated that, in glioblastoma, PTEN level is not only modulated by miRNAs but also by other miRNA modulators, such as PALB2, WNT7A and WIPF2.

In summary, we carried out a series of analysis in order to understand the post-transcriptional regulation of miRNA biogenesis and miRNA function. MiRNAs are known to be involved in many biological pathways, and 30% to 60% of transcriptome are under miRNA modulation. Therefore, this analysis could lead us to a clearer view of miRNA profile. Most of miRNA regulators are able to modulate a subset of miRNAs or miRNA-target interactions, which implicate the dysregulation of components in miRNA pathway may lead to severe defects, such as developmental abnormality, embryonic lethality and tumorigenesis (Kumar, Lu et al. 2007; Stefani and Slack 2008; Yi, Pasolli et al. 2009).

Lu *et al.* first profiled miRNA expression from various tumor types and the adjacent normal tissues and showed that miRNAs are generally less abundant in tumors compared to their normal tissue counterparts (Lu, Getz et al. 2005). This finding leads to the proposal that there are defects in miRNA biogenesis pathway, which may lead to tumorigenesis. Subsequent studies showed around 27% of various tumors are found to have a hemizygous deletion in DICER1 (Kumar, Pester et al. 2009). Additionally, targeting DICER1, DROSHA and its cofactor DGCR8 knocks down global miRNA expression and increases the oncogenic potential of cancer cell lines and accelerates the tumor formation (Kumar, Lu et al. 2007).

Aberrant expression of DICER1 leads to the impaired miRNA processing has also been shown to increase the rate of tumor formation in two different cancer mouse models, a KRAS-driven lung cancer and an Rb-driven retinoblastoma (Kumar, Pester et al. 2009; Lambertz, Nittner et al. 2010). Therefore, DICER1 is considered a haploinsufficient tumor suppressor, requiring partial deletion for its associated tumorigenesis phenotype. The phosphorylation of the DICER1 cofactor TARBP2 by the mitogen-activated protein kinase ERK enhances pre-miRNA processing of oncogenic miRNAs, such as miR-21, and decreases production of tumor suppressor let-7a (Paroo, Ye et al. 2009). Moreover, TARBP2 is mutated in some colon and gastric cancers with microsatellite instability, and TARBP2 frameshift mutations correlate with DICER1 destabilization; in cell lines and xenografts with TARBP2 mutations, reintroduction of wild type TARBP2/DICER1 slowed tumor growth (Melo, Ropero et al. 2009; Garre, Perez-Segura et al. 2010). Finally, DICER1 was also recently implicated as a metastasis suppressor (Garre, Perez-Segura et al. 2010). Another core miRNA regulator, DROSHA, was found to be a predictor of breast cancer patient prognosis. A gradual loss of cytoplasmic DROSHA was found to gradually lose during tumor progression from ductal carcinoma in situ to invasive or even metastatic tumor. Loss of cytoplasmic DROSHA is associated with BRCA1 and ESR1 expression in breast cancer (Khoshnaw, Rakha et al. 2013). Note that miRNA precursors are processed by DROSHA in the nucleus and then the pre-miRNAs are transported to cytoplasm by XPO5; therefore, the connection between miRNA and cytoplasmic DROSHA needs to be further examined.

Besides the core components in the miRNA biogenesis pathway, post-transcriptional regulators also associate with tumor initiation and progression. LIN28A is

the most well studied miRNA biogenesis regulator. LIN28A blocks processing of tumor suppressor pri-let-7 in nucleus and pre-let-7 in cytoplasm by recognizing a motif on the terminal loop, thus maintaining expression of genes that drive self-renewal and proliferation; tumors that express LIN28A were indeed shown to be poorly differentiated and more aggressive than LIN28A-negative tumors (Newman, Thomson et al. 2008; Piskounova, Viswanathan et al. 2008; Rybak, Fuchs et al. 2008; Viswanathan, Daley et al. 2008; Viswanathan, Powers et al. 2009). Another example is heterodimer DDX5/DDX17. The helicases DDX5 and DDX17 both carry DEAD box can stimulate the processing of one third of all murine miRNAs by interacting with DROSHA and miRNA precursors or by acting as a scaffold protein to recruit other factors, such as TP53, SMAD5, to the DROSHA complex. Dimerization of DDX17 and DDX5 RNA helicases through interactions is mediated by the tumor suppressor TP53. Interactions of DDX5/DD17 and TP53 to the DROSHA-DGCR8 complex facilitate the conversion of some pri-miRNAs to pre-miRNAs. Specifically, the DDX5-mediated interaction of the DROSHA complex with the tumor suppressor TP53 was shown to have a stimulatory effect on the tumor suppressor pri-miR-16-1, pri-miR-143 and pri-miR-145 processing in response to DNA damage in cancer cells (Suzuki, Yamagata et al. 2009). Thus, mutations and deletion of TP53, which are often observed in multiple types of malignancies, might lead to a widespread decrease in pre-miRNA production. Oncogenic R-SMADs (SMAD1, SMAD3 and SMAD5), downstream effectors of the TGF-β/BMP2 pathway, have been shown to modulate DROSHA-mediated miRNA maturation through interaction with DDX5 and target pri-miRNAs. The binding of SMADs to miRNA precursors promotes the expression of a subset of miRNAs bearing the R-SMAD binding motif, including

oncogenic miR-21, once TGF-β pathway is stimulated. KSRP promotes the biogenesis of a subset of miRNAs, including let-7a, by serving as a component of both DICER1 and DROSHA complexes affecting proliferation, apoptosis and differentiation (Davis, Hilyard et al. 2008; Trabucchi, Briata et al. 2009). In a final example, inactivating mutations of XPO5 in tumors with microsatellite instability result in the nuclear retention of miRNAs (Melo, Moutinho et al. 2010). Restoration of XPO5 function reverses the impaired export of pre-miRNAs and has tumor suppressor features.

In addition to the miRNA regulators mentioned above, we identified IGF2BP2, IGF2BP3 and SMAD6 as novel miRNA biogenesis regulators. Their expression level associates with patient prognosis. High level of IGF2BP2 and IGF2BP3 associate with poor-prognosis of patients with glioblastoma. We also proved IGF2BP3 suppress the tumor suppressor gene, PTEN, via miRNA pathway, and knocking down IGF2BP3 slows down the proliferation of SNB19. More importantly, the clinical data from REMBRANDT database showed the level of IGF2BP3 is associated with the prognosis glioblastoma patients. Patients with high IGF2BP3 level (>2X) has worse prognosis than patients with normal or low IGF2BP3 level.

EGFR was identified as a candidate miRNA activity modulator in breast cancer. From computational screening using validated and predicted miRNA-target interactions, I showed EGFR increases miRNA activity. Preliminary experimental validation demonstrated AGO2 is positively regulated by EGFR, and this observation could explain why EGFR possesses a strong and broad effect on miRNA-mediated regulation. EGFR is highly expressed in basal subtype, but not in luminal A and B subtype of breast cancer. Notably, patients in basal subtype have worse prognosis compared to patients in luminal

A and B subtype. It is important to know if EGFR contributes to the bad prognosis of breast cancer patients in basal subtype through miRNA-mediated regulation.

Regulatory network usually contains only interactions between transcription factor and targets, protein and protein, signaling proteins and substrates. The knowledge we obtained from this study provides a useful resource for miRNA biology, systems biology and oncology. This resource can be utilized for building a complete regulatory network and searching for potential therapeutic targets. This method can also be applied to different contexts in order to study tissue-specific miRNA regulatory mechanisms.

Reference:

Baek, D., J. Villen, et al. (2008). "The impact of microRNAs on protein output." Nature **455**(7209): 64-71.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-297.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**: 215-233.

Berezikov, E., W. J. Chung, et al. (2007). "Mammalian mirtron genes." Molecular cell **28**(2): 328-336.

Bhattacharyya, S. N., R. Habermacher, et al. (2006). "Relief of microRNA-mediated translational repression in human cells subjected to stress." Cell **125**(6): 1111-1124.

Carro, M. S., W. K. Lim, et al. (2010). "The transcriptional network for mesenchymal transformation of brain tumours." Nature **463**: 318-325.

Chatterjee, S. and H. Grosshans (2009). "Active turnover modulates mature microRNA activity in Caenorhabditis elegans." Nature **461**: 546-549.

Cheloufi, S., C. O. Dos Santos, et al. (2010). "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis." Nature **465**: 584-589.

Chen, L., H. Li, et al. (2011). "Expression and function of miR-27b in human glioma." Oncol Rep **26**(6): 1617-1621.

Chen, Y., W. Liu, et al. (2008). "MicroRNA-21 down-regulates the expression of tumor suppressor PDCD4 in human glioblastoma cell T98G." Cancer letters **272**(2): 197-205.

Chi, S. W., J. B. Zang, et al. (2009). "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps." Nature **460**: 479-486.

Cifuentes, D., H. Xue, et al. (2010). "A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity." Science **328**(5986): 1694-1698.

Consortium, E. P., B. E. Bernstein, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Davis, B. N., A. C. Hilyard, et al. (2008). "SMAD proteins control DROSHA-mediated microRNA maturation." Nature **454**: 56-61.

Davis, B. N., A. C. Hilyard, et al. (2010). "Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha." Molecular cell **39**: 373-384.

Desgrosellier, J. S., N. A. Mundell, et al. (2005). "Activin receptor-like kinase 2 and Smad6 regulate epithelial-mesenchymal transformation during cardiac valve formation." Dev Biol **280**(1): 201-210.

Dews, M., J. L. Fox, et al. (2010). "The myc-miR-17~92 axis blunts TGF{beta} signaling and production of multiple TGF{beta}-dependent antiangiogenic factors." Cancer Res **70**(20): 8233-8246.

Dvinge, H., A. Git, et al. (2013). "The shaping and functional consequences of the microRNA landscape in breast cancer." Nature **497**(7449): 378-382.

Edbauer, D., J. R. Neilson, et al. (2010). "Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132." Neuron **65**(3): 373-384.

Ernst, A., B. Campos, et al. (2010). "De-repression of CTGF via the miR-17-92 cluster upon differentiation of human glioblastoma spheroid cultures." Oncogene **29**(23): 3411-3422.

Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nature biotechnology **28**(8): 817-825.

Ernst, J., P. Kheradpour, et al. (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.

Fisher, R. A. (1925). Statistical methods for research workers. Edinburgh, London,, Oliver and Boyd.

Flynt, A. S., J. C. Greimann, et al. (2010). "MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila." Molecular cell **38**: 900-907.

Franceschini, A., D. Szklarczyk, et al. (2013). "STRING v9.1: protein-protein interaction networks, with increased coverage and integration." Nucleic Acids Research **41**(Database issue): D808-815.

Franceschini, A., D. Szklarczyk, et al. (2013). "STRING v9.1: protein-protein interaction networks, with increased coverage and integration." Nucleic Acids Res **41**(Database issue): D808-815.

Friedman, R. C., K. K. Farh, et al. (2009). "Most mammalian mRNAs are conserved targets of microRNAs." Genome Res **19**(1): 92-105.

Fukuda, T., K. Yamagata, et al. (2007). "DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs." Nature cell biology **9**: 604-611.

Fukuda, T., K. Yamagata, et al. (2007). "DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs." Nat Cell Biol **9**(5): 604-611.

Galgano, A., M. Forrer, et al. (2008). "Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system." PloS one **3**(9): e3164.

Garcia, D. M., D. Baek, et al. (2011). "Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs." Nature structural & molecular biology.

Garre, P., P. Perez-Segura, et al. (2010). "Reassessing the TARBP2 mutation rate in hereditary nonpolyposis colorectal cancer." Nat Genet **42**(10): 817-818; author reply 818.

Gennarino, V. A., M. Sardiello, et al. (2009). "MicroRNA target prediction by expression analysis of host genes." Genome research **19**: 481-490.

Glesne, D. and E. Huberman (2006). "Smad6 is a protein kinase X phosphorylation substrate and is required for HL-60 cell differentiation." Oncogene **25**(29): 4086-4098.

Gregory, R. I., K.-P. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." Nature **432**: 235-240.

Grimson, A., K. K. Farh, et al. (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing." Mol Cell **27**(1): 91-105.

Guo, H., N. T. Ingolia, et al. (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels." Nature **466**: 835-840.

Hafner, M., M. Landthaler, et al. (2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP." Cell **141**: 129-141.

Hagan, J. P., E. Piskounova, et al. (2009). "Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells." Nature structural & molecular biology **16**(10): 1021-1025.

Hammell, C. M., I. Lubin, et al. (2009). "nhl-2 Modulates microRNA activity in Caenorhabditis elegans." Cell **136**(5): 926-938.

Han, J., Y. Lee, et al. (2006). "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex." Cell **125**(5): 887-901.

Heale, B. S., L. P. Keegan, et al. (2009). "Editing independent effects of ADARs on the miRNA/siRNA pathways." The EMBO journal **28**(20): 3145-3156.

Helwak, A., G. Kudla, et al. (2013). "Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding." Cell **153**(3): 654-665.

Heo, I., C. Joo, et al. (2008). "Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA." Molecular cell **32**: 276-284.

Heo, I., C. Joo, et al. (2009). "TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation." Cell **138**(4): 696-708.

Hsu, S. D., F. M. Lin, et al. (2011). "miRTarBase: a database curates experimentally validated microRNA-target interactions." Nucleic Acids Res **39**(Database issue): D163-169.

Huse, J. T., C. Brennan, et al. (2009). "The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo." Genes & development **23**: 1327-1337.

Jones, M. R., L. J. Quinton, et al. (2009). "Zcchc11-dependent uridylation of microRNA directs cytokine expression." Nature cell biology **11**(9): 1157-1163.

Katoh, T., Y. Sakaguchi, et al. (2009). "Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2." Genes & development **23**: 433-438.

Kawahara, Y., M. Megraw, et al. (2008). "Frequency and fate of microRNA editing in human brain." Nucleic acids research **36**(16): 5270-5280.

Kawahara, Y., B. Zinshteyn, et al. (2007). "RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex." EMBO reports **8**(8): 763-769.

Kawahara, Y., B. Zinshteyn, et al. (2007). "Redirection of silencing targets by adenosine-to-inosine editing of miRNAs." Science **315**(5815): 1137-1140.

Kawai, S. and A. Amano (2012). "BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex." J Cell Biol **197**(2): 201-208.

Kawaji, H., J. Severin, et al. (2011). "Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation." Nucleic Acids Res **39**(Database issue): D856-860.

Kedde, M., M. J. Strasser, et al. (2007). "RNA-binding protein Dnd1 inhibits microRNA access to target mRNA." Cell **131**(7): 1273-1286.

Kefas, B., J. Godlewski, et al. (2008). "microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma." Cancer research **68**(10): 3566-3572.

Khoshnaw, S. M., E. A. Rakha, et al. (2013). "The microRNA maturation regulator Drosha is an independent predictor of outcome in breast cancer patients." Breast Cancer Res Treat **137**(1): 139-153.

Kim, T.-M., W. Huang, et al. (2011). "A developmental taxonomy of glioblastoma defined and maintained by MicroRNAs." Cancer research **71**: 3387-3399.

Kim, T. M., W. Huang, et al. (2011). "A developmental taxonomy of glioblastoma defined and maintained by MicroRNAs." Cancer Res **71**(9): 3387-3399.

Kim, V. N., J. Han, et al. (2009). "Biogenesis of small RNAs in animals." Nature reviews. Molecular cell biology **10**: 126-139.

Kim, Y.-K. and V. N. Kim (2007). "Processing of intronic microRNAs." The EMBO journal **26**: 775-783.

Kodzius, R., M. Kojima, et al. (2006). "CAGE: cap analysis of gene expression." Nat Methods **3**(3): 211-222.

Kozomara, A. and S. Griffiths-Jones (2011). "miRBase: integrating microRNA annotation and deep-sequencing data." Nucleic acids research **39**(Database issue): D152-157.

Krol, J., I. Loedige, et al. (2010). "The widespread regulation of microRNA biogenesis, function and decay." Nature reviews. Genetics **11**: 597-610.

Kumar, M. S., J. Lu, et al. (2007). "Impaired microRNA processing enhances cellular transformation and tumorigenesis." Nat Genet **39**(5): 673-677.

Kumar, M. S., R. E. Pester, et al. (2009). "Dicer1 functions as a haploinsufficient tumor suppressor." Genes Dev **23**(23): 2700-2704.

Lages, E., A. Guttin, et al. (2011). "MicroRNA and target protein patterns reveal physiopathological features of glioma subtypes." PLoS One **6**(5): e20600.

Lagos-Quintana, M., R. Rauhut, et al. (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-858.

Lambertz, I., D. Nittner, et al. (2010). "Monoallelic but not biallelic loss of Dicer1 promotes tumorigenesis in vivo." Cell Death Differ **17**(4): 633-641.

Lebedeva, S., M. Jens, et al. (2011). "Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR." Molecular cell **43**(3): 340-352.

Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-854.

Lehrbach, N. J., J. Armisen, et al. (2009). "LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans." Nature structural & molecular biology **16**(10): 1016-1020.

Lewis, B. P., C. B. Burge, et al. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**(1): 15-20.

Li, Y., F. Guessous, et al. (2009). "MicroRNA-34a inhibits glioblastoma growth by targeting multiple oncogenes." Cancer Res **69**(19): 7569-7576.

Lu, J., G. Getz, et al. (2005). "MicroRNA expression profiles classify human cancers." Nature **435**(7043): 834-838.

Lu, J., G. Getz, et al. (2005). "MicroRNA expression profiles classify human cancers." Nature **435**: 834-838.

Lu, X., P. Zhao, et al. (2009). "Analysis of miR-221 and p27 expression in human gliomas." Molecular medicine reports **2**(4): 651-656.

Malzkorn, B., M. Wolter, et al. (2010). "Identification and functional characterization of microRNAs involved in the malignant progression of gliomas." Brain Pathol **20**(3): 539-550.

Margolin, A. a., I. Nemenman, et al. (2006). "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context." BMC bioinformatics **7 Suppl 1**: S7.

Margolin, A. A., K. Wang, et al. (2006). "Reverse engineering cellular networks." Nature protocols **1**(2): 662-671.

Medina, R., S. K. Zaidi, et al. (2008). "MicroRNAs 221 and 222 bypass quiescence and compromise cell survival." Cancer research **68**(8): 2773-2780.

Melo, S. A., C. Moutinho, et al. (2010). "A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells." Cancer Cell **18**(4): 303-315.

Melo, S. a., S. Ropero, et al. (2009). "A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function." Nature genetics **41**: 365-370.

Melo, S. A., S. Ropero, et al. (2009). "A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function." Nat Genet **41**(3): 365-370.

Michlewski, G., S. Guil, et al. (2008). "Posttranscriptional regulation of miRNAs harboring conserved terminal loops." Molecular cell **32**: 383-393.

Mishima, Y., A. J. Giraldez, et al. (2006). "Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430." Current biology : CB **16**(21): 2135-2142.

Moller, H. G., A. P. Rasmussen, et al. (2013). "A systematic review of microRNA in glioblastoma multiforme: micro-modulators in the mesenchymal mode of migration and invasion." Mol Neurobiol **47**(1): 131-144.

Mortensen, R. D., M. Serra, et al. (2011). "Posttranscriptional activation of gene expression in Xenopus laevis oocytes by microRNA-protein complexes (microRNPs)." Proc Natl Acad Sci U S A **108**(20): 8281-8286.

Mukherjee, N., D. L. Corcoran, et al. (2011). "Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability." Molecular cell **43**(3): 327-339.

Mullokandov, G., A. Baccarini, et al. (2012). "High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries." Nat Methods **9**(8): 840-846.

Neumuller, R. A., J. Betschinger, et al. (2008). "Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage." Nature **454**(7201): 241-245.

Newman, M. A., J. M. Thomson, et al. (2008). "Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing." RNA **14**(8): 1539-1549.

Newman, M. A., J. M. Thomson, et al. (2008). "Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing." Spring: 1539-1549.

Nolde, M. J., N. Saka, et al. (2007). "The Caenorhabditis elegans pumilio homolog, puf-9, is required for the 3'UTR-mediated repression of the let-7 microRNA target gene, hbl-1." Developmental biology **305**(2): 551-563.

Noushmehr, H., D. J. Weisenberger, et al. (2010). "Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma." Cancer cell **17**: 510-522.

Obernosterer, G., P. J. F. Leuschner, et al. (2006). "Post-transcriptional regulation of microRNA expression." RNA (New York, N.Y.) **12**: 1161-1167.

Okamura, K., J. W. Hagen, et al. (2007). "The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila." Cell **130**(1): 89-100.

Okamura, K., M. D. Phillips, et al. (2008). "The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution." Nature structural & molecular biology **15**: 354-363.

Olivey, H. E., N. A. Mundell, et al. (2006). "Transforming growth factor-beta stimulates epithelial-mesenchymal transformation in the proepicardium." Dev Dyn **235**(1): 50-59.

Orom, U. A., F. C. Nielsen, et al. (2008). "MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation." Mol Cell **30**(4): 460-471.

Ozsolak, F., L. L. Poling, et al. (2008). "Chromatin structure analyses identify miRNA promoters." Genes & development **22**: 3172-3183.

Papagiannakopoulos, T., A. Shapiro, et al. (2008). "MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells." Cancer research **68**(19): 8164-8172.

Paris, O., L. Ferraro, et al. (2012). "Direct regulation of microRNA biogenesis and expression by estrogen receptor beta in hormone-responsive breast cancer." Oncogene **31**(38): 4196-4206.

Paroo, Z., X. Ye, et al. (2009). "Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling." Cell **139**(1): 112-122.

Pasero, M., M. Giovarelli, et al. (2012). "Bone morphogenetic protein/SMAD signaling orients cell fate decision by impairing KSRP-dependent microRNA maturation." Cell Rep **2**(5): 1159-1168.

Pasquinelli, A. E., B. J. Reinhart, et al. (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." Nature **408**(6808): 86-89.

Piskounova, E., C. Polytarchou, et al. (2011). "Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms." Cell **147**: 1066-1079.

Piskounova, E., S. R. Viswanathan, et al. (2008). "Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28." J Biol Chem **283**(31): 21310-21314.

Piskounova, E., S. R. Viswanathan, et al. (2008). "Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28." The Journal of biological chemistry **283**: 21310-21314.

Quastler, H. (1953). Essays on the use of information theory in biology. Urbana,, University of Illinois Press.

Quintavalle, C., M. Garofalo, et al. (2012). "miR-221/222 overexpession in human glioblastoma increases invasiveness by targeting the protein phosphate PTPmu." Oncogene **31**(7): 858-868.

Rao, S. A., V. Santosh, et al. (2010). "Genome-wide expression profiling identifies deregulated miRNAs in malignant astrocytoma." Mod Pathol **23**(10): 1404-1417.

Rodriguez, A., S. Griffiths-Jones, et al. (2004). "Identification of mammalian microRNA host genes and transcription units." Genome research **14**: 1902-1910.

Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-86.

Rybak, A., H. Fuchs, et al. (2009). "The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2." Nature cell biology **11**(12): 1411-1420.

Rybak, A., H. Fuchs, et al. (2008). "A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment." Nat Cell Biol **10**(8): 987-993.

Sakamoto, S., K. Aoki, et al. (2009). "The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway." Molecular and cellular biology **29**: 3754-3769.

Schraivogel, D., L. Weinmann, et al. (2011). "CAMTA1 is a novel tumour suppressor regulated by miR-9/9* in glioblastoma stem cells." EMBO J **30**(20): 4309-4322.

Seggerson, K., L. Tang, et al. (2002). "Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation." Dev Biol **243**(2): 215-225.

Selbach, M., B. Schwanhäusser, et al. (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature **455**: 58-63.

Setty, M., K. Helmy, et al. (2012). "Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma." Mol Syst Biol **8**: 605.

Shin, C., J. W. Nam, et al. (2010). "Expanding the microRNA targeting code: functional sites with centered pairing." Mol Cell **38**(6): 789-802.

Shiohama, A., T. Sasaki, et al. (2007). "Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins." Experimental cell research **313**: 4196-4207.

Singh, P., P. S. Soon, et al. (2012). "Dysregulation of microRNAs in adrenocortical tumors." Mol Cell Endocrinol **351**(1): 118-128.

Siomi, H. and M. C. Siomi (2010). "Posttranscriptional regulation of microRNA biogenesis in animals." Molecular cell **38**(3): 323-332.

Stefani, G. and F. J. Slack (2008). "Small non-coding RNAs in animal development." Nat Rev Mol Cell Biol **9**(3): 219-230.

Sumazin, P., X. Yang, et al. (2011). "An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma." Cell **147**(2): 370-381.

Suzuki, H. I., K. Yamagata, et al. (2009). "Modulation of microRNA processing by p53." Nature **460**(7254): 529-533.

Suzuki, H. I., K. Yamagata, et al. (2009). "Modulation of microRNA processing by p53." Nature **460**: 529-533.

Thomson, J. M., M. Newman, et al. (2006). "Extensive post-transcriptional regulation of microRNAs and its implications for cancer." Genes & development **20**: 2202-2207.

Thornton, J. E. and R. I. Gregory (2012). "How does Lin28 let-7 control development and disease?" Trends Cell Biol **22**(9): 474-482.

Trabucchi, M., P. Briata, et al. (2009). "The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs." Nature **459**: 1010-1014.

van Kouwenhove, M., M. Kedde, et al. (2011). "MicroRNA regulation by RNA-binding proteins and its implications for cancer." Nature reviews. Cancer **11**(9): 644-656.

van Kouwenhove, M., M. Kedde, et al. (2011). "MicroRNA regulation by RNA-binding proteins and its implications for cancer." Nature reviews. Cancer.

Vasudevan, S., Y. Tong, et al. (2007). "Switching from repression to activation: microRNAs can up-regulate translation." Science **318**(5858): 1931-1934.

Verhaak, R. G., K. A. Hoadley, et al. (2010). "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1." Cancer cell **17**(1): 98-110.

Viswanathan, S. R., G. Q. Daley, et al. (2008). "Selective blockade of microRNA processing by Lin28." Science **320**(5872): 97-100.

Viswanathan, S. R., J. T. Powers, et al. (2009). "Lin28 promotes transformation and is associated with advanced human malignancies." Nat Genet **41**(7): 843-848.

Volk, N. and N. Shomron (2011). "Versatility of MicroRNA Biogenesis." PloS one **6**: e19391.

Wang, K., M. Saito, et al. (2009). "Genome-wide identification of post-translational modulators of transcription factor activity in human B cells." Nature biotechnology **27**: 829-839.

Wightman, B., I. Ha, et al. (1993). "Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans." Cell **75**(5): 855-862.

Wuchty, S., D. Arjona, et al. (2011). "Prediction of Associations between microRNAs and Gene Expression in Glioma Biology." PLoS One **6**(2): e14681.

Xiao, F., Z. Zuo, et al. (2009). "miRecords: an integrated resource for microRNA-target interactions." Nucleic Acids Res **37**(Database issue): D105-110.

Yamagata, K., S. Fujiyama, et al. (2009). "Maturation of microRNA is hormonally regulated by a nuclear receptor." Molecular cell **36**: 340-347.

Yang, J.-S., T. Maurin, et al. (2010). "Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis." Proceedings of the National Academy of Sciences of the United States of America **107**: 15163-15168.

Yi, R., H. A. Pasolli, et al. (2009). "DGCR8-dependent microRNA biogenesis is essential for skin development." Proc Natl Acad Sci U S A **106**(2): 498-502.

Zhang, C. Z., J. X. Zhang, et al. (2010). "MiR-221 and miR-222 target PUMA to induce cell survival in glioblastoma." Molecular cancer **9**: 229.

Zhang, J., L. Han, et al. (2010). "miR-221/222 promote malignant progression of glioma through activation of the Akt pathway." International journal of oncology **36**(4): 913-920.

Zhang, Q. C., D. Petrey, et al. (2013). "PrePPI: a structure-informed database of protein-protein interactions." Nucleic Acids Research **41**(Database issue): D828-833.