12-2014

IDENTIFICATION OF CELL SIGNALING PATHWAY REGULATED BY MICRORNAS IN CANCER CELLS USING A SYSTEMS BIOLOGICAL APPROACH

sangbae Kim

IDENTIFICATION OF CELL SIGNALING PATHWAY REGULATED BY MICRORNAS IN CANCER CELLS USING A SYSTEMS BIOLOGICAL APPROACH

by

Sangbae Kim, M.S.

APPROVED:

_____

Ju-Seog Lee, Supervisory Professor

_____

Shiaw-Yih Lin, Ph.D.

_____

Prahlad Ram, Ph.D.

_____

Peng Huang, M.D. Ph.D.

_____

Gabor Balazsi , Ph.D.

APPROVED:

_____

Dean, The University of Texas

Graduate School of Biomedical Sciences at Houston

IDENTIFICATION OF CELL SIGNALING PATHWAY REGULATED BY MICRORNAS IN CANCER CELLS USING A SYSTEMS BIOLOGICAL APPROACH

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

And

The University of Texas

MD Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Sangbae Kim, M.S.

Houston, Texas

Dec, 2014

DEDICATION

I dedicate this dissertation to my dearest family:

Serin Kim, David Kim and Mira Jeong,

my parents: Sun-Ja Park and Hee-San Kim, and

My parents in law: Bong-Yeon Choi and Jung-Hwa Jeong

ACKNOWLEDGEMENTS

IDENTIFICATION OF CELL SIGNALING PATHWAY REGULATED BY MICRORNAS IN CANCER CELLS USING A SYSTEMS BIOLOGICAL APPROACH

Sang-Bae Kim, Ph.D.

Supervisory Professor: Ju-Seog Lee, Ph.D.

MicroRNAs (miRNAs) are single-stranded, non-coding RNA molecules that regulate gene expression via imperfect binding of the miRNA to specific sites in the 3' untranslated region of the mRNAs. Because prediction of miRNA targets is an essential step for understanding the functional roles of miRNAs, many computational approaches have been developed to identify miRNA targets. However, identifying targets remains challenging due to the inherent limitation of current prediction approaches based on imperfect complementarity between miRNA and its target mRNAs. To overcome these current limitations, we developed a novel correlation-based approach that is sequence independence to predict functional targets of miRNAs by step-wise integration of the expression data of miRNAs, mRNAs, and proteins from NCI-60 cell lines. A correlation matrix between expression of miRNAs and mRNAs was first generated and later integrated with the correlation matrix between expression of mRNAs and signaling proteins. Because these integrated matrices reflect the association of miRNAs and signaling pathways, they were used to predict potential signaling pathways regulated by certain miRNAs. We implemented a web-based tool, miRPP, based on our approach. As validation of our approach, we also

demonstrated that miR-500 regulates the MAPK pathway in melanoma and breast cancer cells as predicted by our algorithms. In additional experiments, we further identified *PPFIA1* as a direct target of miR-500 that regulates *MAP2K1* in the MAPK pathway. In conclusion, we developed a systematic analysis approach that can predict signaling pathways regulated by particular miRNAs. Our approach can be used to investigate the unknown regulatory role of miRNAs in signaling pathways and gene regulatory networks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AS   Association score

ASM   Association score matrix

CD64   Fc Fragment Of IgG, High Affinity Ia, Receptor

CLCN5   Chloride Channel, Voltage-Sensitive 5

Dicer   Double-Stranded RNA-Specific Endoribonuclease

Drosha   Double-Stranded RNA-Specific Ribonuclease Type III

E2F   Transcription factors in higher eukaryotes

ECL   Enhanced chemiluminescence

ERK   Extracellular signal-regulated kinases

Exp5   Exported into the cytoplasm by the karyopherin exportin 5

GFP   Green fluorescent protein

GTP   Guanosine-5'-triphosphate

KEGG   Kyoto Encyclopedia of Genes and Genomes

MAP2K1 Mitogen-activated protein kinase kinase 1

MAPK   Mitogen-activated protein kinase

miRNA   MicroRNA

MOI   Multiplicity of infection

NCI	National Cancer Institute
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institutes of Health
PAS	Pathway Association Score
PASM	Pathway Association Score Matrix
PPFIA1	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1
PPP2R5E	Protein Phosphatase 2, Regulatory Subunit B', Epsilon Isoform
pRb	Retinoblastoma
Pre-miRNA	premature microRNA
pri-miRNA	Primary microRNA
Ran	Ras-related nuclear protein
RISC	RNA-induced silencing complex
RPPA	Reverse phase protein array

CHAPTER 1

INTRODUCTION

1.   MicroRNAs

Noncoding RNAs (ncRNAs) are functional RNA molecules which are not translated into proteins (Cech and Steitz 2014). NcRNAs play critical roles in regulating gene expression at transcription, RNA processing, and translation levels (Bogunovic et al. 2009) in various biological processes. Depending on the length and functions, ncRNAs can be classified into three groups: very small RNAs (18 - 25 nucleotides) – microRNAs (miRNAs) and small interfering RNAs (siRNAs); small RNAs (smRNAs, 20-200 nucleotides); and medium and large RNAs (piRNAs, 200-10000 nucleotides) (Table 1) (Wurdinger and Costa 2007; Esteller 2011) .

MicroRNAs are single-stranded noncoding RNA molecules and approximately 21–25 nucleotides in length, that play crucial roles in posttranscriptional regulation of gene expression (Bartel 2004). MicroRNAs are partially or fully sequence-complementary to mRNA targets, and their main function is reduce stability, expression and/or translation of mRNAs in a variety of manners, including mRNA cleavage, translational repression, and deadenylation (Bartel 2009). More than 28,000 of miRNAs have been discovered in various organisms through both experimental sequencing and computational prediction in miRBase (http://miRBase.org), which provides easy access to miRNA sequence data, annotation data, target prediction and nomenclature (Kozomara et al. 2014).

Table 1. Non-coding RNA types

Name	Size	Location	Number in humans	Functions
Short ncRNAs				
miRNAs	19–24 bp	Encoded at widespread locations	>1,424	Targeting of mRNAs and many others
piRNAs	26–31bp	Clusters, intragenic	23,439	Transposon repression, DNA methylation
tiRNAs	17–18bp	Downstream of TSSs	>5,000	Regulation of transcription?
Mid-size ncRNAs				
snoRNAs	60–300 bp	Intronic	>300	rRNA modifications
PASRs	22–200 bp	5′ regions of protein-coding genes	>10,000	Unknown
TSSa-RNAs	20–90 bp	–250 and +50 bp of TSSs	>10,000	Maintenance of transcription?
PROMPTs	<200 bp	–205 bp and –5 kb of TSSs	Unknown	Activation of transcription?
Long ncRNAs				
lincRNAs	>200 bp	Widespread loci	>1,000	Examples include scaffold DNA–chromatin complexes
T-UCRs	>200 bp	Widespread loci	>350	Regulation of miRNA and mRNA levels?
Other lncRNAs	>200 bp	Widespread loci	>3,000	Examples include X-chromosome inactivation, telomere regulation, imprinting

1-1 Biogenesis

Most of miRNAs has known to be located in the intron region of their host genes and share their mRNAs and regulatory elements resulting in a similar expression pattern (Bartel 2009; Saj and Lai 2011). Figure 1 shows canonical biogenesis of miRNA (Sullivan and Ganem 2005; Ruby et al. 2007; Winter et al. 2009). The process has several steps to produce a mature form of a miRNA from its host gene. The miRNA host gene is transcribed to a large RNA precursor (pri-miRNA) with of a 5' cap and poly-A tail3 in nucleus, and then processed to a precursor miRNA (pre-miRNA) by the complex of Drosha4 (RNase III enzyme) and Pasha/DGCR85 (double-stranded-RNA-binding protein). The pre-miRNAs becomes the imperfect stem-loop structure of about 70-nucleotides in length, and is exported into the cytoplasm by the karyopherin exportin 5 (Exp5) and Ras-related nuclear protein (Ran)-GTP complex. Next pre-miRNA is further processed to a miRNA-miRNA duplex with 22 nucleotides by the RNAse III enzyme Dicer. After one strand of this duplex is degraded, a mature miRNA with approximately 22 nucleotides is generated and binds to RNA-induced silencing complex (RISC) leading to gene silencing and RNA interference (Paroo et al. 2007).

Host genome

m7G ———— AAA
pri-miRNA

Nucleus

Drosha

DGCR8

pre-miRNA

Exp 5

Cytoplasm

pre-miRNA

TRBP

Dicer

miRNA:miRNA*
duplex

mature miRNA   5' ————— 3'

miRISC

AGO

AAA ————————— m7G

**Translational
repression**

**mRNA
degradation**

**Translational
upregulation**
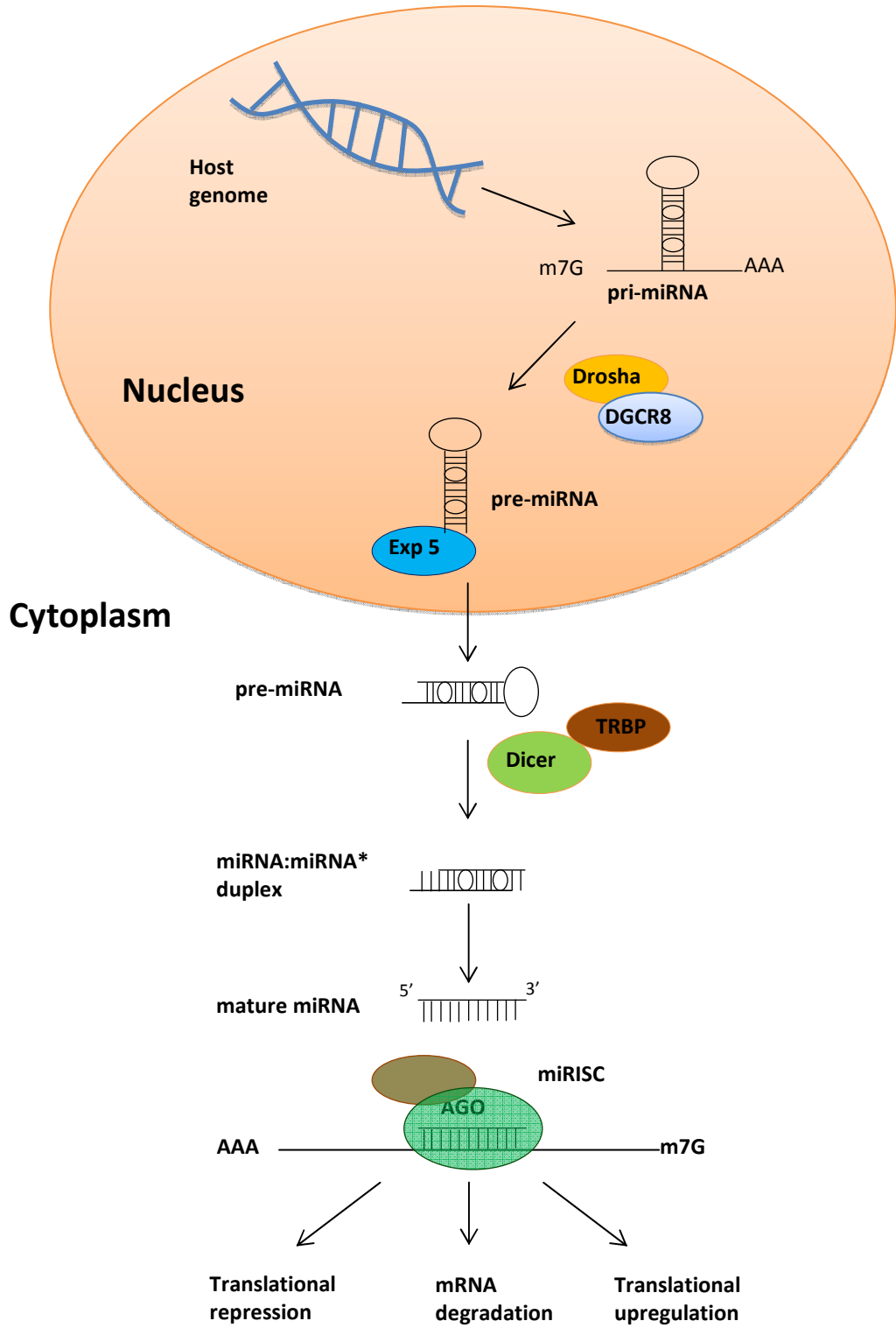
Figure 1. **Canonical biogenesis of miRNA.** miRNA genes are expressed by RNA polymerase II in the nucleus forming the primary miRNA (pri-miRNA). The stem–loop is cleaved by Drosha in conjunction with Pasha (= DGCR8 in vertebrates). The precursor miRNA (pre-miRNA) is transported into the cytoplasm by Exportin-5 (Exp 5) where it is further processed by Dicer in association with the loquacious protein to produce a ~22 nt miRNA:miRNA* duplex. The passenger strand, miRNA*, is usually degraded and the guide strand, miRNA, becomes incorporated into the RNA-induced silencing complex (Sontheimer 2005) containing the argonaute (Ago) protein. The miRNA–RISC complex interacts with the target sequences leading to repression of translation, mRNA degradation, or upregulation of transcript levels (Asgari 2011).

Modified from Asgari, Front Physiol. 2011 under permission from Frontiers

1-2 MicroRNA and mRNA target interaction

In miRNA target prediction using sequence based computational approaches, the critical step is to identify the miRNA-mRNA target interaction. There are four main characteristics to predict this interaction as followings: seed match, conservation, free energy and site accessibility (Cancer Genome Atlas Research et al. 2013). The seed sequence of a miRNA is the first 2-8 nucleotides from the 5' end (Asgari 2011). Most of miRNA target prediction tools use this seed sequence to identify the targets of a miRNA by Watson-Crick (WC) match (Asgari 2011) (**Fig.2**). Base paring pattern is important to predict miRNA targets (Maziere and Enright 2007). The potential binding sites can be defined into three groups such as 5'-dominant seed canonical, 5'-dominant seed only , and 3'-compensatory (Enright et al. 2003; Lewis et al. 2005) (**Fig.3**).

Depending on the prediction algorithms, several types of seed matches are used: 6 mer, 7-mer-m8, 7mer-A1, and 8mer (Brennecke et al. 2005). Sequence conservation across species is one of factors to predict miRNA targets. Generally miRNA seed regions are highly conserved compared with non-seed region in a miRNA (Cancer Genome Atlas 2012). Gibbs free energy in binding between a miRNA and its target mRNA can be used for a prediction measure. If it has the lower energy the stability is increased meaning more likely to be a true target (Nair et al. 2014). Site accessibility of a miRNA to a mRNA target is one of measurements for target prediction. Depending on the secondary structure of a target mRNA, miRNA:mRNA hybridization can be predicted (Cancer Genome Atlas Research 2008). Although many miRNA target prediction algorithms have

been developed, it is still challengeable due to the thousands of binding

possibility per a miRNA depending on various conditions.
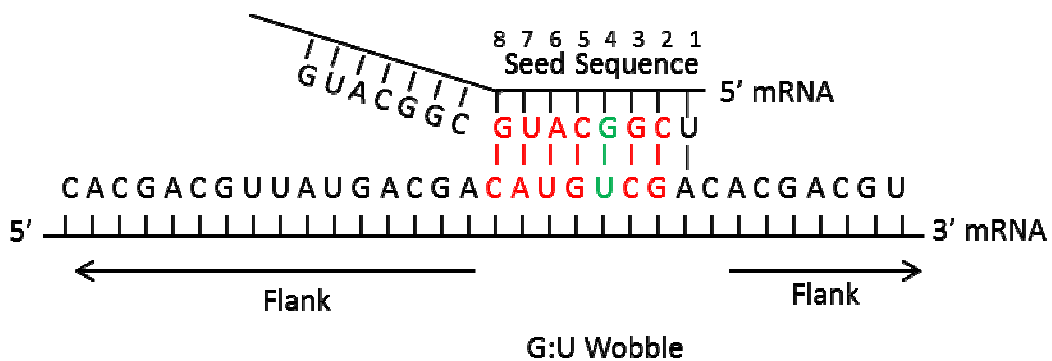
Figure 2. Schematic overview of a miRNA-mRNA target interaction. The

nucleotides 2-8 of the miRNA seed region play an important role in binding to a

target mRNA. Frank represents the outside sequence of the seed region. A G-U

wobble pair in green is shown in the middle of the interaction.

Modified from Peterson et al., Front Genet. 2014 Feb 18;5:23 with permission

from Frontiers

A

```
                                    C
                                  /   \
A-C-G-A-U-C-G-A-U-A-G-U    U-U-A-G-A-A-G-A-G-G   5' miRNA
U-G-U-U-A-G-U-U-A-U-C-A    A-A-U-C-U-U-C-U-C-U   3' UTR
                                  \   /
                                    C
                Bulge          Seed region
```

B

```
                    U — G
                  /        \
                A            U
U-U-A-G                A-G-U      /U\            U
A-A-U-C                U-C-A    U-U-A-C-U-A-G   5' miRNA
        \   C    C   /       A-A-U-G-G-U-C   3' UTR
          C — G         \ C /              C
                        Bulge  Seed
```

C

```
                    U — G
                  /        \
                A            U            /C\
G-U-A-C — C-G          A-G-G-G-U-U    A   5' miRNA
C-A-U-G   G-C          U-U-C-C-A-A    U   3' UTR
      \C/    \          \C/       \C/
           C — G              Seed region
```
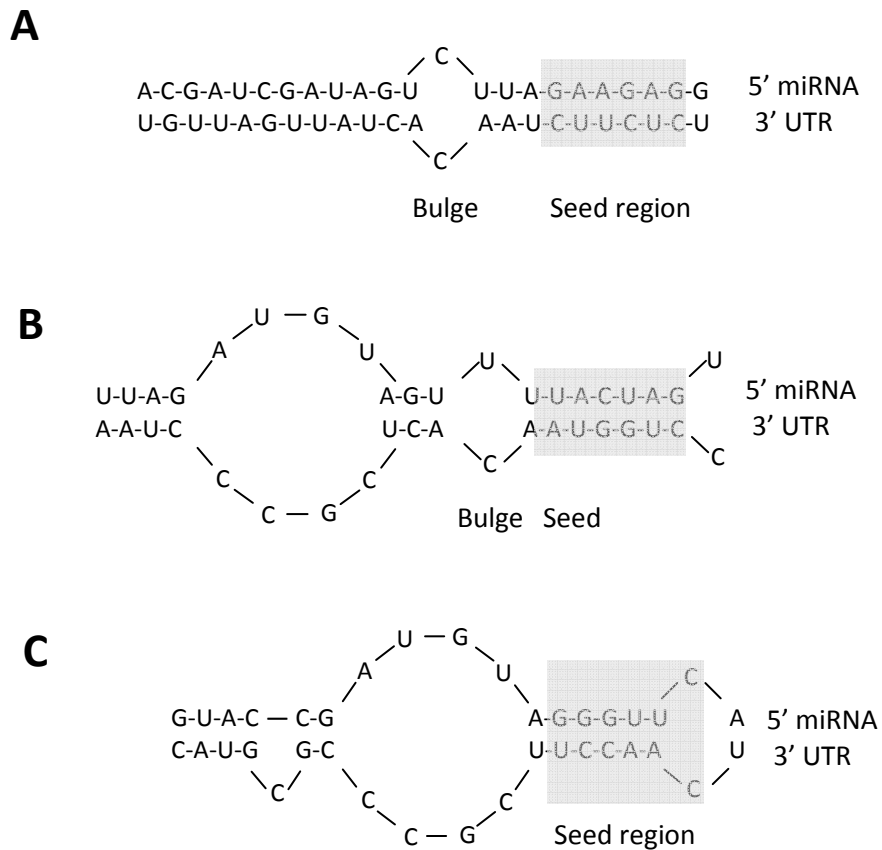
Figure 3. Major three types of secondary structures of miRNA-mRNA duplex. (A) Canonical sites with a perfect base paring , a bulge, and an extensive base pairing. (B) Dominant seed sites with perfect complementarity in the seed region, but poor complementarity in the 3' end. (C) Compensatory sites with a mismatch or G:U wobble in the seed region, but have extensive base pairing to the 3' end.

Modified from Min and Yoon, Experimental & Molecular Medicine (2010) 42, 233-244 with permission from BioMed Central.

1-3 Nomenclature of miRNA

For thousands of miRNAs across other species, the nomenclature system has been used before publication of their discovery. There are conventions for naming miRNA (Ambros et al. 2003) (**Fig.4**). In annotation a miRNA experimentally examined has a number followed by the prefix 'mir' or 'miR' and a dash such as 'mir-123'. The 'mir-' and 'miR' represent pre-miRNA and the mature form, respectively. miRNAs with similar sequences are annotated by adding lower case letter such as miR-1a and miR-1b. The same miRNA from different loci are annotated with an additional number such as miR-1-1 and miR-1-2. The species information is preceded to a miRNA. For example, hsa-miR-142 represents the miR-142 of homo sapience. The miRNAs originated from the 3' or 5' end are annotated with a -3p or 5p suffix such as miR-142-5p and miR-142-3p.
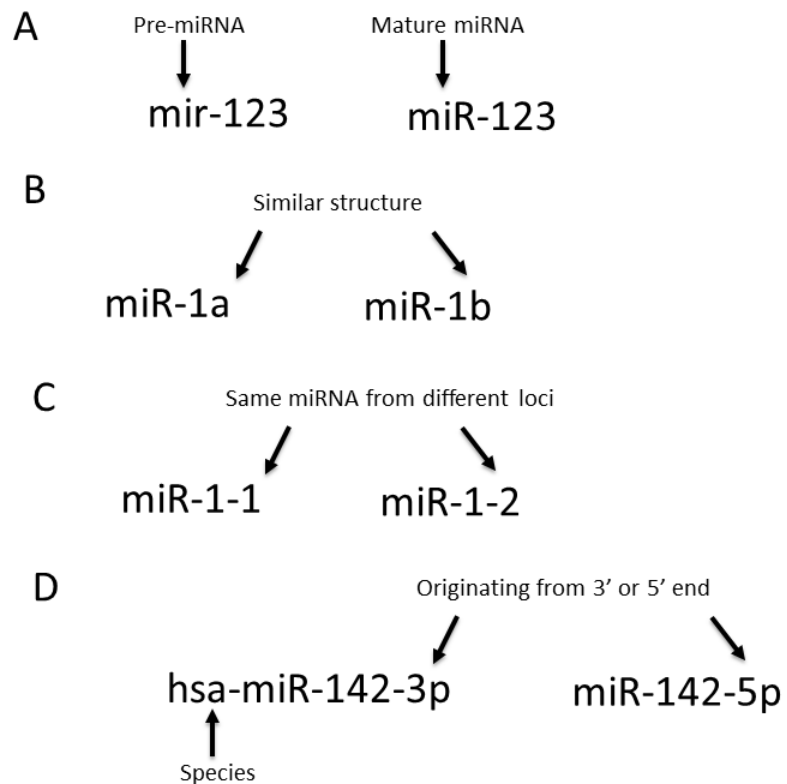
Figure 4. Nomenclature of miRNA. (A) Prefix "mir" and "miR" represent premature and mature form of miRNA. (B) Lower case letter represents similar structure of miRNA with 1 or 2 nucleotide differences (C) Additional number means the same miRNA from different loci. (D) "3p" and "5p" represent originating for the 3 and 5 end. Species is preceded to miRNA nomenclature.

1-4   miRNA action mechanism

Currently, 28645 miRNAs has been registered in miRBase (Kozomara et al. 2014). miRNAs silence gene expression by making partial base-paring with 3' untranslated region of target mRNAs (Bartel 2009). Because imperfect base-pairing with target mRNAs is sufficient for inhibition, single miRNA can target a number of genes and also multiple miRNAs can target single mRNA (Vlachos et al. 2012).

Figure 5 shows main regulatory mechanism of miRNAs. MicroRNAs can repress mRNA translation and destabilize mRNA transcripts in the processing body (P-body) in which miRNA-target mRNAs are isolated from translational process and degraded (Fazi and Nervi 2008; Romero-Cordoba et al. 2014). Furthermore miRNA regulation mechanisms can be classified to *cis*- and *trans*-regulation (Liu et al. 2009) (**Fig.6**). In *cis*-regulation, miRNAs directly bind to target mRNA sequences and regulate the gene expression and translation. In *trans*-regulation, miRNAs can indirectly regulate from gene to protein levels by targeting the mRNAs of transcription factors, RNA regulating proteins and interacting proteins.

Recent studies estimated that each miRNA can regulate more than 200 genes (Krek et al. 2005; Bussey et al. 2006; Gennarino et al. 2009), implying that miRNAs regulates a large number of biological processes that are frequently altered in many human diseases. Therefore, to understand the functional roles of miRNAs in disease, it will be axiomatic to accurately identify target mRNAs.
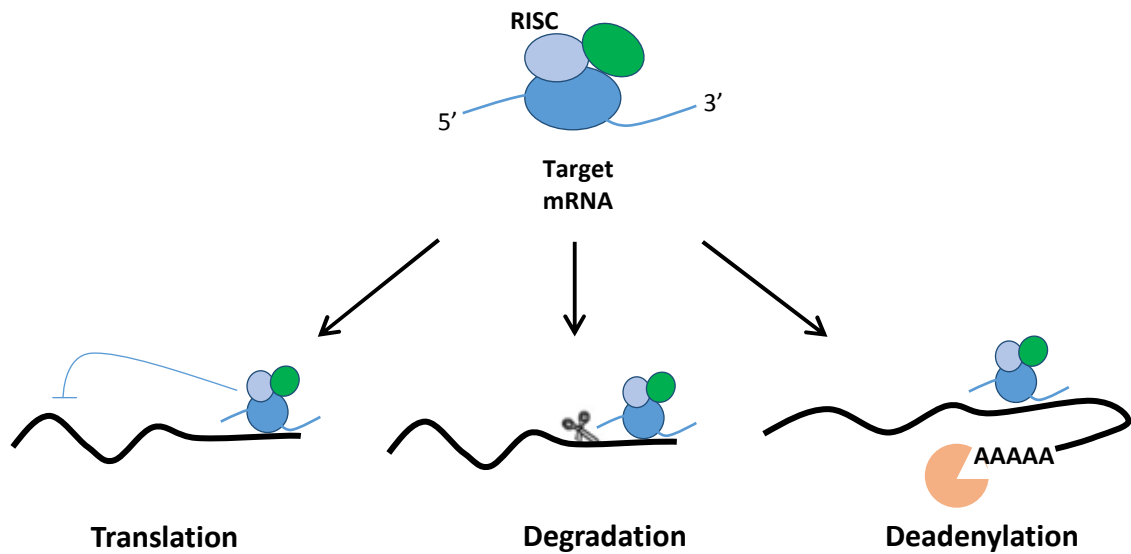
Figure 5. Major functions of miRNAs. Mature miRNAs guide the RISC complex to the 3′ untranslated regions (3′-UTR) of the complementary mRNA targets. The complexes repress mRNA translation, degrade mRNAs and destabilize by de-adenylation. Scissors indicate the cleavage on pri-miRNA or mRNA. RISC: RNA-induced silencing complex.

A. Cis-regulation

miRNA

RNA-directed DNA methylation?
Targeting gene promoter?

Targeting mRNA
1. mRNA decay
2. Inhibit translation

Gene

mRNA

B. Trans-regulation

miRNA

mRNA encoding
transcription

mRNA encoding protein
Regulating mRNA

mRNA encoding
Regulating mRNA

Transcription factor

RNA regulating protein

Interacting protein
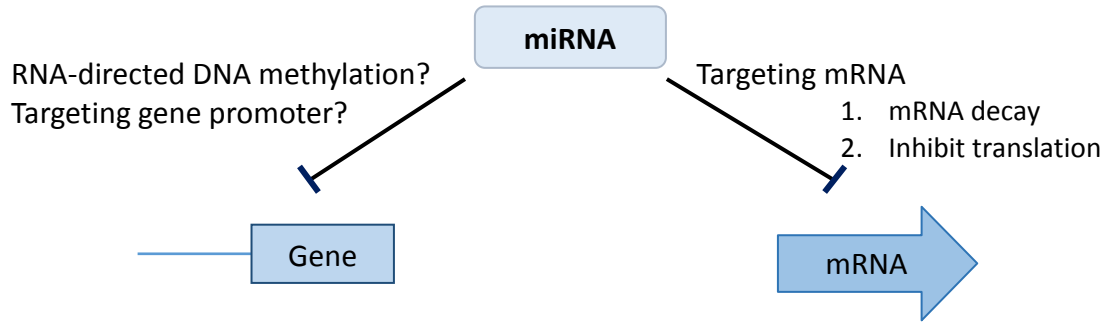
Gene

mRNA

Protein

Figure 6. Potential microRNA regulation mechanisms by multifactorial and encompassing interactions. (A) *Cis-regulation.* MicroRNAs directly target the mRNA, and control the expression of the target gene at post-transcriptional levels by mRNA degradation and inhibiting translation. (B) *Transregulation.* MicroRNAs regulate the expression changes of the targeted specific genes such as transcription factors, RNA regulating protein coding genes, and interacting protein coding genes. *Modified from Liu et al., Comp Funct Genomics. 2009:837514 with permission from Hindawi publishing corporation.*

1-5   miRNA and Cancer

miRNAs are frequently dysregulated in many human diseases including cancer (Zhang et al. 2007; Lu et al. 2008; Macfarlane and Murphy 2010). Throughout regulating the expression of key genes in these critical pathways, miRNAs can modulate the cell cycle, cellular senescence, and the DNA damage response on tumorigenesis. The previous studies have shown that miRNAs play critical roles in cancer by targeting oncogenes or tumor suppressor genes (Garzon et al. 2006; Zhang et al. 2007) (**Fig.7**). When miRNAs are oncogenic, the tumor suppressor genes are repressed by the miRNAs resulting in tumor formation (Paranjape et al. 2009). For example, mir-17-92, known as oncomir, was reported that enforced expression of mir-17-19b collaborated with the c-myc oncogene to activate B lymphomagenesis (He et al. 2005).

While miRNAs also can play roles as tumor suppressor genes. In this case the miRNA downregulation can induce tumors. The previous studies showed miRNAs as a tumor suppressor gene are repressed in human cancers: colon cancer (Michael et al. 2003), lung cancer (Takamizawa et al. 2004), breast cancer (Iorio et al. 2005), and renal cell cancer (Liu et al. 2010a). For example, miR-20a and miR-125b have been reported to regulate the expression of E2F transcription factors in the retinoblastoma pathway (O'Donnell et al. 2005).
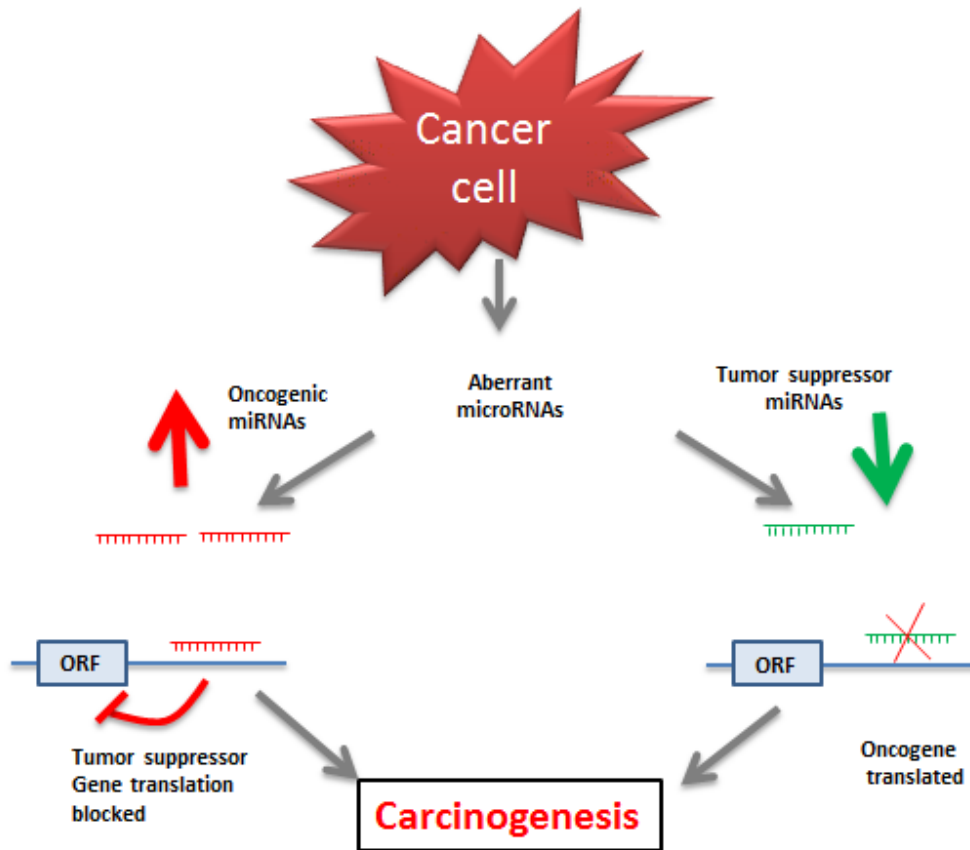
Figure 7. MicroRNA functions as tumor suppressors and oncogenes. In cancer, aberrant miRNAs function as oncogenes by inhibiting the translation of tumor suppressor genes. On the other hand, tumor suppressor miRNAs are down-regulated resulting in the increase of oncogene expression. The both events lead to tumor formation.

1-6   miRNA cluster

Around 50 % of all miRNA genes are located within 50 kb in length on the genome and transcribed together as a cluster (Megraw et al. 2007; Becker et al. 2012; Chan et al. 2012). These clusters range from 2 to several dozens of miRNAs. miRNAs in a cluster frequently shows similar sequence homology in the seed region. This results in identical targets of a miRNA cluster. The length of miRNA clusters depend on species (Chan et al. 2012). For example, mir-17 cluster family including mir-17-92, mir-106-92, and mir-106-25, and located by tandem duplications (Olive et al. 2010) (**Fig.8**). This cluster family functions in cell proliferation, apoptosis, development and cancer oncogenesis (Mogilyansky and Rigoutsos 2013). Moreover, miRNA clusters can coordinately regulate the different genes or the downstream effectors such as transcription factors in a specific signaling pathway or protein complex (Inui et al. 2010).

Because individual miRNA cluster can regulate the expression of multiple genes belonging to various signaling pathways, it could be a critical work to predict the its cellular function to target specific signaling pathways (Kuhn et al. 2008). A miRNA cluster has been identified as cooperative regulatory RNAs through targeting multiple biological processes (Guo et al. 2014). They were evolutionary correlated and simultaneously expressed (Liang et al. 2014). Although the expression pattern of individual miRNA in a cluster or a family can be different due to complex maturation and degradation mechanisms, a specific biological process can be regulated by the cluster (Xu and Wong 2008; Guo et al.

2014). Therefore, the integrative analysis based on clustered miRNAs might be a useful approach to discover the potential functional roles in tumorigenesis.
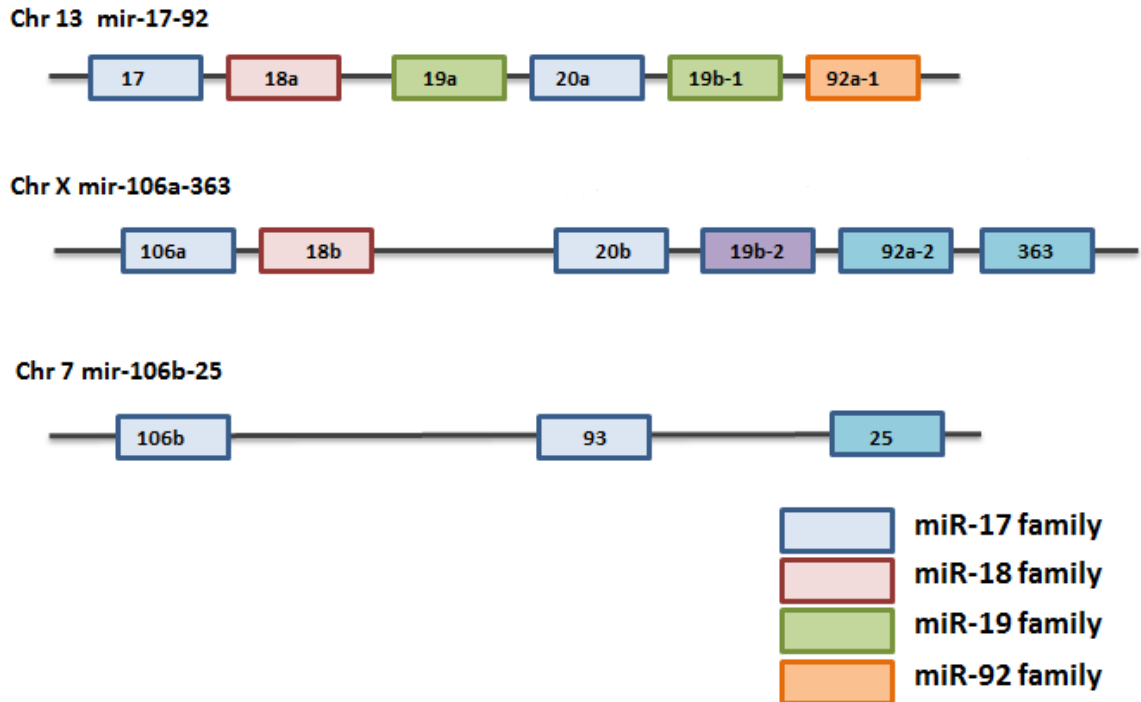
Figure 8. Human mir-17-92 as a polycistronic miRNA cluster and its homologues. Gene structure of mir-17-92 consist of miR-17, -18a, -19a, -20a, -19b-1, and -92a-1 on chromosome 13. The mir-17-92 cluster has two paralogs such as mir-106a-363 and mir-106b-25 clusters. Based on seed sequences, these miRNA clusters can be divided to four miRNA families as followings: miR-17 family (miR-17, miR-20a, miR-20b, miR-106a, miR-106b, and miR-93), the miR-18 family (miR-18a and miR-18b), the miR-19 family (miR-19a, miR-19b-1, and miR-19b-2) and the miR-92 family (miR-92a-1, miR-92a-2, miR-383, and miR-25). *Modified from Olive et al., Int J Biochem Cell Biol. 2010 Aug;42(8): 1348-54 with permission from Elsevier.*

1-7   miRNA and clinical application

MicroRNAs play critical roles in cancer and can be used as prognostic or diagnostic markers (Hung et al. 2014; Shah and Chen 2014; Ye and Cao 2014). Therapeutic approaches using miRNAs have been performed by re-introduction of miRNAs as tumor suppressors, or inhibition of oncogenic miRNAs (Jansson and Lund 2012). The expression profiles of miRNAs have been shown stable and unique signatures in the different tissue types and the stages of cancers (Lu et al. 2005; Olson et al. 2009; Iorio and Croce 2012a; Iorio and Croce 2012b) . This tissue specificity of miRNA expression patterns showed the possibility to identify the primary origin of metastatic cancer (Rosenfeld et al. 2008; Paranjape et al. 2009; Ferracin et al. 2011). MicroRNAs also can be applied to discover non-invasive biomarkers, and can measure therapeutic effect in cancer (Chen et al. 2008; Ng et al. 2009; Paranjape et al. 2009). For these clinical application, multi-omic data analysis including miRNA data might be useful in deciphering cancer biology (**Fig.9**).
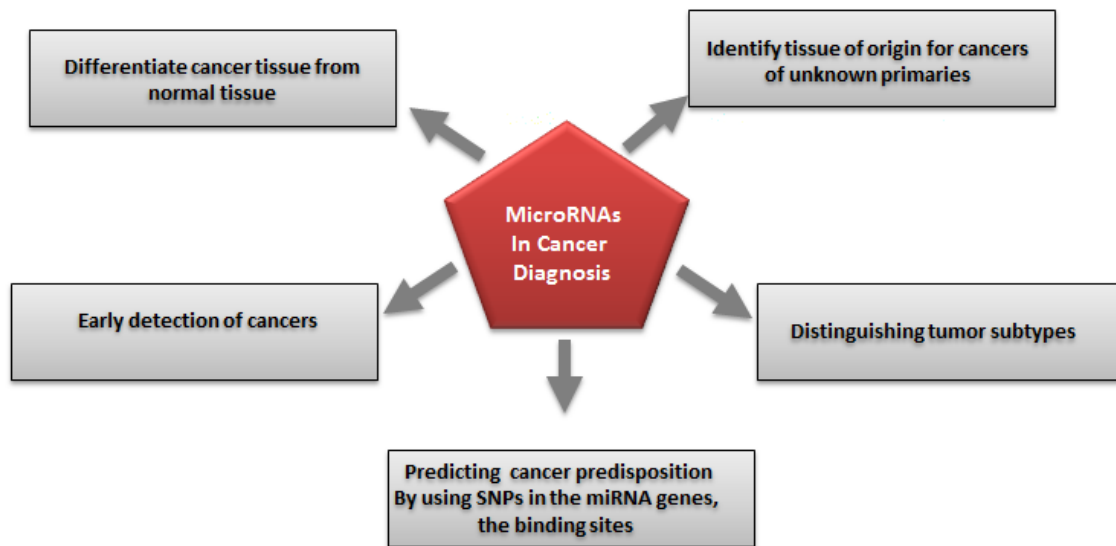
Figure 9. MicroRNAs as potential diagnostic biomarkers. In cancer diagnosis, miRNAs can be a useful tool to provide several features. MicroRNA expression profile in cancer can be used to identify cancer from normal tissues and the original tissue for cancer. Moreover, their expression patterns in one cancer type show unique, and it allows to distinguishing tumor subtype. For a diagnostic test blood-based miRNA expression patterns can be used as a not-invasive method. SNPs in miRNA genes, binding sites, and the genes of a specific pathway also can be used to predict cancer predisposition. *Modified from Paranjape et al.,Gut 2009;58:1546-1554 with permission from BMJ Publishing Group Ltd & British Society of Gastroenterology.*

2.   miRNA database and target prediction

Many computational prediction approaches using sequence complementarity have been developed with moderate success in identifying miRNA targets (Chu et al. 2013(Krek et al. 2005; Griffiths-Jones et al. 2008; Maragkakis et al. 2009; Betel et al. 2010; Dweep et al. 2011; Iorio and Croce 2012b; Hamzeiy et al. 2014). However, finding true targets is still extremely challenging because of inherited limitation of current prediction approaches based on imperfect complementarity between miRNA and its target mRNAs and a lack of a sufficiently large group of experimentally validated targets of miRNAs that can be used as a robust training set for target prediction (Rajewsky 2006).

Therefore the integrated approaches using experimental and computational analysis has been developed to identify a true target of a miRNA and miRNA databases have been constructed (Fig.10) (Table2,3).
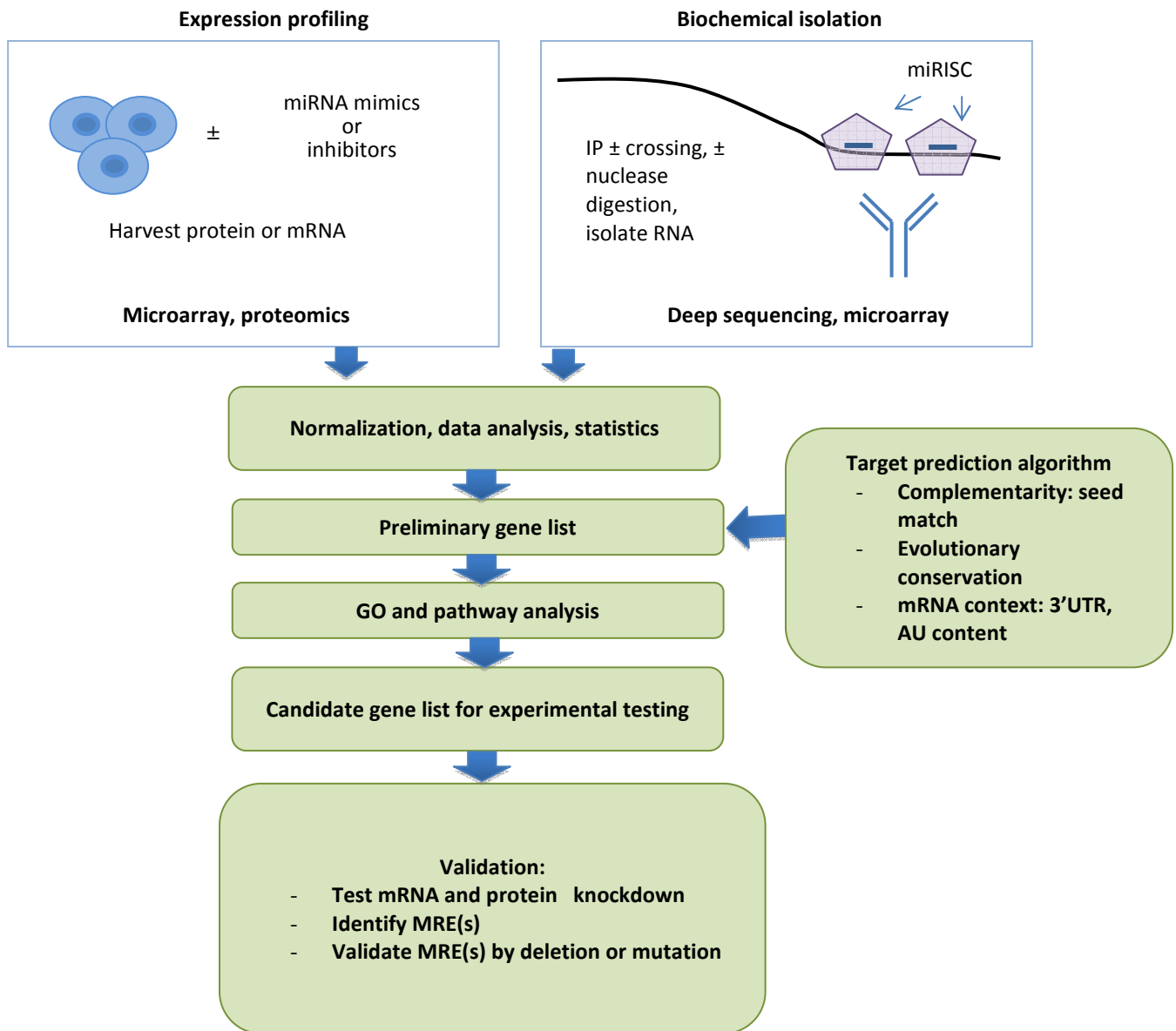
Figure 10. Schematic flow to identify miRNA targets using in-vitro and in-silico approaches. Putative target genes predicted by biochemical isolation of the miRISC or by target prediction algorithms are more than hundreds of candidates. Then, these genes can be evaluated experimentally. *Modified from Thomas et al., Nature Structural & Molecular Biology **17**, 1169–1174 (2010) with permission from Nature publishing group.*

2-1   Prediction tools

Many miRNA target prediction tools have been developed based on their own detection algorithms (Table.3). miRbase is a well-known miRNA database to offer the information of miRNA sequences, annotations and computationally predicted targets linked to other prediction tools (Bussey et al. 2006). For predicting miRNA targets RNAHybrid used the minimum free energy (MFE) of hybridization between target genes and miRNA sequences   (Kruger and Rehmsmeier   2006)   (http://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid). TargetScan used the sequence conservation scores calculated from the 6-8 mer sites of the seed region on target genes against a miRNA (Lewis et al. 2005). On the other hand, Tarbase provides experimentally validated target information (Deschenes-Simard et al. 2014).

Table 2. MicroRNA databases

Name of the database	URL	Main features	Reference
mirBase	http://microrna.sanger.ac.uk	miRNA sequences, annotations and computationally predicted targets	(Bussey et al. 2006)
Argonaute	http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface/	Detailed information about known miRNAs and their targets	(Liu et al. 2010b)
miRNAMap	http://mirnamap.mbc.nctu.edu/tw	Known as computationally predicted miRNAs and their targets	(Nishizuka et al. 2003)
Tarbase	http://www.diana.pcbi.upenn.edu/tarbase.html	Experimentally verified miRNA targets	(Deschenes-Simard et al. 2014)
Arabidopsis Small RNA project Database	http://asrp.cgrb.oregonstate.edu/	Arabidopsis miRNA sequences and corresponding target genes in addition to other small RNAs	(Akbani et al. 2014)

Table 3 . MicroRNA target prediction algorithms

Software Name	URL	Reference(s)
TargetScan, TargetScanS	http://genes. mit.edu/targetscan/	(Asgari 2011)
miRanda	http://www.microrna.org/ http://www.ebi.ac.uk/enrightsrv/microcosm/htdocs/targets/v5/	(Fazi and Nervi 2008)
Pictar	http://pictar.bio.nyu.edu/	(Olive et al. 2010)
TargetBoost	http://demo1.interagon.com/targetboost/	(Peterson et al. 2014)
DIANA-microT	http://diana.pcbi.upenn.edu/DIANA-microT	(Maragkakis et al. 2009; Maragkakis et al. 2011)
Rna22	http://cbsrv.watson.ibm.com/rna22.html	(Miranda et al. 2006)
PITA	http:// genie.weizmann.ad.il/pubs/mir07	(Tibes et al. 2006)

2-2   Limitation of DNA sequence based prediction

For miRNA target prediction, most computational methods used the nucleotide sequences with perfect seed complementarity to miRNA, and the predicted secondary mRNA structure or energetically favorable hybridization sites (Martin et al. 2007; Deschenes-Simard et al. 2014). Table 3 and 4 show the features of the prediction tools publicly available such as TargetScan (Lewis et al. 2005), miRanda (John et al. 2004), PicTar (Lall et al. 2006), DIANA-micorT (Maragkakis et al. 2009), rna22 (Miranda et al. 2006). The prediction of miRNA targets based on nucleotide sequence generated many putative targets that cannot be experimentally validated (Sethupathy et al. 2006). They have high false positive rates in target prediction up to 40% (Table 4).

Table 4. Summary for miRNA target prediction

Name	Target species	Algorithms	performance	Distinguishing feature
TargetScan	Vertebrates	Seed complementarity	FPR: 22% (mammal)	Requires 6-nt seed match and conserved Adenosine
miRanda	Flies, vertebrates	Complementarity	FPR: 24-39% (Fly)	Also provides the expression profile of miRNA in various tissues
PicTar	Vertebrates, flies	Thermodynamics	FPR: 30%	Uses cross-species comparisons to filter out false positives
DIANA-microT	Any	Thermodynamics	Precision: 66%	Target structure comes before seed complementarity
rna22	Any	Pattern recognition	FPR: 19-25.7% Sensitivity: 83%	Eliminates the use of cross-species conservation filtering, and leads to putative targets sites in 5' UTRs and ORF

Modified from Experimental & Molecular Medicine (2010) 42, 233-244 with permission from Nature publishing group

2-3　Modified prediction

Recent studies proposed new approaches that integrate transcriptomic data into complementary seed sequence data to overcome the current limitation by using a combination of target predictions and paired miRNA-mRNA expression patterns in the same tissues and cells (Liu et al. 2007; Wang et al. 2009; Wang and Li 2009) (Table 5).　However, predicted outcomes of many different prediction algorithms are frequently inconsistent, suggesting that finding a true miRNA target is still very challenging (Witkos et al. 2011). Moreover, understanding molecular functions of miRNAs are further hampered by the fact that many of predicted targets are poorly annotated.

2-4　New challenges

Recent studies demonstrated an association between miRNAs and expression of target proteins (Betel et al. 2010), suggesting that protein expression profile data can be used to verify putative functional targets and potential molecular networks regulated by miRNA (Baek et al. 2008; Selbach et al. 2008). Because proteins are functional end products of miRNA targets, proteomic approaches for identifying miRNAs targets would significantly improve our understanding on functional roles of miRNAs.

Table 5. Experimental Evaluation Results and Assessment of Commonly Used Algorithms in miRNA Target Prediction

Target prediction algorithm	Features		Experimental evaluation results			
	Parameters contributing to the final score	Cross-species conservation	Sethupathy et al. 2006 sensitivity	Baek et al. 2008 log2-fold change	Alexiou et al. 2009 precision	Alexiou et al. 2009 sensitivity
miRanda	complementarity and free energy binding	conservation filter is used	49%	0.14	29%	20%
TargetScan	seed match, 3' complementarity local AU content and position contribution	given scoring for each result	21%	0.326	51%	12%
TargetScanS	seed match type	only conservative sites are considered	48%	–	49%	8%
PicTar	binding energy, complementarity and conservation	required pairing at conserved positions	48%	0.26	49%	10%
DIANA-microT	free energy binding and complementarity	dataset of conserved UTRs among human and mouse is used	10%	–	48%	12%
PITA	target site accessibility energy	user-defined cut-off level	–	0.046	26%	6%
Rna22	pattern recognition and folding energy	not included	–	0.09	24%	6%

(Hennessy et al. 2010). RPPA data provides protein expression profiles in huge number of samples at the same time against high-quality antibodies in a

quantitative manner. As a major advantage, RPPA allows to assess target protein expression quantitatively in large sample sets while requiring only a very low amount of biological sample making this platform attractive for the analysis of clinical materials and biomarker discovery.

Furthermore, because proteomic data include critical signaling events such as phosphorylation, acetylation, and sumoylation of proteins (Nishizuka et al. 2003; Park et al. 2010), functional roles of miRNAs in key signaling pathways can be readily uncovered by correlating miRNAs with proteomic data.

4.   NCI-60 cells and data sets

In this study, we sought to integrate expression data of miRNAs, mRNAs, and proteins to uncover signaling networks regulated by miRNAs and key regulatory genes (mRNAs) of the networks directly targeted by miRNAs. We used data from panel of NCI-60 cancer cells that have been extensively used for genomic and pharmacological studies (Ross et al. 2000; Tibes et al. 2006; Blower et al. 2007; Shankavaram et al. 2007). The NCI-60 cells consist of leukemia, lymphoma, and ovarian, renal, breast, prostate, colon, lung, and central nervous system carcinoma cells(**Table.6**). NCI-60 is also the most extensively profiled set of cells, and investigators have widely used them in many cancer-related studies. In recent study (Park et al. 2010), we generated extensive proteomic data from NCI-60 cell lines by using the RPPA technique.

5. Motivation and aims

Identification of miRNAs' target genes has been challenging because of inherited limitation of current prediction methods that are largely based on imperfectly matched short sequences between miRNAs and their target sequences. Because mRNAs indirectly regulate protein expression or activation by targeting their mRNAs, we aimed to develop prediction approach that would mimic biological process of regulatory circuits in cells.

To uncover the complicated intracellular signaling networks regulated by miRNAs, we carried out multi-step integrated analyses with both transcriptome and proteome data from cancer cell lines.

In this study we aimed as the following:

- First, to develop a new approach for the integration of three independent datasets in NCI-60 data.
- Second, to establish a prediction model for identifying putative direct or indirect miRNA targets by integrating genomic and proteomic data, and to construct a web-based database
- Third, to validate a prediction model that miRNAs are highly associated with biological signaling pathway by in-vitro experiment.
- Last, to construct a web-based database to provide the data analysis outputs by our approach
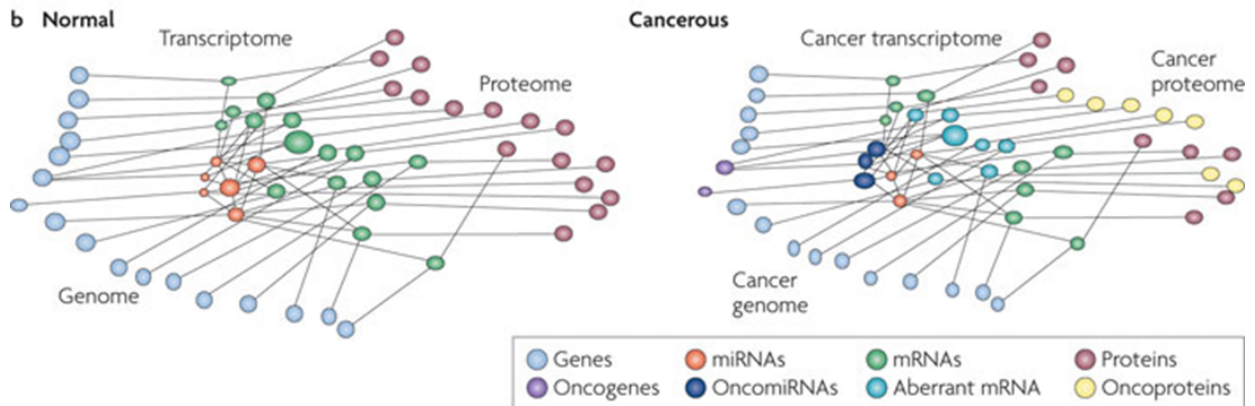
Figure11. Gene–protein network in normal tissues and in cancer. miRNAs are transcribed from miRNA non-coding genes, MicroRNA coordinately regulate multiple mRNAs affecting the output of many proteins. miRNAs have a crucial role in keeping the gene–protein network interconnected. In cancer aberrant miRNA and mRNA expression can occurs and induce the expression of oncogenic proteins that cause a certain cancer phenotype. Thus miRNAs coordinate gene regulatory networks at genomic and proeomic level.

Adopted from Garzon et al., Nature Reviews Drug Discovery 9, 775-789. 2010 with permission from Nature publishing group

CHAPTER 2

MATERIALS AND METHODS

1.   The datasets

The mRNA expression data of NCI-60 cell lines (**Table.6**) were generated by using Affymetrix HG-U133 44K platform and available from Gene Express Omnibus (GEO) database (accession number, GSE5720). MiRNA expression data were generated by using custom-made microarray from Ohio State University Comprehensive Cancer Center (OSU-CCC-hsa-miRNA-chip-V3) and available from ArrayExpress (accession number E-MEXP-1029) (Blower et al. 2007). This microarray contains 627 probes representing 423 unique miRNAs. Protein expression data were generated by using custom-made RPPA from University of Texas MD Anderson Cancer Center as described in Park et al (Park et al. 2010). RPPA platform contains 167 unique protein features.

2.   Data preprocessing

The mRNA and miRNA data sets were log2 transformed and normalized using quantile normalization in Bioconductor (www.bioconductor.org). As a large fraction of mRNAs and miRNAs were either not expressed or non-detectable, we filtered out 30% of total probes with low variance across all samples to reduce potential noise. The final number of mRNA and miRNA probes reduced to 25,306 and 436 probes for final data analysis respectively. Multiple values from multiple probes map to single genes by calculating the average of their values.

Table 6. Tumor tissue types of NCI-60 cell lines*

* http://discover.nci.nih.gov/cellminer

Tissue of origin	Cell Line Name	Age	Sex	Histology	Source	p53	doubling time
Breast	BT_549	72	F	Papillary infiltrating ductal carcinoma-mammary gland; breast	Metastasis	-	53.9
Breast	HS578T	74	F	Carcinosarcoma-mammary gland; breast	Primary	MT	53.8
Breast	MCF7	69	F	Adenocarcinoma- mammary gland; breast; metastatic site: pleural effusion;	Pleural effusion	WT	25.4
Breast	MCF_7/AdrR	NA	F	Adenocarinoma	NA	MT	34
Breast	MDA_MB_231	51	F	Adenocarcinoma-mammary gland; breast; epithelial; metastatic site	Pleural effusion	MT	41.9
Breast	MDA_MB_435	31	F	Ductal carcinoma- mammary gland; breast; duct; metastatic site	Pleural effusion	MT	25.8
Breast	T47D	54	F	infiltrating ductal carcinoma	NA	MT	45.5
CNS	SF_268	24	F	Glioblastoma, ud	NA	MT	33.1
CNS	SF_295	67	F	Glioblastoma, ud	NA	MT	29.5
CNS	SF_539	34	F	Glial cell neoplasm	NA	WT	35.4
CNS	SNB_19	47	M	Glioblastoma, ud	NA	MT	34.6
CNS	SNB_75	NA	F	Astrocytoma	NA	MT	62.8
CNS	U251	75	M	Glioblastoma,ud	NA	MT	23.8
Colon	COLO205	70	M	Adenocarcinoma	Ascites	MT	23.8
Colon	HCC_2998	NA	NA	carcinoma	NA	MT	31.5
Colon	HCT_116	NA	M	carcinoma-vpd	NA	-	17.4
Colon	HCT_15	NA	NA	Adenocarcinoma p/md	NA	-	20.6
Colon	HT29	44	F	Adenocarcinoma-md	Primary	MT	19.5
Colon	KM12	NA	NA	Adenocarcinome-pd	NA	MT	23.7
Colon	SW_620	51	M	Carcinoma-ud	NA	MT	20.4
Leukemia	CCRF_CEM	4	F	ALL	NA	MT	26.7
Leukemia	HL_60	36	M	Pro myelocytic leukemia	PBL	MT	28.6
Leukemia	K_562	53	F	CML	Pleural effusion	MT	19.6
Leukemia	MOLT_4	19	M	ALL (cells were taken when patient was in relapse)	PB	WT	27.9
Leukemia	RPMI_8226	61	M	Myeloma	PB	WT	33.5
Leukemia	SR	11	M	Lymphoma	NA	-	28.7
Melanoma	LOXIMVI	58	M	Malignant amelanotic melanoma	NA	WT	20.5

Melanoma	M14	NA	NA	Melanotic melanoma	NA	MT	26.3
Melanoma	MALME_3M	43	M	Malignant melanotic melanoma	Metastasis	WT	46.2
Melanoma	SK_MEL_2	60	M	Malignant melanotic melanoma	Metastasis	WT	45.5
Melanoma	SK_MEL_28	51	M	Malignant melanotic melanoma	NA	MT	35.1
Melanoma	SK_MEL_5	24	F	Malignant melanotic melanoma	Metastasis	WT	25.2
Melanoma	UACC_257	NA	NA	Melanotic melanoma	NA	WT	38.5
Melanoma	UACC_62	NA	NA	Melanotic melanoma	NA	WT	31.3
Non-Small Cell Lung	A549	58	M	Adenocarcinoma-p/md	NA	WT	22.9
Non-Small Cell Lung	EKVX	NA	M	Adenocarcinoma-md	NA	MT	43.6
Non-Small Cell Lung	NCI_H226	NA	M	Squamous cell carcinoma-vpd	NA	MT	61
Non-Small Cell Lung	NCI_H23	NA	M	Adenocarcinoma-ud	NA	MT	33.4
Non-Small Cell Lung	NCI_H322M	52	M	Small cell Bronchioalveolar Carcinoma	NA	MT	35.3
Non-Small Cell Lung	NCI_H460	NA	M	Large Cell Carcinoma-ud	Pleural effusion	WT	17.8
Non-Small Cell Lung	NCI_H522	NA	M	Adenocarcinoma-vpd	NA	MT	38.2
Non-Small Cell Lung	HOP_62	60	F	adenocarcinoma-ud	NA	MT	39
Non-Small Cell Lung	HOP_92	62	M	Large cell-ud	NA	MT	79.5
Ovarian	IGROV1	47	F	Cystoadenocarcinoma-pd	NA	MT	31
Ovarian	OVCAR_3	60	F	Adenocarcinoma-md	Ascites	MT	34.7
Ovarian	OVCAR_4	42	F	Adenocarcinoma-md	NA	WT	41.4
Ovarian	OVCAR_5	67	F	Adenocarcinoma-wd	NA	MT	48.8
Ovarian	OVCAR_8	64	F	Carcinoma-ud	NA	MT	26.1
Ovarian	SK_OV_3	64	F	Adenocarcinoma-vpd	Ascites	-	48.7
Prostate	DU_145	69	M	prostate; metastatic site: brain; carcinoma	Metastasis	-	32.3
Prostate	PC_3	62	M	Adenocarcinoma- prostate; metastatic site: bone;	NA	MT	27.1
Renal	786_0	58	M	Adenocarcinoma	NA	MT	22.4
Renal	A498	52	F	Adenocarcinoma	NA	WT	66.8
Renal	ACHN	22	M	Renal cell carcinoma-p/md	NA	WT	27.5
Renal	CAKI_1	49	M	Clear cell carcinoma	Metastasis	WT	39
Renal	RXF_393	54	M	hypernephroma-pd	NA	MT	62.9
Renal	SN12C	43	M	Renal cell carcinoma-pd	NA	MT	29.5
Renal	TK_10	43	M	Renal Spindle cell carcinoma	NA	MT	51.3
Renal	UO_31	NA	F	Renal cell carcinoma-vpd	NA	WT	41.7

3. Correlation analysis between miRNA-mRNA and mRNA-RPPA pairs

To identify the functional targets of miRNA, Pearson's correlation coefficient approach applied to the miRNA, mRNA and protein array data generated from NCI-60 cell lines. The log-normalized values were used as input data. To explore the association among the expression profiles of NCI-60 data sets we calculated the correlation coefficient values among all miRNA, mRNA and protein expression profiles; all-against-all miRNA-mRNA and mRNA-protein pairs. Overall data process is shown in Figure 1 and 2. The Pearson correlation between two data sets was calculated by

$$ r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)S_x S_y} $$

where X denotes expression value of a miRNA or a protein, and Y denotes expression value of a mRNA. and are the mean of X and Y. Sx and Sy are the standard deviations of all X and Y values of each probe. We defined the miRNA-mRNA correlation coefficient *P*-value less than 0.005 to be statistically significant.

4.   Generation of miRNA and Pathway association

To establish the association between miRNAs and signaling pathways, we carried out multi-step data analyses and integrated analyzed outcomes. We started this process by generating association score. First, we computed correlation between miRNAs and mRNAs by using Pearson correlation test with stringent cut-off ($P < 0.005$). Likewise, correlation between protein features and mRNAs were computed by using Pearson correlation test ($P < 0.005$). Second, in order to estimate the association between miRNA and protein, two independently generated correlation lists (miRNAs and protein features) were merged together by using correlated mRNAs as denominators. Computed association scores were saved in matrix format (222 RPPA probes x 436 miRNA = 96,792) for further analysis (association score matrix or ASM). In ASM, correlation between miRNAs and protein features were represented as number of commonly shared correlated mRNAs. For example, if miRNA-X and RPPA-Y has 5 correlated mRNAs respectively and share 3 mRNAs, association score between miRNA-X and RPPA-Y will be 3 (**Fig. 12**).
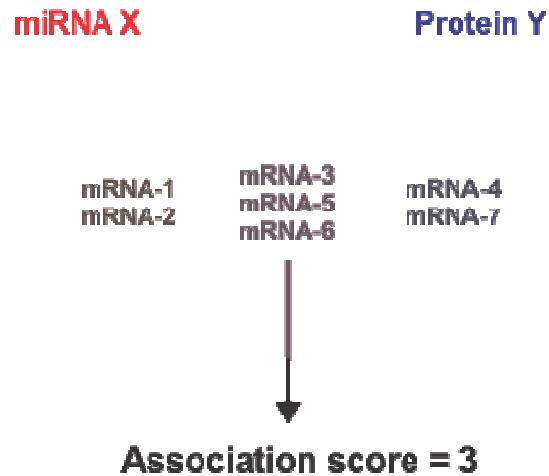
Figure 12. Schematic diagram for Association Scores (AS) was estimated by number of commonly shared correlated mRNAs between miRNAs and protein features. For example, if miRNA-X and RPPA-Y has 5 correlated mRNAs respectively and share 3 mRNAs, association score between miRNA-X and RPPA-Y will be 3. Finally Association Scores Matrix (ASM) was generated between miRNAs and RPPA proteins.

Next, we assessed significance of signaling events in NCI-60 cells by estimating correlated number of mRNAs with phosphorylation of signaling proteins. To estimate the signaling influence we assumed that a phosphorylated protein might be more impact on gene expression more than its unmodified form in signaling pathway.

RPPA data includes 25 pairs of antibodies that can specifically recognize phosphorylated or unmodified same proteins. Establishing a Pearson's correlation test P-value of less than 0.005 as indicative of significance with expression or phosphorylation patterns of the each protein feature, we first generated lists of correlated mRNAs for expression (unmodified form)) and phosphorylation level (modified form) of proteins. Second, with paired numbers of mRNAs significantly associated with phosphorylated and unmodified protein, P/U ratios (number of mRNAs associated with phosphorylated protein/number of mRNAs associated with unmodified protein) were generated and used as indicator of signaling activity of particular protein in NCI-60 cell lines. For example, if number of mRNAs correlated with phosphorylation of particular protein (P value) is 9 and number of mRNAs correlated with unmodified same proteins (U value) is 3, then P/U ratio of the protein will be 3 (**Fig.13**).

We also identified signaling pathways that are potentially regulated by 25 signaling proteins in RPPA data (**Table 7**).
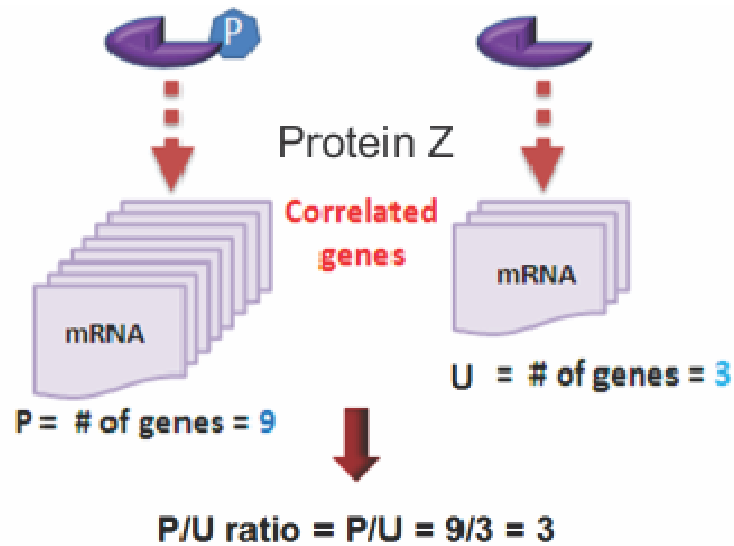
Figure 13 . Schematic diagram for generation of P/U ratios. Estimating cell signaling influence weight, P/U ratio. It was calculated by P/U ratios (number of mRNAs associated with phosphorylated protein/number of mRNAs associated with unmodified protein). For example, if number of mRNAs correlated with phosphorylation of particular protein (P value) is 9 and number of mRNAs correlated with unmodified same proteins (U value) is 3, then P/U ratio of the protein will be 3.

Table 7. Phosphorylation specific probe information of RPPA data

No	Gene	Phospho-sepcific probe
1	ACACA	PS79
2	AKT1	PS473/ PT308
3	EIF4EBP1	PS65/ PT37.46
4	ERBB2	PY1248
5	ESR1	PS167/ PS118
6	FOXO3	PS318.S321
7	FRAP1	PS2448
8	GSK3A	PS21.S9
9	IRS1	PS307
10	MAP2K1	PS217.S221
11	MAPK1	PMAPK1/PMAPK8/JNK
12	MAPK14	PT180.Y182
13	MYC	PT58.S62
14	PDK1	PS241
15	PRKAA1	PT172
16	PRKCA	PS567
17	RPS6	PS235.S236/ PS240.S244
18	RPS6KA1	PT389.S363
19	RPS6KB1	PT389
20	SGK	PS78
21	SRC	PY416/ PY527
22	STAT3	PS727/ PY705
23	STAT5A	PY694
24	STAT6	PY641
25	TSC2	PT1462

To do this, we carried out pathway enrichment analysis after mapping 25 proteins in signaling pathways by using the DAVID bioinformatics resource (http://david.abcc.ncifcrf.gov/) in NIAID, NIH. We used a significance threshold *P*-value adjusted by Benjamini of 0.05 for KEGG and BIOCARTA pathways to control the false discovery rate (Huang da et al. 2009). This analysis showed that 40 signaling pathways can be potentially regulated by 25 signaling proteins (**Table 8**).

Table 8. Significantly enriched signaling pathways with 25 P/U pair proteins†

Pathway	DB	Count	Genes –NP/P pair protein	PValue	Bonferroni	Benjamini
ErbB signaling pathway	KEGG	12	PRKCA, AKT1, MAPK1, EIF4EBP1, MAP2K1, ERBB2, STAT5A, MAPK8, RPS6KB1, MTOR, MYC, SRC	2.20E-14	1.47E-12	1.47E-12
Insulin signaling pathway	KEGG	11	AKT1, MAPK1, EIF4EBP1, MAP2K1, TSC2, ACACA, MAPK8, PRKAA1, RPS6KB1, MTOR, IRS1	1.06E-10	7.12E-09	2.37E-09
mTOR signaling pathway	KEGG	8	AKT1, MAPK1, EIF4EBP1, RPS6KA1, TSC2, PRKAA1, RPS6KB1, MTOR	1.67E-09	1.12E-07	2.79E-08
Neurotrophin signaling pathway	KEGG	9	PDK1, AKT1, MAPK1, RPS6KA1, MAP2K1, MAPK14, MAPK8, FOXO3, IRS1	3.60E-08	2.41E-06	4.82E-07
Regulation of eIF4e and p70 S6 Kinase		8	PRKCA, AKT1, MAPK1, EIF4EBP1, MAPK14, RPS6KB1, MTOR, IRS1	2.04E-08	2.85E-06	2.85E-06
Fc epsilon RI signaling pathway	KEGG	7	PDK1, PRKCA, AKT1, MAPK1, MAP2K1, MAPK14, MAPK8	8.80E-07	5.90E-05	6.55E-06
Adipocytokine signaling pathway	KEGG	6	AKT1, MAPK8, PRKAA1, MTOR, IRS1, STAT3	9.56E-06	6.41E-04	5.34E-05
VEGF signaling pathway	KEGG	6	PRKCA, AKT1, MAPK1, MAP2K1, MAPK14, SRC	1.67E-05	0.001118	8.60E-05
GnRH signaling pathway	KEGG	6	PRKCA, MAPK1, MAP2K1, MAPK14, MAPK8, SRC	6.13E-05	0.0041	2.42E-04
MAPK signaling pathway	KEGG	8	PRKCA, AKT1, MAPK1, RPS6KA1, MAP2K1, MAPK14, MAPK8, MYC	1.21E-04	0.008048	4.49E-04
Human Cytomegalovirus and Map Kinase Pathways	BIOCARTA	5	AKT1, MAPK1, MAP2K1, MAPK14, RB1	3.74E-05	0.005223	0.001744
Links between Pyk2 and Map Kinases	BIOCARTA	6	PRKCA, MAPK1, MAP2K1, MAPK14, MAPK8, SRC	2.77E-05	0.003871	0.001938
Bioactive Peptide Induced Signaling Pathway	BIOCARTA	6	PRKCA, MAPK1, MAP2K1, MAPK14, STAT5A, MAPK8	5.77E-05	0.008041	0.002016
Fc gamma R-mediated phagocytosis	KEGG	5	PRKCA, AKT1, MAPK1, MAP2K1, RPS6KB1	7.71E-04	0.050366	0.002458
Toll-like receptor signaling pathway	KEGG	5	AKT1, MAPK1, MAP2K1, MAPK14, MAPK8	9.71E-04	0.063027	0.002955
Multiple antiapoptotic pathways from IGF-1R signaling lead to BAD phosphorylation	BIOCARTA	5	AKT1, MAPK1, RPS6KA1, MAP2K1, IRS1	1.10E-04	0.015258	0.00307
Transcription factor CREB and its extracellular signals	BIOCARTA	5	PRKCA, AKT1, MAPK1, RPS6KA1, MAPK14	1.10E-04	0.015258	0.00307
Chemokine signaling pathway	KEGG	6	AKT1, MAPK1, MAP2K1, GSK3A, FOXO3, STAT3	0.00125	0.080358	0.003345
T cell receptor signaling pathway	KEGG	5	PDK1, AKT1, MAPK1, MAP2K1, MAPK14	0.001248	0.080275	0.003481
NFAT and Hypertrophy of the heart (Transcription in the broken heart)	BIOCARTA	6	AKT1, MAPK1, MAP2K1, MAPK14, MAPK8, RPS6KB1	1.64E-04	0.02276	0.00383

Top 20 of 40 Pathways (Bejamini-Hochberg procedure, corrected **P-value < 0.05**)

† Proteins with a pair of non-phospho and phospho protein in the protein array data

By using analyzed outcomes from previous steps, we next to find signaling pathways associated with each miRNA. First, we generated matrix of miRNAs vs. P/U ratios of each proteins by using association scores in ASM from previous analysis. Among 222 antibodies used in RPPA data, we only selected 25 pairs of antibodies for analysis. P/U ratios were generated by using number of associated mRNAs from ASM. In this new matrix (436 miRNAs x 25 P/U ratios of signaling proteins), signaling strength of each proteins per each miRNA is presented as P/U ratios. Second, under the assumption that a protein with higher P/U ratio would play more active roles in regulation of signaling pathway, we computed the sum of P/U ratios of the proteins in each signaling pathway per miRNA, generating Pathway Association Score (PAS) (**Fig.14**). PAS*i* per a miRNA, *i*, was calculated as following:

$$\text{PAS}i = \sum_{k=1}^{n} R$$

Where n is the total number of protein pairs belonging to a specific pathway and R is the P/U ratio value of each protein pair. For example, if there are 12 proteins with P/U ratios in a specific pathway, we can calculate the sum of the 12 P/U ratio values as a PAS per a miRNA. We computed PAS values for 40 signaling pathways for all miRNAs, yielding new matrix of 436 miRNAs x 40 signaling pathways (Pathway Association Score Matrix, PASM). In later analysis, we used the ranked PAS values as indicator of signaling strength associated with miRNAs.

Figure 14. miRNA-Pathway association score. MicroRNA-Pathway association score. After generating Association Score Matrix (ASM) between miRNAs (Mi) and P/U ratios (Ri) of each protein, we computed the sum of P/U ratios of the proteins in each signaling pathway per each miRNA and calculated Pathway Association Score (PAS) that indicated the association of miRNAs to signaling pathways. Next, the matrix of PAS (PASM) was constructed. Diagram represents the sample pathway as following: circle-pathway.I, square-pathway.II, and triangle-pathway. AS: Association Score, PAS: Pathway Association Score, Mi: a miRNA, Ri: AS ratio of P/U protein pair, and Sik: PAS between miRNA Mi and Pathway k.

5.   miRNA cluster enrichment analysis

Previous study identified miRNA clusters in human genome (52 clusters) (Ruepp et al. 2010) (**Table 9**). Because miRNAs of the same cluster have been known to share common target mRNAs and similar biological function (Gurtan and Sharp 2013), we carried out enrichment analysis of miRNAs with pathways by using values in PASM and miRNA genome cluster information in order to identify the signaling pathways potentially regulated by miRNA.

For the miRNA-pathway enrichment test we used the modified parametric analysis of gene set enrichment algorism (Kim and Volsky 2005). For each pathway, *j*, we calculated the enrichment score (Kim and Volsky 2005) based on PASM as following:

$$Ej = \frac{(Ux - Ut) \times \sqrt{N}}{S}$$

where U*x* denotes the mean PAS values within each miRNA cluster. U*t* denotes the mean of PAS against all miRNAs. N denotes the number of all miRNAs and S denotes the standard deviation of all PAS values against each pathway. Enrichment score f converted into *P*-value by applying the cumulative standard normal distribution function using T-profiler algorithm (Boorsma et al. 2005) in R. Finally we generated the pathway-miRNA cluster association matrix (52 miRNA clusters x 40 pathways). Significant pathway enriched miRNA clusters were

selected based on the $P < 0.05$. All analyses were performed using the R

Bioconductor statistical programming platform (http://bioconductor.org).

Table 9. MicroRNA cluster family

Cluster	miRNA cluster
1	hsa-mir-371|hsa-mir-372|hsa-mir-373|hsa-mir-512-1|hsa-mir-512-2|hsa-mir-498|hsa-mir-520e|hsa-mir-515-1|hsa-mir-519e|hsa-mir-520f|hsa-mir-515-2|hsa-mir-519c|hsa-mir-520a|hsa-mir-526b|hsa-mir-519b|hsa-mir-525|hsa-mir-523|hsa-mir-518f|hsa-mir-520b|hsa-mir-518b|hsa-mir-526a-1|hsa-mir-520c|hsa-mir-518c|hsa-mir-524|hsa-mir-517a|hsa-mir-519d|hsa-mir-521-2|hsa-mir-520d|hsa-mir-517b|hsa-mir-520g|hsa-mir-516b-2|hsa-mir-526a-2|hsa-mir-518e|hsa-mir-518a-1|hsa-mir-518d|hsa-mir-516b-1|hsa-mir-518a-2|hsa-mir-517c|hsa-mir-520h|hsa-mir-521-1|hsa-mir-522|hsa-mir-519a-1|hsa-mir-527|hsa-mir-516a-1|hsa-mir-516a-2|hsa-mir-519a-2|hsa-mir-1323|hsa-mir-1283-1|hsa-mir-1283-2
2	hsa-mir-134|hsa-mir-154|hsa-mir-299|hsa-mir-376c|hsa-mir-369|hsa-mir-376a-1|hsa-mir-377|hsa-mir-379|hsa-mir-380|hsa-mir-381|hsa-mir-382|hsa-mir-323|hsa-mir-329-1|hsa-mir-329-2|hsa-mir-453|hsa-mir-409|hsa-mir-412|hsa-mir-410|hsa-mir-376b|hsa-mir-485|hsa-mir-487a|hsa-mir-494|hsa-mir-495|hsa-mir-496|hsa-mir-539|hsa-mir-544|hsa-mir-376a-2|hsa-mir-487b|hsa-mir-411|hsa-mir-654|hsa-mir-655|hsa-mir-656|hsa-mir-758|hsa-mir-668|hsa-mir-1185-2|hsa-mir-1185-1|hsa-mir-300|hsa-mir-541|hsa-mir-889|hsa-mir-543|hsa-mir-1197
3	hsa-mir-16-2|hsa-mir-15b
4	hsa-mir-105-1|hsa-mir-105-2|hsa-mir-767
5	hsa-mir-127|hsa-mir-136|hsa-mir-370|hsa-mir-337|hsa-mir-431|hsa-mir-433|hsa-mir-493|hsa-mir-432|hsa-mir-770|hsa-mir-665
6	hsa-mir-34b|hsa-mir-34c
7	hsa-mir-144|hsa-mir-451
8	hsa-mir-365-1|hsa-mir-193b
9	hsa-mir-224|hsa-mir-452
10	hsa-mir-421|hsa-mir-374b
11	hsa-mir-449a|hsa-mir-449b
12	hsa-mir-296|hsa-mir-298
13	hsa-mir-424|hsa-mir-450a-1|hsa-mir-450a-2|hsa-mir-503|hsa-mir-542|hsa-mir-450b
14	hsa-mir-215|hsa-mir-194-1
15	hsa-mir-221|hsa-mir-222
16	hsa-mir-141|hsa-mir-200c
17	hsa-let-7c|hsa-mir-99a
18	hsa-mir-195|hsa-mir-497
19	hsa-mir-143|hsa-mir-145
20	hsa-mir-23a|hsa-mir-24-2|hsa-mir-27a
21	hsa-let-7g|hsa-mir-135a-1
22	hsa-mir-181b-1|hsa-mir-181a-1
23	hsa-let-7a-3|hsa-let-7b
24	hsa-let-7a-2|hsa-mir-100|hsa-mir-125b-1
25	hsa-mir-25|hsa-mir-93|hsa-mir-106b
26	hsa-mir-24-1|hsa-mir-23b|hsa-mir-27b
27	hsa-mir-181c|hsa-mir-181d
28	hsa-mir-191|hsa-mir-425

29	hsa-mir-30c-1|hsa-mir-30e
30	hsa-let-7a-1|hsa-let-7d|hsa-let-7f-1
31	hsa-mir-193a|hsa-mir-365-2
32	hsa-mir-30a|hsa-mir-30c-2
33	hsa-mir-15a|hsa-mir-16-1
34	hsa-mir-19b-2|hsa-mir-92a-2|hsa-mir-106a|hsa-mir-363|hsa-mir-18b|hsa-mir-20b
35	hsa-mir-302a|hsa-mir-302b|hsa-mir-302c|hsa-mir-302d|hsa-mir-367
36	hsa-mir-216a|hsa-mir-217|hsa-mir-216b
37	hsa-mir-29a|hsa-mir-29b-1
38	hsa-mir-212|hsa-mir-132
39	hsa-mir-199a-2|hsa-mir-214
40	hsa-mir-29b-2|hsa-mir-29c
41	hsa-mir-188|hsa-mir-362|hsa-mir-500|hsa-mir-501|hsa-mir-502|hsa-mir-532|hsa-mir-660
42	hsa-let-7e|hsa-mir-125a|hsa-mir-99b
43	hsa-mir-1-2|hsa-mir-133a-1
44	hsa-mir-200b|hsa-mir-200a|hsa-mir-429
45	hsa-mir-192|hsa-mir-194-2
46	hsa-mir-17|hsa-mir-18a|hsa-mir-19a|hsa-mir-19b-1|hsa-mir-20a|hsa-mir-92a-1
47	hsa-mir-96|hsa-mir-182|hsa-mir-183
48	hsa-mir-30d|hsa-mir-30b
49	hsa-mir-133a-2|hsa-mir-1-1
50	hsa-mir-181a-2|hsa-mir-181b-2
51	hsa-mir-206|hsa-mir-133b
52	hsa-let-7f-2|hsa-mir-98

For the strong assumption regarding as the log-normality and equal variance for PASM we performed a normalization based on z-score transformation in R. The P-values were calculated against 1000 randomized samples.

6.   Cell culture and miRNAs transfection

MDA-MB-231, MCF7 and UACC-257 were cultured in liquid culture with Dulbecco's modified eagle medium (DMEM; GIBCO Laboratories, Grand Island, NY, USA) and RPMI supplemented with heat-inactivated 10% fetal bovine serum (FBS; Equitech-Bio, Kerrville, TX, USA) and a 1% antibiotic antimycotic solution (Invitrogen, CA, USA). The cells were maintained at 37°C in a humidified atmosphere with 5% CO_2. Mature microRNAs mimics from miR-500 were purchased from Ambion Inc. To estimate transfection efficiency, the negative control FAM (carboxyfluorescein) labeled RNA oligonucleotide (Ambion, TX, USA) were used. The oligonucleotides were individually transfected with Oligofectamine reagent (Invitrogen, CA, USA) at the final concentration of 100 $p$mol/well for a 12-well plate. After 48-h post-transfection, cells were harvested and analyzed.

7.   Microarray and data analysis

Two breast cancer cell lines (MDA-MB-231 and MCF7) were used for analysis of gene expression data after transfecting miR-500 mimic (Ambion, TX, USA) and control. Total RNA was extracted by using the mirVana miRNA Isolation kit

(Ambion, TX, USA). 750 ng of total RNA was used for labeling. Sample labeling was performed with an RNA amplification kit according to the manufacturer's instructions (Applied Biosystems; Foster City, CA). We used the HumanHT-12 v4 expression beadchip containing 48000 probes of 25000 annotated genes from Illumina Inc (San Diego, CA). After hybridization according to the manufacturer's protocols (Illumina®), the bead chip was scanned with a BeadArray Reader (Illumina®) and microarray data were log2-transformed and normalized using the quantile normalization method in the Linear Models for Microarray Data package in Bioconductor. Primary microarray data are available in the National Center for Biotechnology Information Gene Expression Omnibus public database (GSE61752). BRB-ArrayTools were used for statistical analysis of gene expression data (Simon et al. 2007). For class comparison between control and test samples treated by miR-500, the t-test was applied to identify the genes significantly different between two groups when compared using BRB ArrayTools (Qi et al. 2009). Gene expression differences were considered significant if $P <$ 0.001. Cluster analysis was performed using the software programs Cluster and Heatmap was generated by Treeview (Eisen et al. 1998).

8.   3'-UTR luciferase reporter assays

To evaluate miR-500 binding to the *PPFIA1* 3′-UTR, 74-bp oligonucleotides spanning the predicted 3′-UTR miRNA binding site flanked by *XhoI* and *NotI* restriction sites were cloned into pmirGLO Dual-Luciferase miRNA target Expression vector (Promega). Oligonucleotides with a mutated binding site were

used as control. Breast cancer and melanoma cells were seeded at a density of 10,000 cells per well in an opaque 6-well plate. Twenty-four hours after seeding, cells were co-transfected with a miR-500 pre-miR (Ambion) or negative control pre-miR (Ambion) in combination with the vectors. Forty-eight hours after transfection, luciferase reporter gene activity was measured using the Dual-Glo Luciferase Assay System (Promega) and a FLUOstar OPTIMA microplate reader (BMG LABTECH).

9. 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay

The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; Sigma-Aldrich) assay was performed as described previously (Mosmann 1983) to assess cell proliferation after mimic miR-500 treatment or control in MDA-MB-231 cells. Cells were plated into 12-well plates, transfected with mimic and control and incubated for 48 hrs. Cells were fixed in 10% formaldehyde and stained in 10% Crystal violet (Sigma-Aldrich) for 10 min at room temperature. The absorbance of individual wells was read using by a VMax kinetic microplate reader (Molecular Devices, Sunnyvale, CA, USA) at 570nm.

10.   shRNAs and lentiviral transduction.

Letiviral-based shRNA against human (Sigma-Aldrich, St. Louis, MO,), human
PPP2R5E and *PPFIA1*, and control shRNA targeting *eGFP* (Cat. No. SHC005V)
were used for knock-down experiments. Lentivirus particles were produced by
transient transfection of shRNA expression vectors along with packaging vectors
pLP1, pLP2, and pLP/VSVG (Invitrogen, Carlsbad, CA) in 293FT cells. The
lentiviral supernatants were collected at 48-72 hours post transfection and frozen
in aliquots. A moderate multiplicity of infection (MOI=3) was used for transduction
of cells to minimize negative effects on cellular proliferation. At 2 to 4 days after
infection, all experiments were performed. Reverse transcription polymerase
chain reaction (RT-PCR) was performed to quantify the mRNA of PPP2R5E and
PPFIA1 to examine the knock-down efficiency with the primers (**Table 10**).

Table 10. Primers for qRT-PCR

Gene	Forward	Reverse
PPFIA1	CCACATCTGTGCATGACCTC	TTCCAAGCGCTCCTGTAACT
PPP2R5E	GTATGTGGCTGTCAGCTCGT	TTGTGGCGTTGGTGTACAAT

11.   Western blot analysis

The culture cell samples were homogenized at 4°C in a protein lysis buffer. Equal amounts of total protein from each sample were resolved through a 10% SDS–PAGE gel and then transferred to a PVDF membrane (Perkin Elmer Life Scientific Inc.). The membranes were blocked with 5% non-fat dried skimmed milk powder solution for 1 h, and then incubated overnight at 4◦C with monoclonal or polyclonal antibodies for Phospho-*MAP2K1* (Ser217/221, 1:1000, Cell Signaling Technology), *MAP2K1* (1:1000, Cell Signaling Technology), and α-tublin (1:1000, Cell Signaling Technology). After washing with TBST (Tris-buffered saline with Tween; 20 mM Tris/HCl, pH 7.6, 150 mM NaCl and 0.05% Tween 20), the membranes were incubated with a secondary antibody against rabbit or mouse IgG. Then the membranes were washed, and protein was detected with an ECL (enhanced chemiluminescence) kit (Bio-Rad).

12.   Database implementation

We implemented a web-based miRNA-RPPA-pathway profiling system (miRPP, http://www.appex.kr/mirpp) based on our approach using a host language, JAVA (http://www.java.com). To provide user friendly and active interfaces, Google web toolkit (GWT, http://www.gwtproject.org/) and GWT extended (GXT, http://www.sencha.com/products/gxt) frameworks were used. The data transporting between client and miRPP server is controlled by GWT remote

procedure call (RPC) method. All statistical analysis methods of miRPP were implemented by R script language (http://www.r-project.org) with Bioconductor plugins (http://www.bioconductor.org). Calling R modules from a host language is managed by RCaller framework (https://code.google.com/p/rcaller). To store and handle the association scores and expression intensities, MySQL database management system was applied (http://dev.mysql.com). In addition, data query on MySQL from a host language is controlled by MyBatis, a XML based SQL mapping framework (https://code.google.com/p/mybatis). All services of miRPP are contained and served on an Apache Tomcat web server (http://tomcat.apache.org).

CHAPTER 3

CONSTRUCTION OF CORRELATION MATRIX OF MICRORNAS, MRNA, AND PROTEIN FEATURES IN NCI-60 CELL LINES.

1. Microarray Expression data sets

In this study, we used three different level of data sets in NCI-60 cells as the followings: mRNA array (45k probes), miRNA array (439 probes), RPPA (222 probes). **Figure.15** shows the overall expression pattern of each data set. We found the tissue specific signature of human tumors in the clustering analysis. For example, in the clustering of miRNA expression data we found the expression signatures depending on tumor types which were reported in the previous study (Olive et al. 2010) (**Fig15C**.).
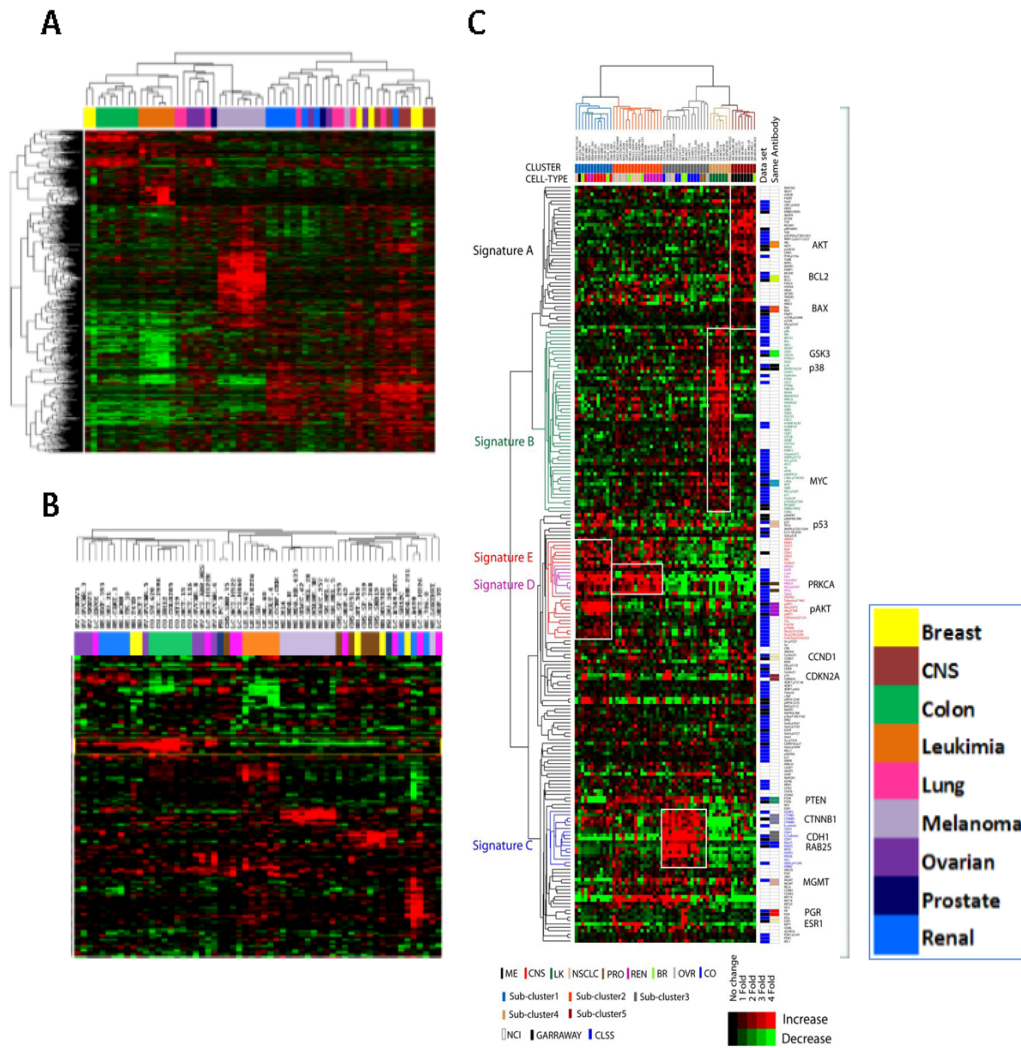
Figure 15. Overall expression pattern of omics data sets in NCI-60 cells. (A) mRNA array data, (B) miRNA array data, (C) RPPA data (Park et al. 2010). After variance filtering of probes with low standard deviation in mRNA data set, hierarchical clustering was performed. MicroRNA and RPPA data were clustered with all probes.

2. Construction of correlation matrix of miRNAs, mRNA, and protein features in NCI-60 cell lines.

For the development of algorithms, we selected data sets of miRNAs, mRNAs, and protein expression from NCI-60 cell lines rather than from other data sets because NCI-60 cell lines are the most extensively characterized lines (Bussey et al. 2006; Liu et al. 2010b; Zeeberg et al. 2012; Varma et al. 2014), and although data size is relatively small but they represent major tumor types well. More importantly, these cell lines can be quickly tested for validation of molecular functions predicted by algorithms.

A schematic overview of our approach is shown in **Figure.16**. Because miRNAs indirectly regulate expression and activation of proteins through targeting mRNAs, we tried to develop an analysis approach that would mimic biological process of regulatory circuits in cells.

In the correlation analysis between miRNA and mRNA data sets we found that top ranked miRNAs by the number of correlated mRNAs tends to be more tumor specific. For example, among these miRNAs in **Table.11** the previous study showed that mir-093-prec and mir-106, mir-107, mir-020, mir-099, and mir-017 were differentially expressed in tumors compared with non-tumorous tissues (Paranjape et al. 2009).

I. miRNA-Protein association

mRNA, miRNA, RPPA array data in NCI-60 cells

Correlation (in trans) analysis

ID mapping- Common gene set selection

Association score matrix

II. miRNA-Pathway association

P/U-ratio Gene effect weight

Pathway information

Pathway association score

Normalized Pathway association score matrix (Z-score, 1000 permutation)

miRNA cluster Information

III. Pathway enriched miRNA

miRNA cluster enrichment analysis

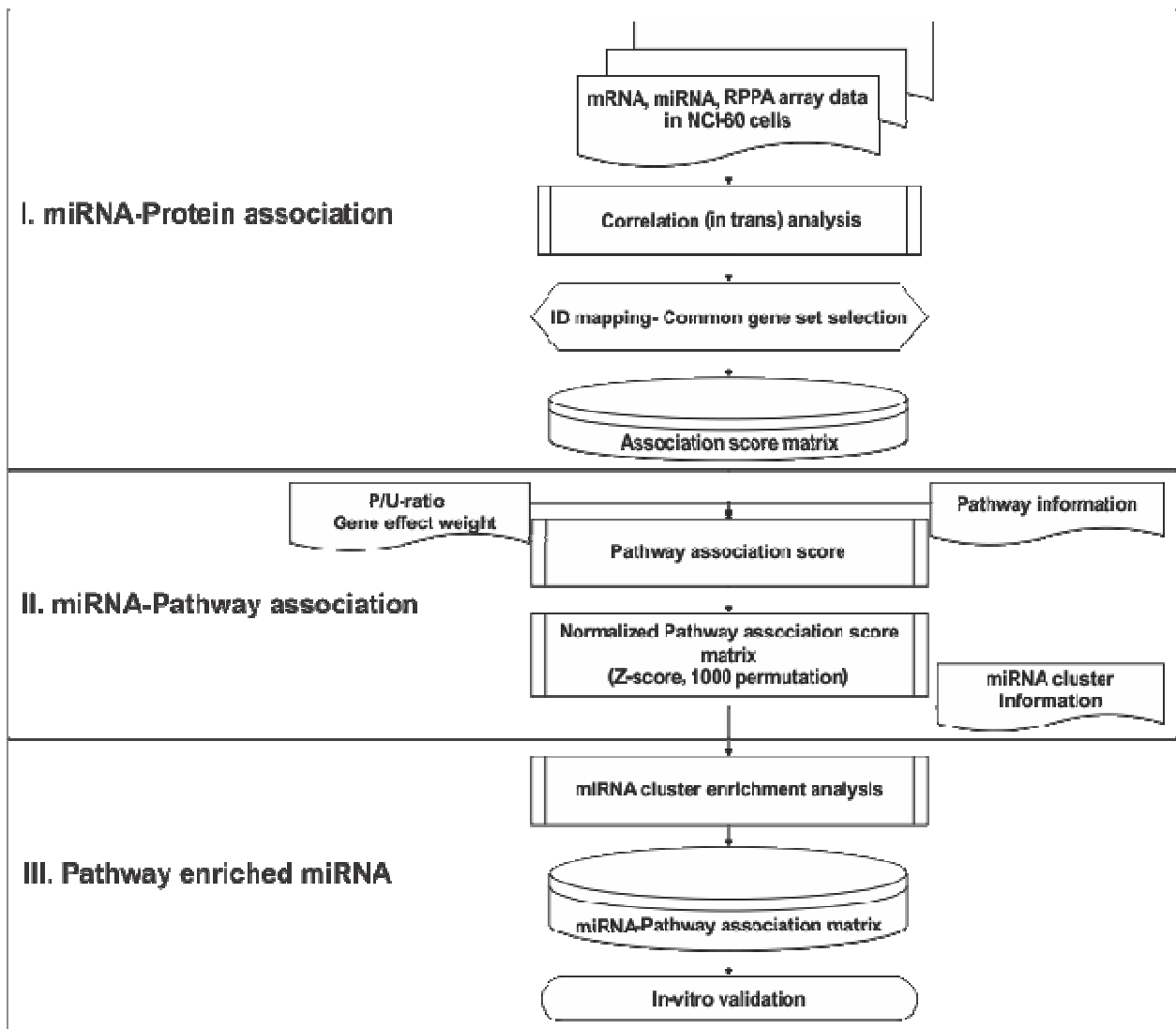miRNA-Pathway association matrix

In-vitro validation

Figure 16. Schematic overview illustrating data analysis process. Data analysis was carried out in three steps. In first step, two correlation matrices (miRNAs vs. mRNAs, and mRNAs vs. protein features) were generated by using Pearson correlation values. Two correlation matrices were integrated later by using mRNAs as denominators, generating association score between miRNAs and protein features. Resulted association scores represent number of commonly correlated mRNAs between miRNAs and protein features in matrix format. In second step, the P/U ratios (number of mRNAs associated with phosphorylated protein/number of mRNAs associated with unmodified protein) of each miRNA are mapped into signaling pathways and used to generate Pathway Association Score (PAS). PAS is later was normalized through *z*-score transformation with 1000 permutation. In last step, miRNA cluster enrichment analysis over signaling pathways was carried out with PASM and miRNA cluster information.

Table 11. Top 20 miRNA ranked by the number of correlated mRNAs

miRNA	# of correlated mRNA probes(%)*
mir-093-prec-7.1=093-1	1841(8.07)
mir-106bNo1	1579(6.92)
mir-125b-2-precNo2	1424(6.24)
mir-106-prec-X	1357(5.95)
mir-142-prec	1337(5.86)
mir-125b-1	1327(5.82)
mir-020-prec	1268(5.56)
mir-106aNo1	1217(5.34)
mir-100No1	1216(5.33)
mir-025-prec	1208(5.3)
mir-509No1	1180(5.17)
mir-032-precNo2	1168(5.12)
mir-20bNo1	1138(4.99)
mir-099-prec-21	1117(4.9)
mir-200cNo1	1099(4.82)
mir-125a-precNo1	1065(4.67)
mir-023a-prec	1053(4.62)
mir-018-prec	1042(4.57)
mir-017-precNo2	1021(4.48)
mir-513-2No1	1018(4.46)

*p-val < 0.005

First, we generated a correlation matrix between expression of miRNAs and mRNAs by using expression data from NCI-60 cell lines (Liu et al. 2010b); 302,712 pairs were significantly correlated ($P < 0.005$). Second, we generated a correlation matrix between expression of mRNAs and protein features (expression and phosphorylation) (Zeeberg et al. 2012; Varma et al. 2014) two correlation matrices were integrated together by using mRNAs as denominators connecting miRNAs to protein features, reflecting indirect regulation of proteins by miRNAs. Briefly, significantly correlated mRNAs for a particular miRNA were cross-compared with significantly correlated mRNAs for all proteins to generate association scores of the miRNA across all proteins. Thus, association scores represent number of commonly correlated mRNAs between miRNAs and protein features in matrix format (**Fig. 17**).
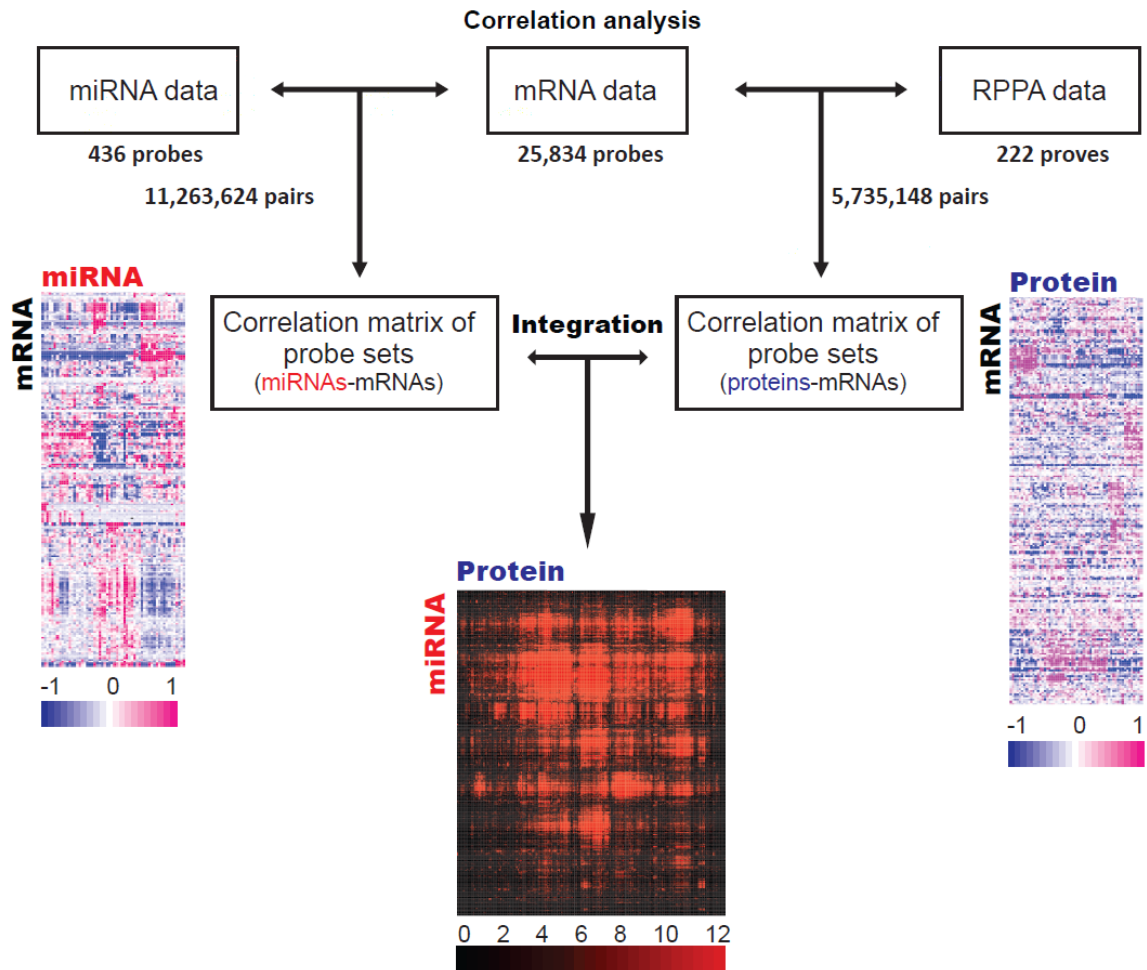
Figure 17. Schematic overview illustrating data integration. Pearson's correlation coefficient analysis was performed to generate two correlated pair sets; miRNA-mRNA and RPPA-mRNA pairs. In correlation analysis, *P*-value less than 0.005 was considered to be statistically significant. To establish the association between miRNA and protein, two independently generated correlation matrices were integrated by counting correlated mRNAs as denominators. Association score represents the log2 transformed number of commonly correlated mRNAs with the both miRNA and RPPA probe.

To explore the distribution of ASM we made a density plot with all association scores between miRNAs and RPPA proteins. It showed a Gaussian distribution over 2 (**Fig.18**). When using ASM integrated together by counting correlated mRNAs between miRNA and RPPA (cut-off >7), we found one of miRNA clusters representing miR-200 cluster (**Fig.19**). This cluster showed the high association with E-cadherin, AKT, PI3K and other proteins. The previous studies showed MiR-200 family play important role in EMT pathway including E-cadherin and also in AKT pathway (Cancer Genome Atlas Research 2014).
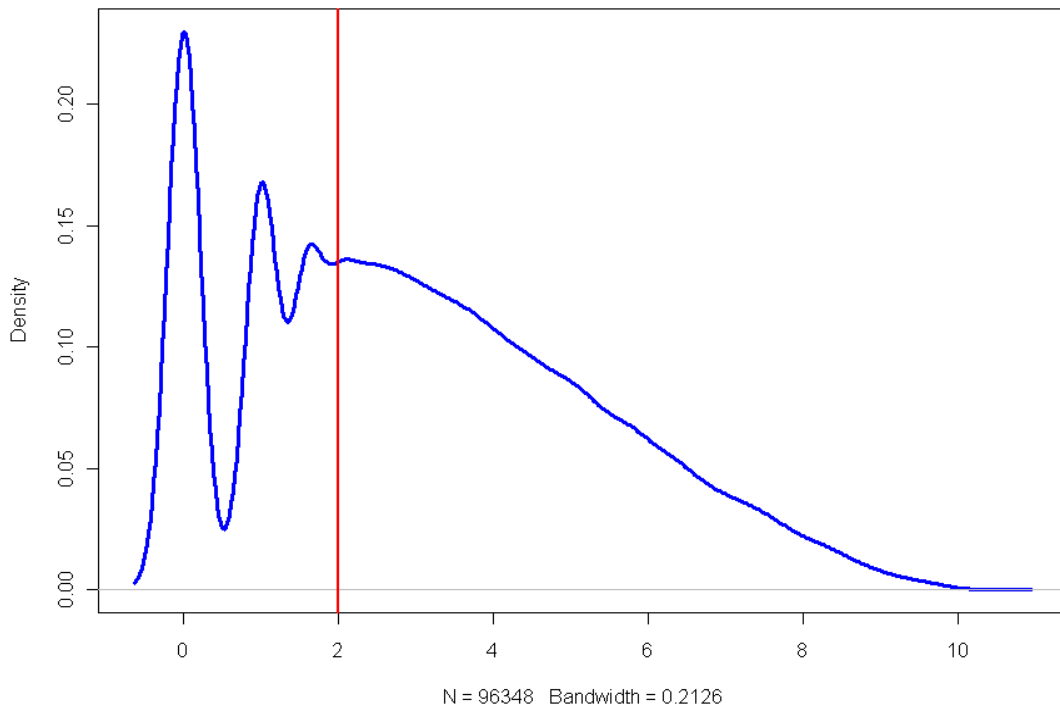
Cut-off(log2)	# of miRNA-RPPA pairs	%
>2	48,242	50.07
>3	36,524	37.91
>4	25,672	26.65
>5	16,654	17.29
>6	9,643	10.01
>7	4,831	5.01
>8	1,809	1.88

Figure 18. Density plot of Association scores. X-axis represents log2 based values. The table shows the number of significantly correlated miRNA-RPPA pairs depending on the cut-off.
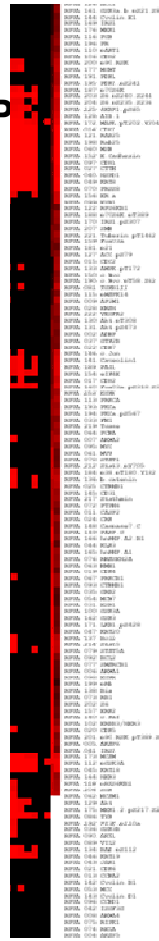
Figure 19 . miR-200 ~~is~~ ~~~~~~~~~~~ A proteins in ASM.
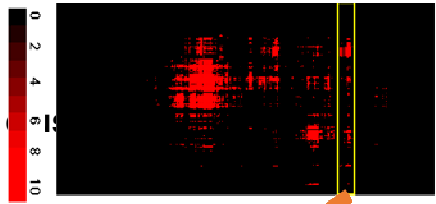
Signaling events of pathways are best reflected in modification of signaling proteins such as for phosphorylation (Karin and Hunter 1995). RPPA has been used for the comprehensive analysis of protein expression levels and activation status in signaling pathways through measuring phosphorylation status of gene products (Tibes et al. 2006). Using information from antibody pairs that can recognize unmodified form or phosphorylated form of the same protein, we assessed the significance of phosphorylation by identifying the number of genes whose expression patterns are significantly correlated with the degree of phosphorylation over expression of the protein.

Because signaling events of pathways are best reflected in modification of signaling proteins like phosphorylation and vast majority of downstream effectors of many signaling pathways are transcription regulators, we hypothesize that cellular signaling activity would be well reflected in number of mRNAs whose expression patterns are correlated with signaling events like phosphorylation of signaling proteins. Therefore, we tried to estimate signaling activities in NCI-60 cells by computing correlated number of mRNAs with phosphorylation of signaling proteins. RPPA used 25 antibody pairs that recognize unmodified or phosphorylated form of signaling proteins (**Table 1**). By selecting out 25 pairs of correlation data between mRNAs and protein features from correlation matrix, we generated P/U ratios (number of mRNAs associated with phosphorylated protein/number of mRNAs associated with unmodified protein) as indicator of signaling activity of signaling proteins in NCI-60 cell lines.

To correlate miRNAs to signaling activity of 25 signaling proteins, we next generated matrix between miRNAs and P/U ratios of each proteins (436 miRNAs x 25 P/U ratios of signaling proteins) by using association scores in ASM from previous analysis. Because 25 signaling proteins are involved in regulation of 40 signaling pathways (**Table 12**), we computed the sum of P/U ratios of the proteins in each signaling pathway per each miRNA, generating Pathway Association Score (PAS) that well reflected the association of miRNAs to signaling pathways and the matrix of PAS (PASM). Significance of miRNAs associated with signaling pathways was estimated by enrichment analysis using association values in PASM and miRNA genome cluster information because miRNAs in the same genomic cluster share target mRNAs and biological function (Gurtan and Sharp 2013). This analysis yielded functionally matched list between signaling pathways and miRNA clusters (**Table 3**). Some mRNA clusters seem to be involved in regulation of many signaling pathways. For example, miRNAs in cluster 5 (hsa-mir-127, -136, -370, -337, -431, -433, -493, -432, -770 and -665) showed frequent association with multiple signaling pathways. Interestingly, substantial fraction of signaling pathways associated with cluster 5 is immune-response pathways, suggesting that miRNAs in cluster 5 might be important regulators of immune systems. For instance, a previous study reported that hsa-miR-127 is involved in B-cell differentiation process (Leucci et al. 2010).

Table 12. MicroRNA cluster enriched pathway

Pathway	DB	Cluster	COUNT	Z-SCORE	P-VALUE	q-VALUE
PELP1 pathway	BIOCARTA	35	14	4.03653	5.43E-05	0.000762
FC EPSILON RI signaling pathway	KEGG	5	14	3.55105	0.000384	0.01867
T-CELL RECEPTOR signaling pathway	KEGG	5	14	3.55105	0.000384	0.01867
CREB pathway	BIOCARTA	5	14	3.52968	0.000416	0.018861
BIOPEPTIDES pathway	BIOCARTA	5	14	3.48198	0.000498	0.015912
ADIPOCYTOKINE signaling pathway	KEGG	46	7	3.31242	0.000925	0.028157
HCMV pathway	BIOCARTA	5	14	3.26827	0.001082	0.0326
TOLL LIKE RECEPTOR signaling pathway	KEGG	5	14	3.26827	0.001082	0.0326
CHEMOKINE signaling pathway	KEGG	5	14	3.26103	0.00111	0.02751
ERK pathway	BIOCARTA	41	9	3.21621	0.001299	0.050206
NFAT pathway	BIOCARTA	5	14	3.1846	0.00145	0.043812
PROGESTERONE MEDIATED OOCYTE MATURATION	KEGG	5	14	3.13745	0.001704	0.053421
BIOPEPTIDES pathway	BIOCARTA	2	38	3.07015	0.002139	0.034201
FC GAMMA R MEDIATED PHAGOCYTOSIS	KEGG	5	14	3.05116	0.00228	0.080329
NEUROTROPHIN signaling pathway	KEGG	5	14	2.99409	0.002753	0.142044
HER2 pathway	BIOCARTA	41	9	2.90519	0.00367	0.041129
MTOR pathway	BIOCARTA	47	6	2.90323	0.003693	0.034864
TPO pathway	BIOCARTA	4	3	2.89114	0.003838	0.199599
PDGF pathway	BIOCARTA	4	3	2.89114	0.003838	0.199599
EGF pathway	BIOCARTA	4	3	2.89114	0.003838	0.199599
IL4 pathway	BIOCARTA	7	3	2.83008	0.004654	0.030358
EIF4 pathway	BIOCARTA	5	14	2.81424	0.004889	0.12102
ADIPOCYTOKINE signaling pathway	KEGG	26	4	2.80944	0.004963	0.053712
VEGF signaling pathway	KEGG	5	14	2.80882	0.004972	0.053927
VEGF signaling pathway	KEGG	23	2	2.69258	0.00709	0.053927
EDG1 pathway	BIOCARTA	5	14	2.67643	0.007441	0.172658
MAPK signaling pathway	KEGG	5	14	2.63849	0.008328	0.216517
ERK pathway	BIOCARTA	4	3	2.58685	0.009686	0.187181
CHEMOKINE signaling pathway	KEGG	23	2	2.52562	0.01155	0.090542
IGF1R pathway	BIOCARTA	5	14	2.52197	0.01167	0.219031

Top 30 overrepresented pathways with p < 0.02 The pathway and cluster are sorted by

p-value of Parametric gene set enrichment analysis

1. Web-based Database system, miRPP

Based on our approach we implemented a web-based tool, miRNA-RPPA-pathway profiling system (miRPP) (**Fig.20**). miRPP provides the following information against a query miRNA in NCI-60 data sets: Overall expression of significantly correlated mRNAs and RPPA protein probes, mRNA mediated Association Score Matrix between miRNAs and RPPA probes, and miRNA associated cell signaling pathways.

Figure 20 . miRNA-RPPA-Pathway Association Profiler (miRPP). miRPP provides the following information against a miRNA in NCI-60 data sets: Overall expression of significantly correlated mRNAs and RPPA protein probes, mRNA mediated Association Score Matrix between miRNAs and RPPA probes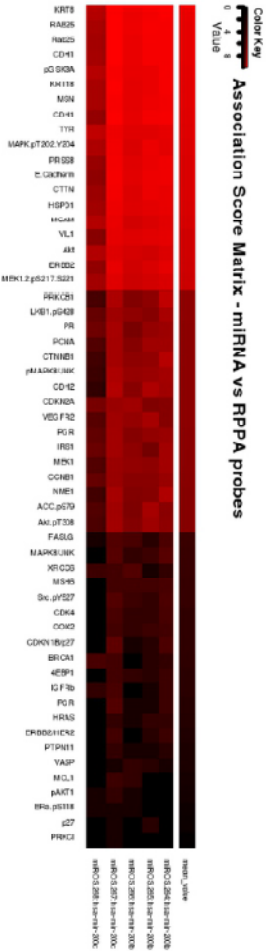, a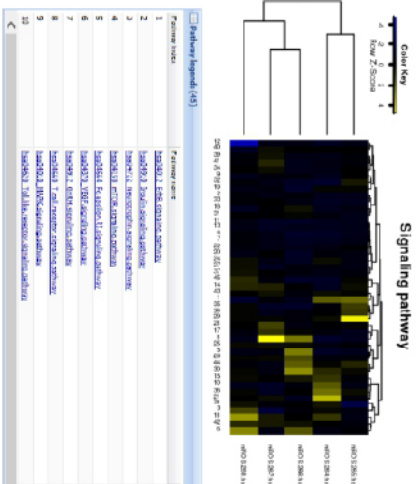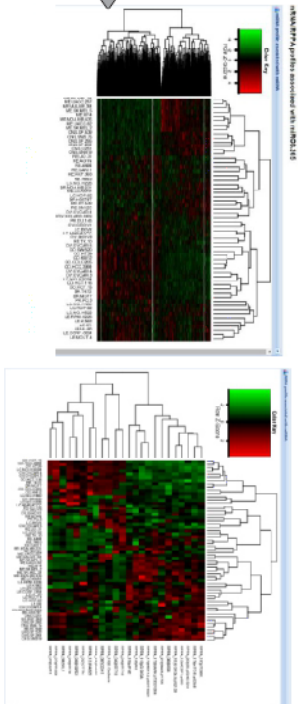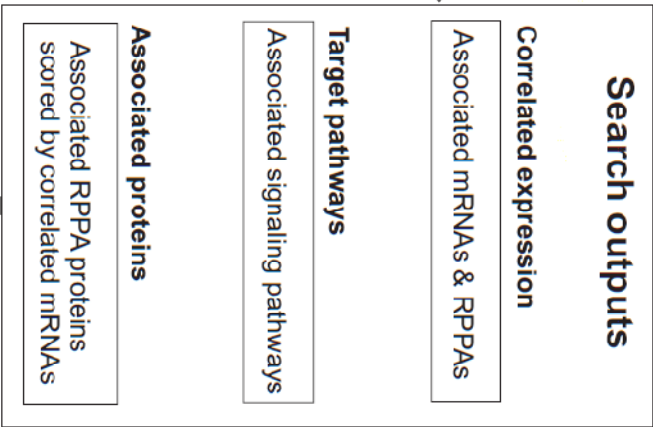nd miRNA associated cell signaling pathways. It was implemented by JAVA. For user friendly and active interfaces, Google web toolkit (GWT) and GWT extended (GXT) frameworks were used. RPPA: Reverse phase protein array.

2. **Biological and clinical significance of miR-500**

Because MAPK pathway is critical for regulation of cell growth and cross-talk with many other signaling pathways (De Luca et al. 2012; Deschenes-Simard et al. 2014), we selected miRNAs (cluster 41; hsa-mir-188, -362, -500, -501, -502, -532, and -660) significantly associated with MAPK pathway for further functional validation (**Table.13**). We found strong association between miR-500 cluster and proteins in MAPK pathway (**Fig.21**). Out of 7 miRNAs in cluster 41, we further selected miR-500 because its expression was most significantly associated with activation of MAPK pathway as judged by correlation between its expression and phosphorylation of *MAP2K1*, a key upstream regulator of *MAPK1* (Sebolt-Leopold 2000; Chowdhury et al. 2014; Neuzillet et al. 2014). We also found that *MAP2K1* had high score of P/U ratio against miR-500 (**Fig. 22**). Among miR-500 cluster members, the phosphorylation of *MAP2K1* was significantly correlated with miR-500 expression in NCI-60 cells (**Fig.23**).

Table 13. Hsa-mir-500 information from miRBase*

Accession	MI0003184
Previous IDs	hsa-mir-500
Symbol	HGNC:MIR500
Description	Homo sapiens miR-500a stem-loop
Gene family	MIPF0000139; mir-500
Genome context	Coordinates (GRCh38) chrX: 50008431-50008514 [+]
Stem-loop	<pre> c c −u uac ug agag ugu gcuc c cucuc aauccuugc c ggug ugc c |||| | ||||| ||||||||| | |||| ||| u cgag g gagag uuaggaacg g ccac acg g a c uc --- gu -gua uaa</pre>
Clustered miRNAs	< 10kb from hsa-mir-500a > hsa-mir-532 chrX: 50003148-50003238 [+] hsa-mir-188 chrX: 50003503-50003588 [+] hsa-mir-500a chrX: 50008431-50008514 [+] hsa-mir-362 chrX: 50008964-50009028 [+] hsa-mir-501 chrX: 50009722-50009805 [+] hsa-mir-500b chrX: 50010672-50010750 [+] hsa-mir-660 chrX: 50013241-50013337 [+] hsa-mir-502 chrX: 50014598-50014683 [+]

*http://www.mirbase.org
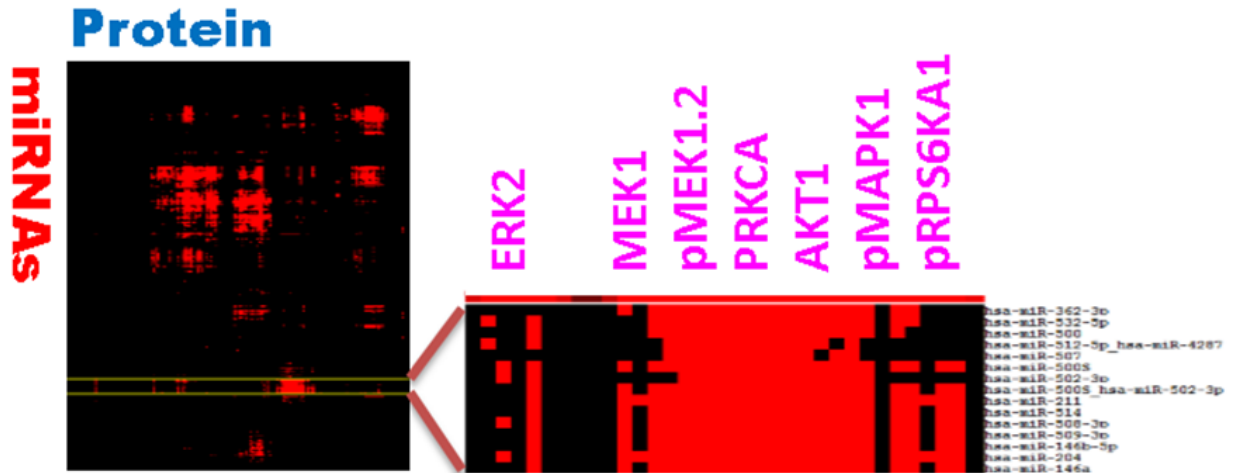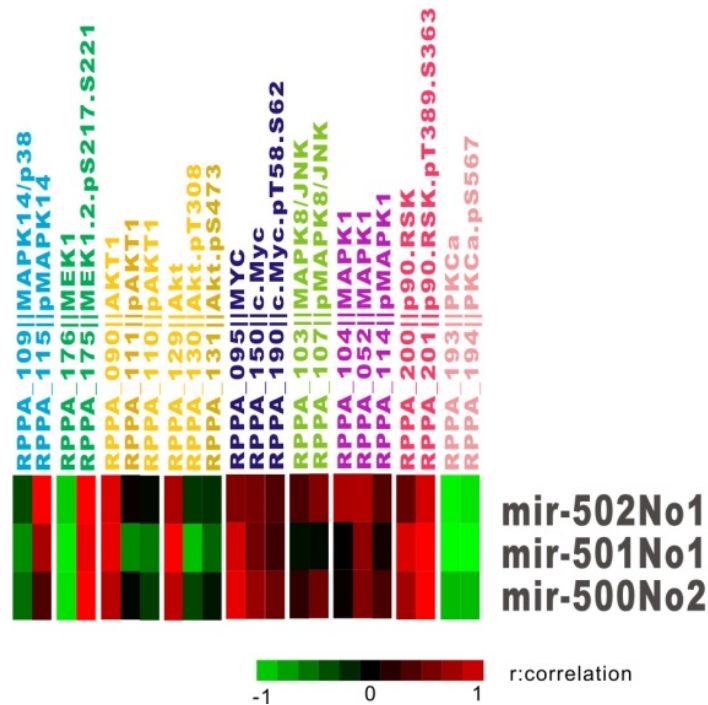
Figure 21. **Association between miR-500s and the proteins in MAPK pathway**. (A) Heatmap of ASM against miR-500s. (B) Direct correlation of expression data between miR-500s and proteins in MAPK pathway. MEK1/2 (official symbol: MAP2K1)
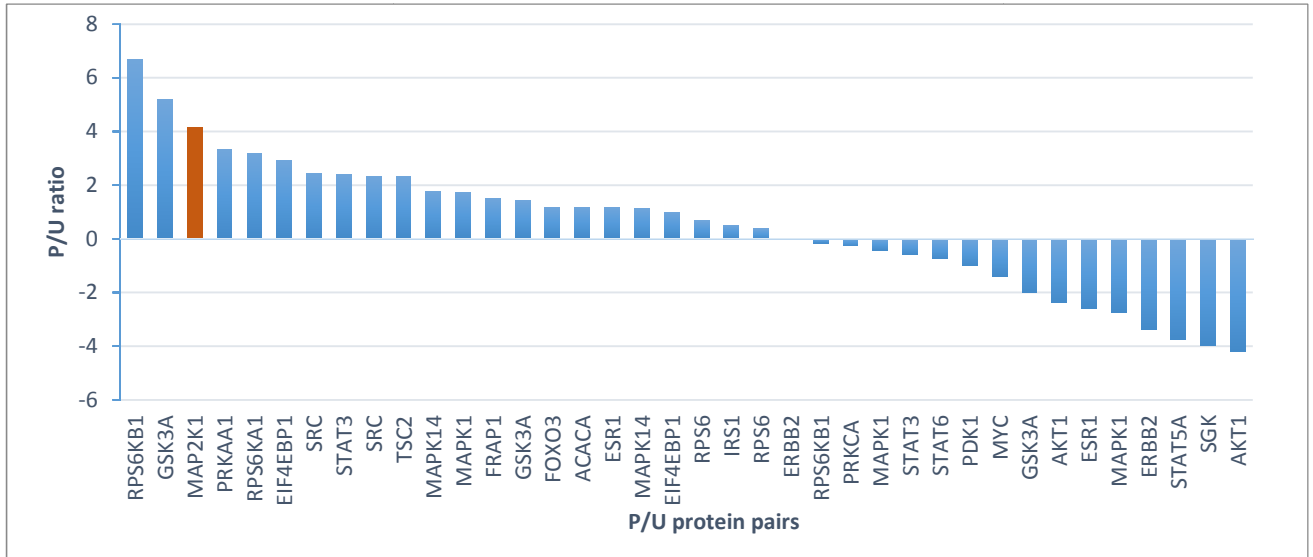
Figure 22. P/U ratio of the protein pairs against miR-500.
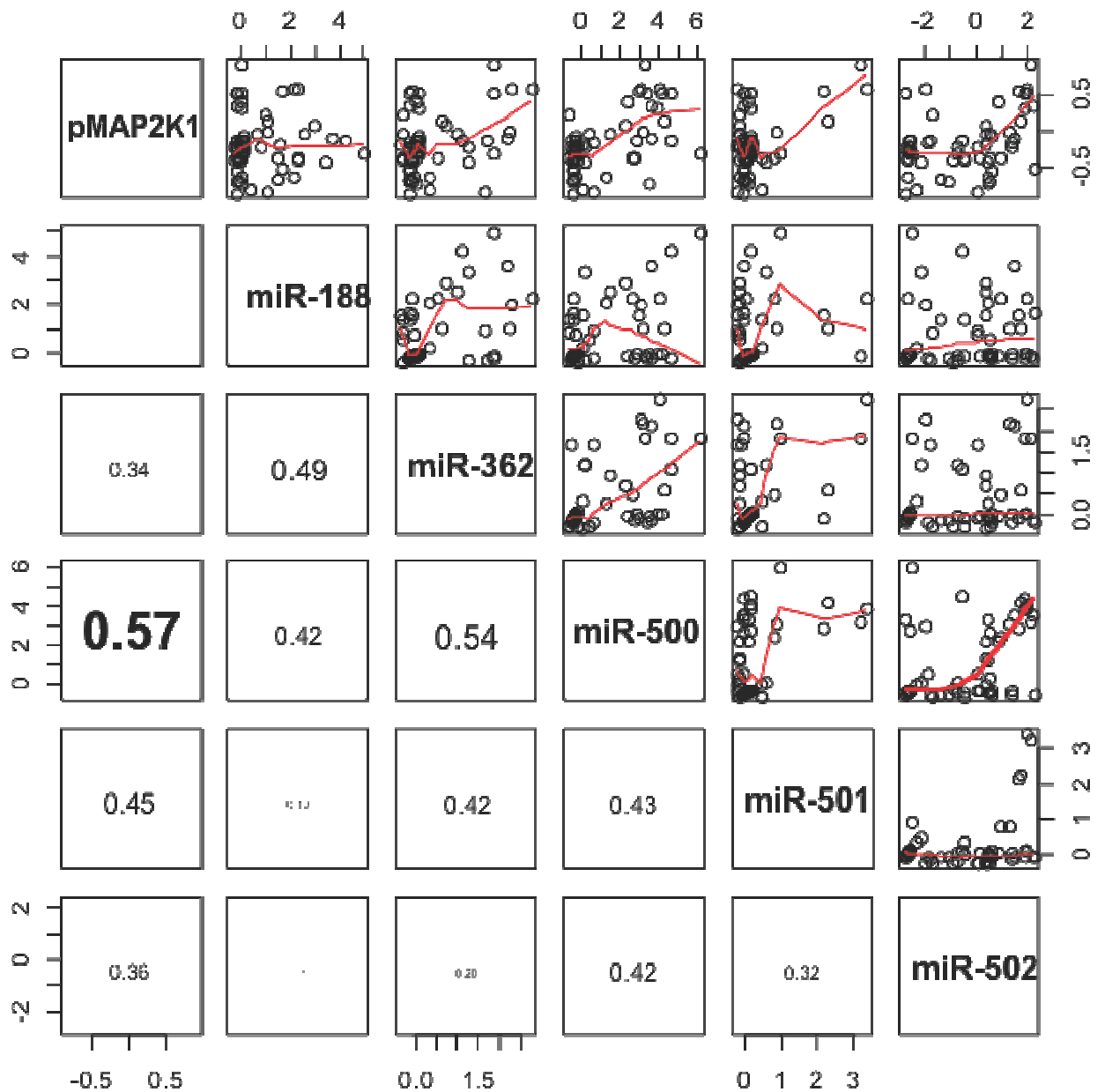
Figure 23. Association between miRNAs in cluster 41 and phosphorylation of MAP2K1in NCI-60 cells. Pearson correlation approach was used to generate correlation matrix between miRNAs in cluster 41 and phosphorylation level of MAP2K1 in NCI-60 cells. Numbers in boxes represent correlation and size of number reflect strength of correlation. pMAP2K1, Phosphorylated MAP2K1.

3. Chloride channel protein 5 (CLCN5)

CLCN5 gene is a member of the CIC family of chloride ion channels and ion transports. It is localized to endosomal membranes. Its mutations induce Dent disease and renal tubular disorders complicnephrolithiasis, and functions as antiport system and exchanges chloride ions against protons (Akbani et al. 2014). CLCN5 also is localized in Golgi apparatus membrane (http://www.genecards.org) and highly expressed in Kidney tissue, moderately in aortic vascular smooth muscle and endothelial cells, and a little higher in the coronary vascular smooth muscle (**Fig.24**).
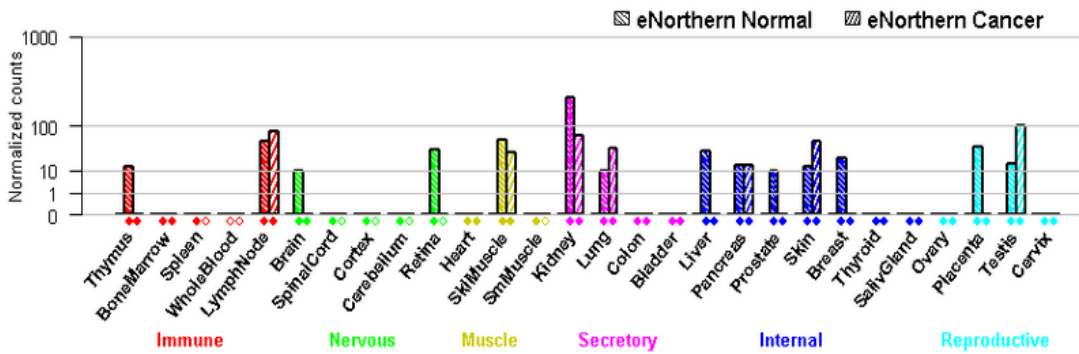
Figure 24. mRNA expression of CLCN5 in nomal and cancer tissues

Interestingly, expression of miR-500 is significantly higher in melanoma and breast cancer cell lines (**Fig. 25A**), suggesting potential roles of miR-500 in regulation of MAPK pathway in these cancer types. Since miRNA cluster 41 is co-localized with *CLCN5* gene and expressed as part of *CLCN5* transcript, we assessed correlation of expression of miRNAs in cluster 41 and *CLCN5* in NCI-60 cell lines (**Fig. 25B**). Interestingly, correlation was highest between miR-500 and *CLCN5*. Because expression of *CLCN5* is highly correlated with expression of miR-500, we next assessed potential clinical relevance of miR-500 in melanoma by using mRNA expression data of tumor tissues from patients with melanoma (GSE19234) (Lorenzi et al. 2009). When patients were dichotomized according to expression of *CLCN5*, high expression of *CLCN5* was significantly associated with shorter overall survival (**Fig. 25C**), strongly suggesting that miR-500 might play roles as oncogene presumably by activating MAPK pathway in melanoma.
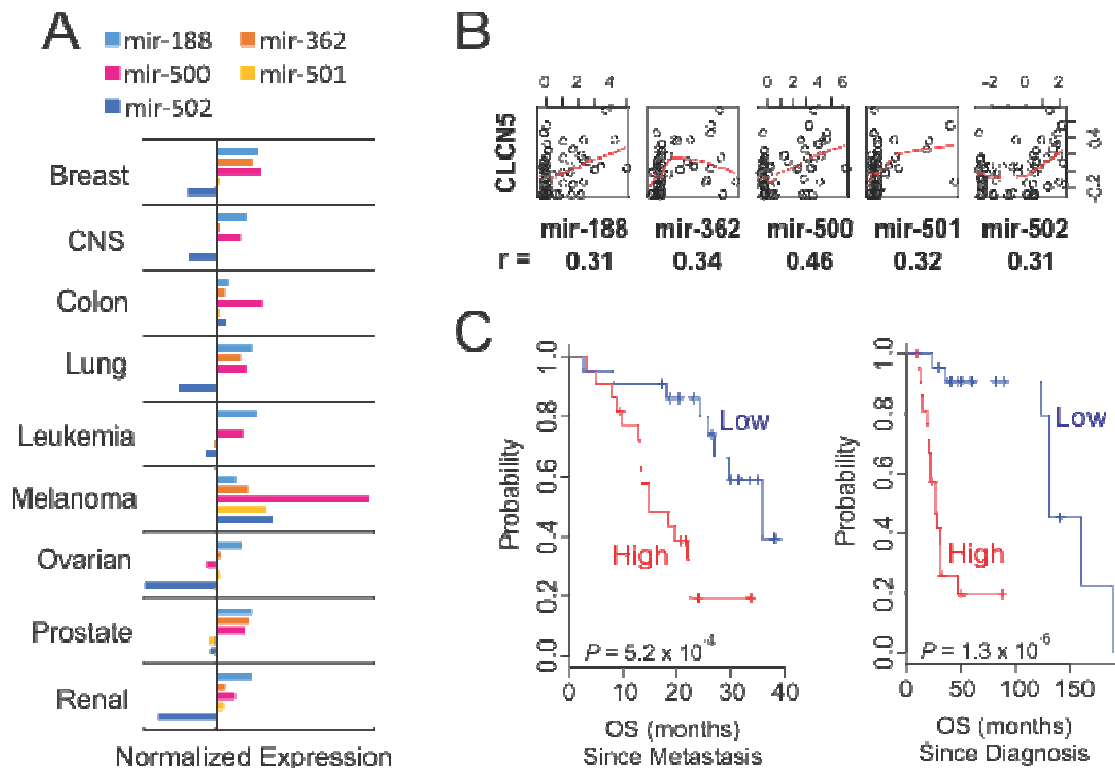
Figure 25. miR-500 is significantly associated with prognosis of patients with melanoma. (A) Average expression of miRNAs in cluster 41 in NCI-60 cell lines. Among 7 miRNAs in cluster 41, probes for 5 miRNAs were available in expression data. The expression values are normalized and averaged in each tissue. (B) Correlation analysis between miRNAs in cluster 41 and the mRNA expression of *CLCN5* in NCI-60 cell lines. (C) Kaplan-Meier plots of overall survival of patients dichotomized according to the expression of *CLCN5*.

CHAPTER 5

IN-VITRO VALIDATION OF PREDICTED TARGETS OF MIR-500

1. Hypothesis and experimental design

In-silico analysis we identified that miR-500 expression is significantly associated with MAPK pathway and positively correlated with the phosphorylation of *MAP2K1*. Therefore, we hypotheses when miR-500 mimic is treated into cells the phosphorylation level of *MAP2K1* will be increased (**Fig.26**).

- **Control**
- **Mimic miR-500**

MB231, MCF7 or UACC257

Phosphorylation level of MAP2K1 (S217.S221) ?
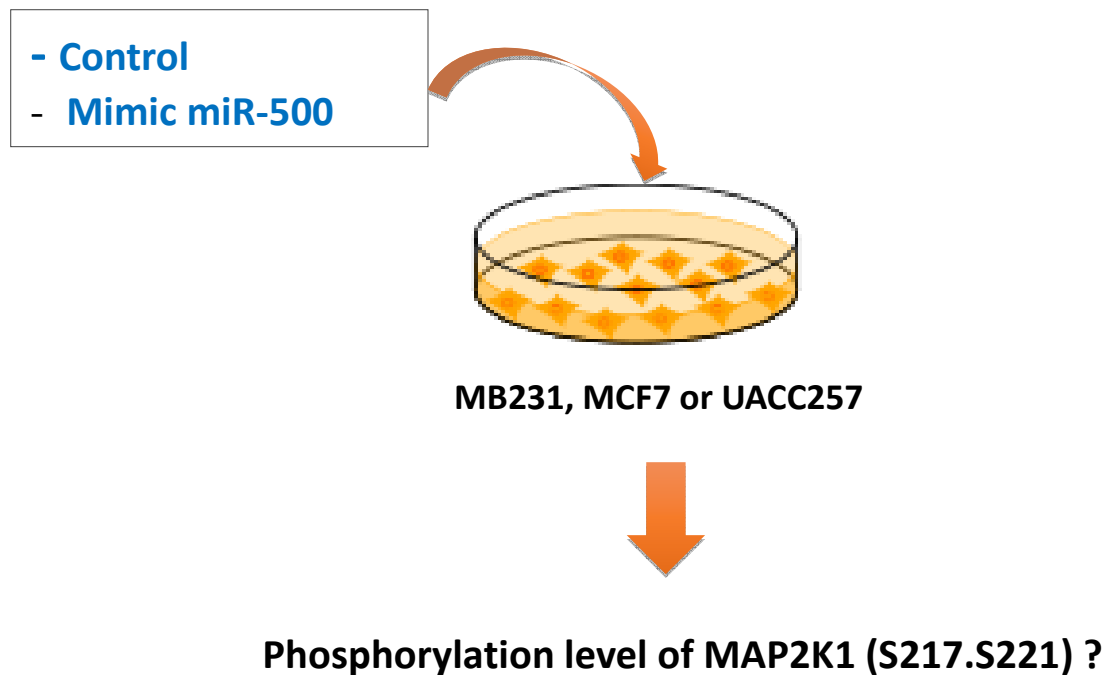
Figure 26. Experimental design to examine the increase of MAP2K1 phosphorylation by miR-500 in melanoma and breast cancer cells.

2. **Screening direct targets of miR-500 through In-vitro assays**

Because our analysis suggested significant correlation of miR-500 with MAPK pathway, we hypothesized that miR-500 regulates *MAP2K1* in melanoma cells. To do this, we transfected miR-500 mimic into UACC-257 melanoma cell and measured phosphorylation of *MAP2K1*, reflecting activation of *MAP2K1* (Mammano et al. 2012; Federici et al. 2013) (**Fig.27**). In good agreement with our hypothesis, *MAP2K1* was activated upon transfection of miR-500 mimic as evidenced by increased phosphorylation of *MAP2K1* (**Fig. 28A**). To verify that the effect was not specific to melanoma cells and was occurred independently of *BRAF* mutation status, we further tested the ability of miR-500 to activate *MAP2K1* in two MCF7 and MDA-MB-231 breast cancer cell lines. MCF7 cells do not have *BRAF* mutation. Consistent with result from UACC-257, we observed activation of *MAP2K1* when miR-500 mimic were transfected into MCF7 and MDA-MB-231 cells, strongly suggesting that miR-500 regulates MAPK pathway by activating *MAP2K1*. In addition, cell growth was significantly increased upon introduction of miR-500 mimic to MDA-MB-231 cells (**Fig. 28B**), further supporting that MAPK pathway is really activated.

NC miRNA-FAM^{TM} : Negative Control FAM (carboxyfluorescein) labeled RNA oligonucleotide
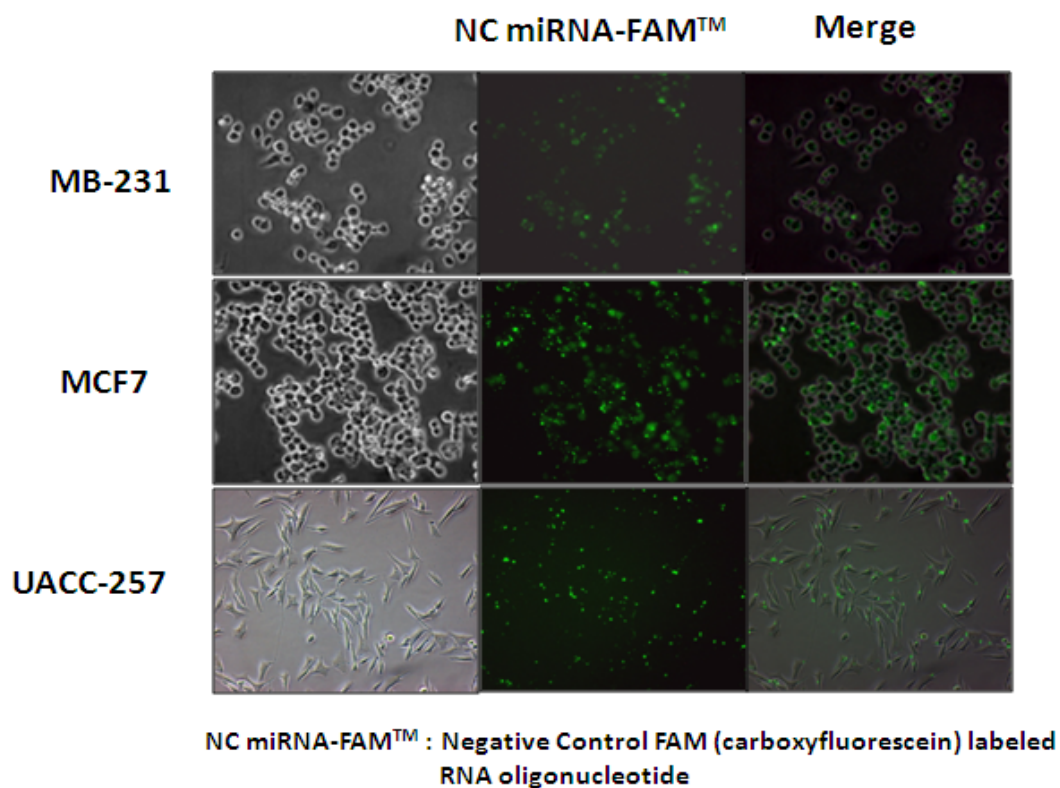
Figure 27. Representative photographs of Negative control miRNA with FAM^{TM}. Control miRNAs were transfected under light (left), epifluorescence (mid), overlay (right) microscopy in MB-231, UACC-257 and MCF7 cells.
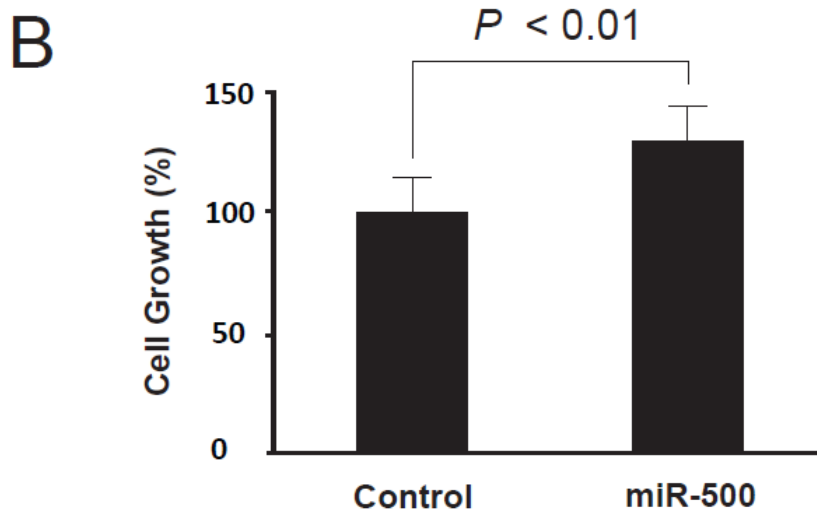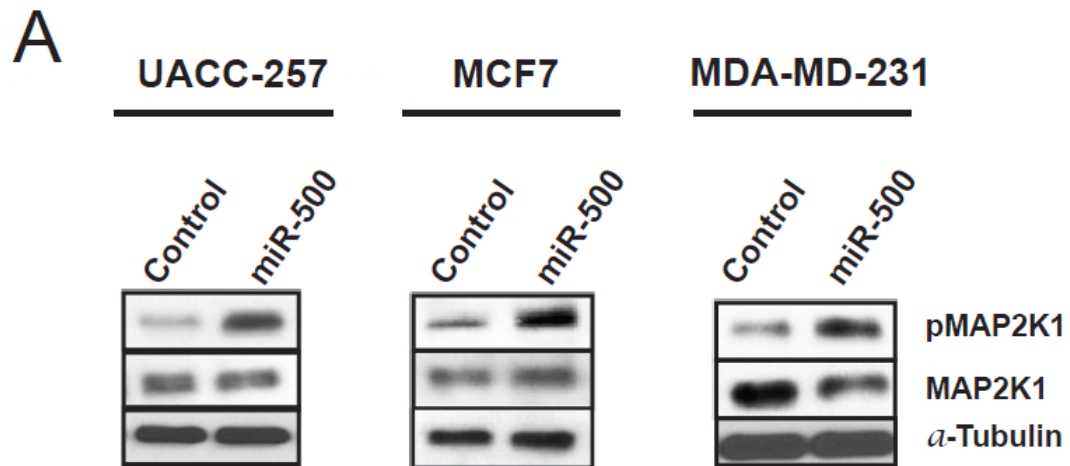
Figure 28. miR-500 regulates phosphorylation of MAP2K1. (A) Western blots for expression and phosphorylation of *MAP2K1* in miR-500 transfected melanoma and breast cancer cells. α-Tubulin was used as loading control. (B) miR-500 significantly increased proliferation of MDA-MB-231 cells. MTT proliferation assay was carried out 48 hours after treating cells with miR-500 mimic or control ($P < 0.01$ by Student t-test).

To determine the molecular mechanism by which miR-500 regulates *MAP2K1*, we performed microarray experiments after transfecting miR-500 mimic into MCF7 and MDA-MB-231 breast cancer cells. Because MAPK pathway can be activated in UACC-257 melanoma cells with a BRAF mutation, we used breast cancer cells to identify more direct downstream effect on the *MAP2K1* phosphorylation by miR-500. Then microarray experiment was performed with extracted RNA.

Analysis of gene expression data revealed that expression of 719 genes was significantly changed in both cell lines ($P < 0.001$, Two sample t-test). Because miRNAs negatively regulates mRNAs by directly targeting them, we further selected genes whose expression is commonly down-regulated by miR-500 (**Fig. 29**). Of 719 genes, 447 genes were commonly downregulated by miR-500 in the both cells. Because miR-500 activated *MAP2K1* while miRNAs negatively regulates their targets, we reasoned that targets of miR-500 for regulation of *MAP2K1* might be negative regulators of *MAP2K1*. Phosphatases are best known negative regulators of kinase-mediated signaling pathways (Roskoski 2012). Therefore, we further selected mRNAs of phosphatase families with predicted target sequence of miR-500 as candidates for key regulators by using starBase database providing miRNA-target interactions (Yang et al. 2011; Chou et al. 2013). There are only two phosphatases or its binding proteins (*PPP2R5E* and *PPFIA1*) with target sequence of miR-500 among down-regulated genes by miR-500. We performed the protein-protein interaction

network analysis related to two putative targets with STRING (http://string-db.org)

(**Fig.30**).
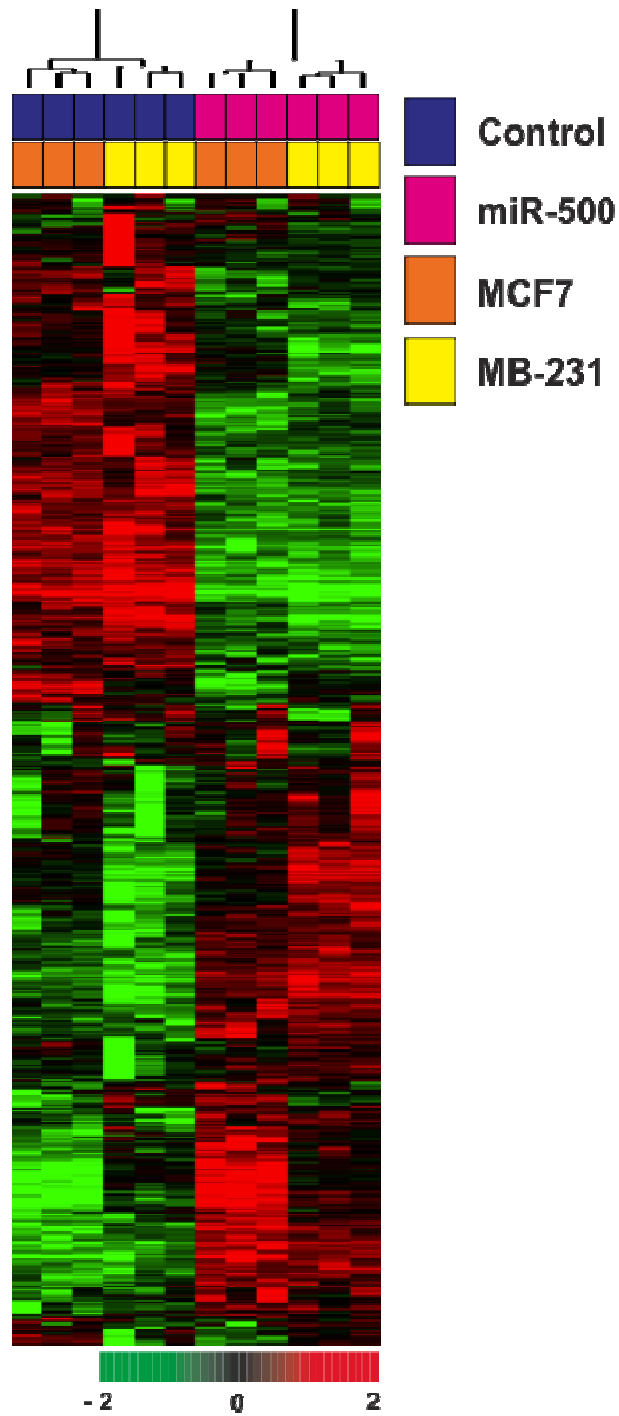
Figure 29. Gene expression of MDA-MB-231 and MCF7 cells after the transfection of mimic miR-500 or control. Gene expression was used for hierarchical clustering after filtering genes with low SD. The data were centralized by subtracting median expression level across samples for clustering analysis. 3047 genes were presented in the data set. The data are presented in matrix format in which rows represent individual gene and columns represent each sample. Each cell in the matrix represents the expression level of a gene feature in an individual sample. The red and green color in cells reflect relative high and low expression levels respectively as indicated in the scale bar (log2 transformed scale).
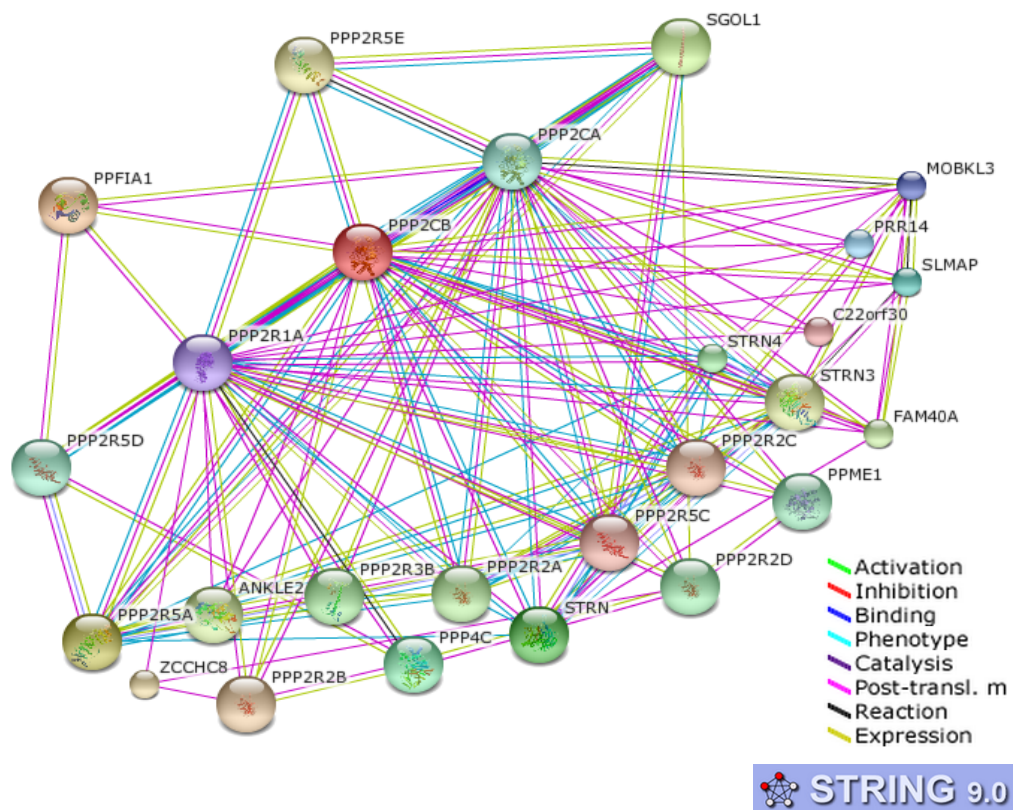
Figure 30. Protein-Protein Interaction network(PPI) of phosphatase associated

proteins from the microarray data analysis. STRING is the database for Known

and predicted protein-protein interactions.

To test if they are involved in regulation of *MAP2K1*, we silenced their expression in MCF7 and UACC-257 cells and measured phosphorylation of *MAP2K1*. Silencing of *PPFIA1* activated *MAP2K1* in both cell lines whereas silencing of PPP2R5E had no effect on *MAP2K1* activity (**Fig. 31B**), suggesting that PPFIA1 might be negative regulator that is regulated by miR-500 for activation of MAPK pathway. To determine whether miR-500 directly targets the 3'UTR of *PPFIA1* mRNA, we used a luciferase reporter vector containing the 3'UTR sequence of *PPFIA1*, including the predicted binding site for miR-500 in UACC-257 cells. Luciferase activity was significantly inhibited by the *PPFIA1* 3'UTR sequence when only miR-500 were co-transfected (**Fig. 31D**). However, luciferase activity was not inhibited by a mutant 3'UTR sequence (**Fig. 31C and 31D**), strongly demonstrating that miR-500 directly targets the 3'UTR sequence of *PPFIA1* mRNA and inhibits the expression of *PPFIA1*.

A

Control vs. miR-500

1555   719   1393

MDA-MB-231   MCF7

Down-regulated
genes (447)

B

■ Control   ■ shRNA1   □ shRNA2

PPFIA1   PPP2R5E

pMAP2K1

α-Tubulin

C

PPFIA1

miR-500   3' AGAGUGGGUCCAUCG**UUCCUAA**G 5'
                               | | | | | | |
WT 3'UTR   5' AAACCAUGACA**AAGGAUU**U 3'
Mutant 3'UTR   5' AAACCAUGACA**AAGCTCU**U 3'
                                 Mutation Sites

D

■ Wild type   □ Mutant

97
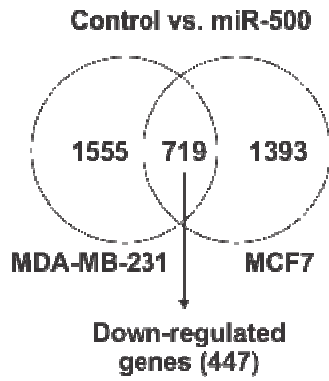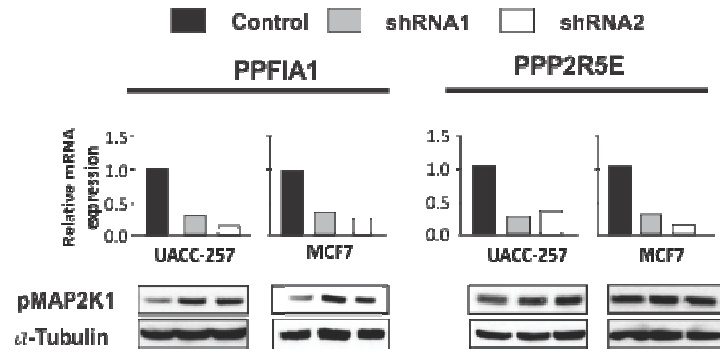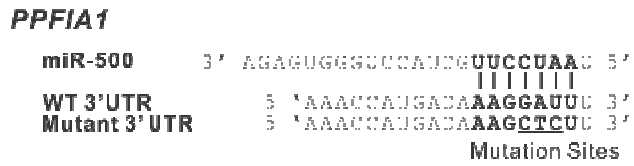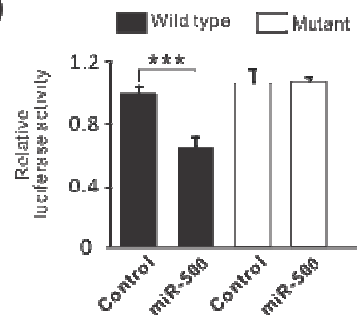
Figure 31. *PPFIA* as a direct target of miR-500 for regulation of *MAP2K1*. (A) Commonly down-regulated genes in MB-231 and MCF7 cells. After transfecting mimic miR-500 and control in MB-231 and MCF7 cells whole genome microarray experiment was performed and differentially expressed genes in two cell lines were selected by two-sample t-test ($P < 0.001$). (B) Western blots for phosphorylation of MAP2K1 after silencing *PPFIA1* and *PPP2R5E* in UACC-527 and MCG7 cells. Cells were transfected with specific shRNAs for 72 hrs. α-Tubulin was used as a loading control. (C) The miR-500 binding site in the 3′-UTR of *PPFIA1* mRNA.(D) Luciferase reporter assays used vectors including WT or Mutant sequences against 3'-UTR region of *PPFIA1*.

CHAPTER 6

DISCUSSION

MicroRNAs have critical roles as a main regulator in post transcriptional level. Because miRNAs have small nucleotide sequences, it is challengeable to characterize their functional roles in abnormal signaling pathways in cancer. In this study we developed a new correlation based approach on the integration of the three levels of omics data including miRNA, mRNA, and protein array data, for miRNA targeted biological pathways. In this approach we generated ASM and PASM for predict miRNA associated pathways. To validate our prediction that miR-500 is associated with MAPK pathway, we performed in-vitro experiment and found that miR-500 increased the phosphorylation of *MAP2K1* by regulating *PPFIA1* as a direct miR-500 target in melanoma and breast cancer cells.

Using three independent data sets (miRNAs, mRNAs, and proteins) from NCI-60 cell lines, we developed a series of systematic integration methods that can uncover functional roles of miRNAs in regulation of signaling pathways important for cancer development and established functional connection map of miRNAs to signaling pathways.

Our approaches have several advantages over conventional approaches. First, our analytical approach mimics natural process of cellular regulation in which miRNAs target mRNAs to regulate protein expression. We estimated the functional association between miRNAs and proteins by using mRNAs as intermediate denominators that were used to establish the functional connection

100

of miRNAs to signaling pathways. In our approach these intermediate denominators might contribute to reflecting biological context for identifying miRNA function in signaling pathway. Second, because miRNAs are connected to signaling pathways that are a functional unit of cellular regulation, interpretation of the data is easier and straightforward. Third, because we used data from NCI-60 cell lines that have been extensively characterized, the identified functional association can be easily tested and validated in cell lines as demonstrated in current study. Lastly, in our study we estimated the function of miRNA clusters in signaling pathways in cancer. Such clusters would more reliably predict the association between miRNAs and signaling pathways compared with using a single miRNA to understand their function and mechanism in various biological processes, because miRNAs in the same cluster are frequently regulated together.

While our approach uncovered new functional connections of miRNAs to signaling pathways, some of identified functional connection of miRNAs to signaling pathways are in good agreement with previous observations. A previous report showed that after the knockdown of the entire miR-183~96~182 cluster, its associated genes were enriched in apoptosis and the PI3K/AKT/mTOR pathway (Weeraratne et al. 2012). In our analysis we predicted that cluster 47, miR-96, -182, and -183 were significantly associated in regulation of mTOR pathway (**Table 4**). We also found miRNAs in cluster 5 (miR-127, -136, -337, -370, 431, -432, -433, -493, -665, and -770) were significantly enriched in

the regulation of immune system associated pathways such as Fc epsilon RI ($P <$ 0.0004), T cell receptor ($P <$ 0.0004), Toll-like receptor ($P <$ 0.0011), Fc gamma R-mediated phagocytosis ($P <$ 0.0023), and chemokine signaling pathways ($P <$ 0.012). In good agreement with our predictions, previous study showed that expression of miR-127 was increased in diffuse large B-cell lymphoma (Robertus et al. 2009) and also plays a role in inhibiting lung inflammation by targeting IgG Fcγ receptorI (CD64) in an IgG immune complex model *in-vivo* (Xie et al. 2012). miR-493 regulated the expression of E2F1 in Lung cancer (Gu et al. 2014) which was involved in controlling Innate immune receptor Toll-like receptor3 in epithelial cells (Taura et al. 2012). MiR-136 is known to be involved in regulation of hematopoietic lineage(Yu et al. 2006) and erythropoiesis (Choong et al. 2007).

In addition to good concordance of our prediction with previously known roles of miRNAs, reliability of our new method is further tested by functional validation of roles of miR-500 in regulation of MAPK pathway. Our data demonstrated that miR-500 function as a positive regulator of *MAP2K1*, upstream regulator of MAPKs, by directly targeting *PPFIA1*, a member of the LAR protein-tyrosine phosphatase-interacting protein (liprin) family. The previous studies showed that it was regulated by *ERK2* in MAPK pathway leading to the inhibition of tumor invasion and progression (von Thun et al. 2012), and identified as a putative invasion suppressor gene in head and neck cancer (Tan et al. 2008). We further demonstrated that *PPFIA1* is negative regulator of *MAP2K1* as evidenced by activation of *MAP2K1* upon depletion of *PPFIA1* in melanoma and breast cancer

cell lines. Although precise mechanism of *MAP2K1* regulation by *PPFIA1* is currently unknown, our data clearly demonstrated that *PPFIA1* is a novel regulator of *MAP2K1*. Taken together with strong concordance of our analysis with previous studies, these data strongly supported the validity of our approach in finding novel functional roles of miRNAs in the regulation of signaling pathways and its potential for identifying novel regulators of signaling pathways.

There are some limitations in our approaches. First, because we generated functional connection map between miRNAs and signaling pathways, extra experiments is necessary to identify direct targets of miRNAs that regulate signaling pathways. Second, our approaches are limited by number of available high quality antibodies used in RPPA experiments. However, because many of antibodies used in RPPA experiment were pre-selected for signaling pathways important for cancer development and progression, substantial portion of signaling pathways is covered by our study. Moreover, antibodies available for RPPA experiments have been steadily increased to discover proteomic biomarkers in cancer (Hennessy et al. 2010; Meric-Bernstam et al. 2014).

There is an limitation of correlation or association based approaches on miRNA target prediction. The correlation between miRNAs and target genes does not mean inferring their causal regulatory relationship. Because in this study we also used the correlation analysis to generate the indirect association between

miRNAs and proteins or biological pathways, it can be limited to examine the causality of miRNA-target genes.

1. Concluding remarks and Future directions

In this study we have developed a novel correlation based approach to characterize potential miRNA target pathways using ASM and PASM generated from genomic and proteomic data. Based on our approach, we found that miR-500 cluster was enriched in MAPK pathway and performed in-vitro validation that miR-500 is involved in MAP2K1 phosphorylation in breast cancer and melanoma cells. In addition, we identified *PPFIA1* as a direct target of miR-500 that regulates *MAPK1* in the MAPK pathway.

Our web-based database, miRPP, constructed with the procedures and approaches described in this study, may open up new opportunities to uncover novel molecular mechanisms regulated by miRNAs in cancer development. This, together with hypothesis-driven validation experiments as demonstrated with miR-500, paves the way for rapidly cataloging functional roles of miRNAs in signaling pathways in general.

2. Future works

Expanding our approach to larger data sets

With availability of The Cancer Genome Atlas data containing all three data sets (mRNAs, miRNAs, and proteins) from same tissues (Cancer Genome Atlas Research Network, 2014; Cancer Genome Atlas Research Network 2013; Cancer Genome Atlas Network, 2012), we expect our approach can be rapidly expanded to identify more functional links of miRNAs to signaling pathways in cancer type specific manner. Moreover, this can be also expanded to entire TCGA data to find generalizable connections of miRNAs to signaling pathways as seen in TCGA Pan-Cancer Project (Cancer Genome Atlas Research et al. 2013b) (Hoadley et al. 2014). Table 14 shows the TCGA multiomics data sets which can be applicable with our approach in cBioPortal (http://www.cbioportal.org). Through the analysis of these massive data using our approaches we expect to elucidate the complicated miRNA regulatory association between miRNAs and signaling pathways.

Improvement of prediction algorithms by including more high-throughput data.
Since our current approach is based on correlation that does not reflect direct functional connections, we will include functional experimental data in future analysis to improve accuracy of our prediction methods. These new data will include photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Bussey et al. 2006) assay, and high-

throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) (Roschke et al. 2003) with the next-generation sequencing.

Other non-cording RNAs

With developing RNA-sequencing technology, huge amount of ncRNA data are produced. However their functions have not been clear and needed to be characterized. lncRNA data from cancer tissues are available in TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). Therefore we can apply our approach to investigate lncRNA functions in signaling pathway using TCGA RNA-seq data in cancer.

miRPP

MiRPP currently provides the miRNA-correlated mRNA and RPPA protein profiles, indirection association between miRNAs and proteins, and the miRNA-associated signaling pathways. For the future works in this system, we will update the analysis ouputs of TCGA data sets to verify our approach in large cancer data sets with the clinical information such as patient survival, tumor recurrence and drug response.

Table 14. TCGA multiomics data sets available in cBioPortal*

*http://www.cbioportal.org (Gao et al. 2013)

CancerStudy	Reference	All	Sequenced	CNA	Tumor mRNA (RNA-Seq v2)	Tumor mRNA (microarray)	Tumor miRNA	Methylation (HM27)	RPPA
Breast Invasive Carcinoma (TCGA, Nature 2012)	TCGA, Nature 2012	825	507	778		526	302	311	410
Breast Invasive Carcinoma (TCGA, Provisional)		1061	976	1033	1037	526	286	313	747
Colorectal Adenocarcinoma (TCGA, Nature 2012)	TCGA, Nature 2012	276	224	257		224	85	236	196
Colorectal Adenocarcinoma (TCGA, Provisional)		600	223	589	352	222	86	233	461
Glioblastoma Multiforme (TCGA, Provisional)		596	283	497	153	500	427	285	214
Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013)	TCGA, Nature 2013.	499	424	436	469	72	171	219	454
Kidney Renal Clear Cell Carcinoma (TCGA, Provisional)		522	417	436	518	72	172	219	454
Lung Squamous Cell Carcinoma (TCGA, Provisional)		494	177	179	489	154	138	133	195
Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)		580	316	569	261	538	519	582	412

3. In summary

We have developed a novel correlation based approach to characterize potential miRNA target pathways using ASM and PASM generated from genomic and proteomic data. Based of our approach, we found that miR-500 cluster was enriched in MAPK pathway and performed in-vitro validation that miR-500 is involved in MEK1/2 phosphorylation in Breast cancer and melanoma cell. In addition, we identified *PPFIA1* as a direct target of miR-500 that regulates *MAP2K1* in the MAPK pathway. This approach can be applied to investigate unknown regulatory role of miRNAs in signaling pathway and miRNA regulatory networks in cancer.

BIBLIOGRAPHY

Akbani R, Ng PK, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang JY, Yoshihara K, Li J et al. 2014. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nature communications* **5**: 3887.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M et al. 2003. A uniform system for microRNA annotation. *Rna* **9**(3): 277-279.

Asgari S. 2011. Role of MicroRNAs in Insect Host-Microorganism Interactions. *Frontiers in physiology* **2**: 48.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455**(7209): 64-71.

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

-. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.

Becker LE, Lu Z, Chen W, Xiong W, Kong M, Li Y. 2012. A systematic screen reveals MicroRNA clusters that significantly regulate four major signaling pathways. *PLoS One* **7**(11): e48474.

Betel D, Koppal A, Agius P, Sander C, Leslie C. 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11**(8): R90.

Blower PE, Verducci JS, Lin S, Zhou J, Chung JH, Dai Z, Liu CG, Reinhold W, Lorenzi PL, Kaldjian EP et al. 2007. MicroRNA expression profiles for the NCI-60 cancer cell panel. *Mol Cancer Ther* **6**(5): 1483-1491.

Bogunovic D, O'Neill DW, Belitskaya-Levy I, Vacic V, Yu YL, Adams S, Darvishian F, Berman R, Shapiro R, Pavlick AC et al. 2009. Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. *Proc Natl Acad Sci U S A* **106**(48): 20429-20434.

Boorsma A, Foat BC, Vis D, Klis F, Bussemaker HJ. 2005. T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic acids research* **33**(Web Server issue): W592-595.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. *PLoS biology* **3**(3): e85.

Bussey KJ, Chin K, Lababidi S, Reimers M, Reinhold WC, Kuo WL, Gwadry F, Ajay, Kouros-Mehr H, Fridlyand J et al. 2006. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol Cancer Ther* **5**(4): 853-867.

Cancer Genome Atlas N. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418): 61-70.

Cancer Genome Atlas Research N. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**(7216): 1061-1068.

-. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**(7511): 543-550.

Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R et al. 2013. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**(7447): 67-73.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**(1): 77-94.

Chan WC, Ho MR, Li SC, Tsai KW, Lai CH, Hsu CN, Lin WC. 2012. MetaMirClust: discovery of miRNA cluster patterns using a data-mining approach. *Genomics* **100**(3): 141-148.

Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X et al. 2008. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research* **18**(10): 997-1006.

Choong ML, Yang HH, McNiece I. 2007. MicroRNA expression profiling during human cord blood-derived CD34 cell erythropoiesis. *Experimental hematology* **35**(4): 551-564.

Chou CH, Lin FM, Chou MT, Hsu SD, Chang TH, Weng SL, Shrestha S, Hsiao CC, Hung JH, Huang HD. 2013. A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC genomics* **14 Suppl 1**: S2.

Chowdhury I, Thompson WE, Thomas K. 2014. Prohibitins role in cellular survival through Ras-Raf-MEK-ERK pathway. *Journal of cellular physiology* **229**(8): 998-1004.

De Luca A, Maiello MR, D'Alessio A, Pergameno M, Normanno N. 2012. The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. *Expert opinion on therapeutic targets* **16 Suppl 2**: S17-27.

Deschenes-Simard X, Kottakis F, Meloche S, Ferbeyre G. 2014. ERKs in cancer: friends or foes? *Cancer research* **74**(2): 412-419.

Dweep H, Sticht C, Pandey P, Gretz N. 2011. miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *Journal of biomedical informatics* **44**(5): 839-847.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**(25): 14863-14868.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in Drosophila. *Genome Biol* **5**(1): R1.

Esteller M. 2011. Non-coding RNAs in human disease. *Nature reviews Genetics* **12**(12): 861-874.

Fazi F, Nervi C. 2008. MicroRNA: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular research* **79**(4): 553-561.

Federici G, Gao X, Slawek J, Arodz T, Shitaye A, Wulfkuhle JD, De Maria R, Liotta LA, Petricoin EF, 3rd. 2013. Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with "-OMIC" data fields and therapeutic response signatures. *Molecular cancer research : MCR* **11**(6): 676-685.

Ferracin M, Pedriali M, Veronese A, Zagatti B, Gafa R, Magri E, Lunardi M, Munerato G, Querzoli G, Maestri I et al. 2011. MicroRNA profiling for the identification of cancers with unknown primary tissue-of-origin. *The Journal of pathology* **225**(1): 43-53.

Garzon R, Fabbri M, Cimmino A, Calin GA, Croce CM. 2006. MicroRNA expression and function in cancer. *Trends in molecular medicine* **12**(12): 580-587.

Gennarino VA, Sardiello M, Avellino R, Meola N, Maselli V, Anand S, Cutillo L, Ballabio A, Banfi S. 2009. MicroRNA target prediction by expression analysis of host genes. *Genome Res* **19**(3): 481-490.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic acids research* **36**(Database issue): D154-158.

Gu Y, Cheng Y, Song Y, Zhang Z, Deng M, Wang C, Zheng G, He Z. 2014. MicroRNA-493 Suppresses Tumor Growth, Invasion and Metastasis of Lung Cancer by Regulating E2F1. *PLoS One* **9**(8): e102602.

Guo L, Yang S, Zhao Y, Zhang H, Wu Q, Chen F. 2014. Global analysis of miRNA gene clusters and gene families reveals dynamic and coordinated expression. *BioMed research international* **2014**: 782490.

Gurtan AM, Sharp PA. 2013. The role of miRNAs in regulating gene expression networks. *Journal of molecular biology* **425**(19): 3582-3600.

Hamzeiy H, Allmer J, Yousef M. 2014. Computational methods for microRNA target prediction. *Methods in molecular biology* **1107**: 207-221.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* **435**(7043): 828-833.

Hennessy BT, Lu Y, Gonzalez-Angulo AM, Carey MS, Myhre S, Ju Z, Davies MA, Liu W, Coombes K, Meric-Bernstam F et al. 2010. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clinical proteomics* **6**(4): 129-151.

Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V et al. 2014. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell*.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**(1): 44-57.

Hung CH, Chiu YC, Chen CH, Hu TH. 2014. MicroRNAs in hepatocellular carcinoma: carcinogenesis, progression, and therapeutic target. *BioMed research international* **2014**: 486407.

Inui M, Martello G, Piccolo S. 2010. MicroRNA control of signal transduction. *Nature reviews Molecular cell biology* **11**(4): 252-263.

Iorio MV, Croce CM. 2012a. Causes and consequences of microRNA dysregulation. *Cancer journal* **18**(3): 215-222.

-. 2012b. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO molecular medicine* **4**(3): 143-159.

Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M et al. 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer research* **65**(16): 7065-7070.

Jansson MD, Lund AH. 2012. MicroRNA and cancer. *Molecular oncology* **6**(6): 590-610.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human MicroRNA targets. *PLoS biology* **2**(11): e363.

Karin M, Hunter T. 1995. Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Current biology : CB* **5**(7): 747-757.

Kim SY, Volsky DJ. 2005. PAGE: parametric analysis of gene set enrichment. *BMC bioinformatics* **6**: 144.

Kozomara A, Hunt S, Ninova M, Griffiths-Jones S, Ronshaugen M. 2014. Target repression induced by endogenous microRNAs: large differences, small effects. *PLoS One* **9**(8): e104286.

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**(5): 495-500.

Kruger J, Rehmsmeier M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic acids research* **34**(Web Server issue): W451-454.

Kuhn DE, Martin MM, Feldman DS, Terry AV, Jr., Nuovo GJ, Elton TS. 2008. Experimental validation of miRNA targets. *Methods* **44**(1): 47-54.

Lall S, Grun D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P et al. 2006. A genome-wide map of conserved microRNA targets in C. elegans. *Current biology : CB* **16**(5): 460-471.

Leucci E, Onnis A, Cocco M, De Falco G, Imperatore F, Giuseppina A, Costanzo V, Cerino G, Mannucci S, Cantisani R et al. 2010. B-cell differentiation in EBV-positive Burkitt lymphoma is impaired at posttranscriptional level by miRNA-altered expression. *International journal of cancer Journal international du cancer* **126**(6): 1316-1326.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15-20.

Liang T, Yu J, Liu C, Guo L. 2014. An exploration of evolution, maturation, expression and function relationships in mir-23 approximately 27 approximately 24 cluster. *PLoS One* **9**(8): e106223.

Liu H, Brannon AR, Reddy AR, Alexe G, Seiler MW, Arreola A, Oza JH, Yao M, Juan D, Liou LS et al. 2010a. Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. *BMC systems biology* **4**: 51.

Liu H, D'Andrade P, Fulmer-Smentek S, Lorenzi P, Kohn KW, Weinstein JN, Pommier Y, Reinhold WC. 2010b. mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. *Mol Cancer Ther* **9**(5): 1080-1091.

Liu T, Papagiannakopoulos T, Puskar K, Qi S, Santiago F, Clay W, Lao K, Lee Y, Nelson SF, Kornblum HI et al. 2007. Detection of a microRNA signal in an in vivo expression set of mRNAs. *PLoS One* **2**(8): e804.

Liu X, Chen Z, Yu J, Xia J, Zhou X. 2009. MicroRNA profiling and head and neck cancer. *Comparative and functional genomics*: 837514.

Lorenzi PL, Reinhold WC, Varma S, Hutchinson AA, Pommier Y, Chanock SJ, Weinstein JN. 2009. DNA fingerprinting of the NCI-60 cell line panel. *Mol Cancer Ther* **8**(4): 713-724.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**(7043): 834-838.

Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. 2008. An analysis of human microRNA and disease associations. *PLoS One* **3**(10): e3420.

Macfarlane LA, Murphy PR. 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics* **11**(7): 537-561.

Mammano E, Galdi F, Pierobon M, Tessari E, Deng J, Pucciarelli S, Agostini M, De Marchi F, Canzonieri V, De Paoli A et al. 2012. Multiplexed protein signal pathway mapping identifies patients with rectal cancer that responds to neoadjuvant treatment. *Clinical colorectal cancer* **11**(4): 268-274.

Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K et al. 2009. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic acids research* **37**(Web Server issue): W273-276.

Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M, Kourtis K, Koziris N, Dalamagas T, Hatzigeorgiou AG. 2011. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic acids research* **39**(Web Server issue): W145-148.

Martin G, Schouest K, Kovvuru P, Spillane C. 2007. Prediction and validation of microRNA targets in animal genomes. *J Biosci* **32**(6): 1049-1052.

Maziere P, Enright AJ. 2007. Prediction of microRNA targets. *Drug discovery today* **12**(11-12): 452-458.

Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG. 2007. miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic acids research* **35**(Database issue): D149-155.

Meric-Bernstam F, Akcakanat A, Chen H, Sahin A, Tarco E, Carkaci S, Adrada BE, Singh G, Do KA, Garces ZM et al. 2014. Influence of biospecimen variables on proteomic biomarkers in breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**(14): 3870-3883.

Michael MZ, SM OC, van Holst Pellekaan NG, Young GP, James RJ. 2003. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Molecular cancer research : MCR* **1**(12): 882-891.

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**(6): 1203-1217.

Mogilyansky E, Rigoutsos I. 2013. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell death and differentiation* **20**(12): 1603-1614.

Mosmann T. 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *Journal of immunological methods* **65**(1-2): 55-63.

Nair VS, Pritchard CC, Tewari M, Ioannidis JP. 2014. Design and Analysis for Studying microRNAs in Human Disease: A Primer on -Omic Technologies. *American journal of epidemiology* **180**(2): 140-152.

Neuzillet C, Tijeras-Raballand A, de Mestier L, Cros J, Faivre S, Raymond E. 2014. MEK in cancer and cancer therapy. *Pharmacology & therapeutics* **141**(2): 160-171.

Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, Poon TC, Ng SS, Sung JJ. 2009. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut* **58**(10): 1375-1381.

Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, Kouros-Mehr H, Bussey KJ, Lee JK, Espina V et al. 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A* **100**(24): 14229-14234.

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**(7043): 839-843.

Olive V, Jiang I, He L. 2010. mir-17-92, a cluster of miRNAs in the midst of the cancer network. *The international journal of biochemistry & cell biology* **42**(8): 1348-1354.

Olson P, Lu J, Zhang H, Shai A, Chun MG, Wang Y, Libutti SK, Nakakura EK, Golub TR, Hanahan D. 2009. MicroRNA dynamics in the stages of

tumorigenesis correlate with hallmark capabilities of cancer. *Genes & development* **23**(18): 2152-2165.

Paranjape T, Slack FJ, Weidhaas JB. 2009. MicroRNAs: tools for cancer diagnostics. *Gut* **58**(11): 1546-1554.

Park ES, Rabinovsky R, Carey M, Hennessy BT, Agarwal R, Liu W, Ju Z, Deng W, Lu Y, Woo HG et al. 2010. Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. *Mol Cancer Ther* **9**(2): 257-267.

Paroo Z, Liu Q, Wang X. 2007. Biochemical mechanisms of the RNA-induced silencing complex. *Cell research* **17**(3): 187-194.

Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. 2014. Common features of microRNA target prediction tools. *Frontiers in genetics* **5**: 23.

Qi Q, Zhao Y, Li M, Simon R. 2009. Non-negative matrix factorization of gene expression profiles: a plug-in for BRB-ArrayTools. *Bioinformatics* **25**(4): 545-547.

Robertus JL, Harms G, Blokzijl T, Booman M, de Jong D, van Imhoff G, Rosati S, Schuuring E, Kluin P, van den Berg A. 2009. Specific expression of miR-17-5p and miR-127 in testicular and central nervous system diffuse large B-cell lymphoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **22**(4): 547-555.

Romero-Cordoba SL, Salido-Guadarrama I, Rodriguez-Dorantes M, Hidalgo-Miranda A. 2014. miRNA biogenesis: Biological impact in the development of cancer. *Cancer biology & therapy* **15**(11): 1444-1455.

Roschke AV, Tonon G, Gehlhaus KS, McTyre N, Bussey KJ, Lababidi S, Scudiero DA, Weinstein JN, Kirsch IR. 2003. Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer research* **63**(24): 8634-8647.

Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, Benjamin H, Shabes N, Tabak S, Levy A et al. 2008. MicroRNAs accurately identify cancer tissue origin. *Nature biotechnology* **26**(4): 462-469.

Roskoski R, Jr. 2012. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacological research : the official journal of the Italian Pharmacological Society* **66**(2): 105-143.

Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M et al. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* **24**(3): 227-235.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**(7149): 83-86.

Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ. 2010. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol* **11**(1): R6.

Saj A, Lai EC. 2011. Control of microRNA biogenesis and transcription by cell signaling pathways. *Current opinion in genetics & development* **21**(4): 504-510.

Sebolt-Leopold JS. 2000. Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene* **19**(56): 6594-6599.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**(7209): 58-63.

Sethupathy P, Corda B, Hatzigeorgiou AG. 2006. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna* **12**(2): 192-197.

Shah NR, Chen H. 2014. MicroRNAs in pathogenesis of breast cancer: Implications in diagnosis and treatment. *World journal of clinical oncology* **5**(2): 48-60.

Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, Reimers MA, Scherf U, Kahn A, Dolginow D et al. 2007. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* **6**(3): 820-832.

Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. 2007. Analysis of gene expression data using BRB-ArrayTools. *Cancer informatics* **3**: 11-17.

Sontheimer EJ. 2005. Assembly and function of RNA silencing complexes. *Nature reviews Molecular cell biology* **6**(2): 127-138.

Sullivan CS, Ganem D. 2005. A virus-encoded inhibitor that blocks RNA interference in mammalian cells. *Journal of virology* **79**(12): 7371-7379.

Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y et al. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer research* **64**(11): 3753-3756.

Tan KD, Zhu Y, Tan HK, Rajasegaran V, Aggarwal A, Wu J, Wu HY, Hwang J, Lim DT, Soo KC et al. 2008. Amplification and overexpression of PPFIA1, a putative 11q13 invasion suppressor gene, in head and neck squamous cell carcinoma. *Genes, chromosomes & cancer* **47**(4): 353-362.

Taura M, Suico MA, Koyama K, Komatsu K, Miyakita R, Matsumoto C, Kudo E, Kariya R, Goto H, Kitajima S et al. 2012. Rb/E2F1 regulates the innate immune receptor Toll-like receptor 3 in epithelial cells. *Molecular and cellular biology* **32**(8): 1581-1590.

Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM. 2006. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther* **5**(10): 2512-2521.

Thomas M, Lieberman J, Lal A. Desperately seeking microRNA targets. Nat Struct Mol Biol. 2010 Oct;17(10):1169-74.

Varma S, Pommier Y, Sunshine M, Weinstein JN, Reinhold WC. 2014. High resolution copy number variation data in the NCI-60 cancer cell lines from

whole genome microarrays accessible through CellMiner. *PLoS One* **9**(3): e92047.

Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, Paraskevopoulou MD, Prionidis K, Dalamagas T, Hatzigeorgiou AG. 2012. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic acids research* **40**(Web Server issue): W498-504.

von Thun A, Birtwistle M, Kalna G, Grindlay J, Strachan D, Kolch W, von Kriegsheim A, Norman JC. 2012. ERK2 drives tumour cell migration in three-dimensional microenvironments by suppressing expression of Rab17 and liprin-beta2. *Journal of cell science* **125**(Pt 6): 1465-1477.

Wang L, Oberg AL, Asmann YW, Sicotte H, McDonnell SK, Riska SM, Liu W, Steer CJ, Subramanian S, Cunningham JM et al. 2009. Genome-wide transcriptional profiling reveals microRNA-correlated genes and biological processes in human lymphoblastoid cell lines. *PLoS One* **4**(6): e5878.

Wang YP, Li KB. 2009. Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. *BMC Genomics* **10**: 218.

Weeraratne SD, Amani V, Teider N, Pierre-Francois J, Winter D, Kye MJ, Sengupta S, Archer T, Remke M, Bai AH et al. 2012. Pleiotropic effects of miR-183~96~182 converge to regulate cell survival, proliferation and migration in medulloblastoma. *Acta neuropathologica* **123**(4): 539-552.

Winter J, Jung S, Keller S, Gregory RI, Diederichs S. 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology* **11**(3): 228-234.

Witkos TM, Koscianska E, Krzyzosiak WJ. 2011. Practical Aspects of microRNA Target Prediction. *Current molecular medicine* **11**(2): 93-109.

Wurdinger T, Costa FF. 2007. Molecular therapy in the microRNA era. *The pharmacogenomics journal* **7**(5): 297-304.

Xie T, Liang J, Liu N, Wang Q, Li Y, Noble PW, Jiang D. 2012. MicroRNA-127 inhibits lung inflammation by targeting IgG Fcgamma receptor I. *Journal of immunology* **188**(5): 2437-2444.

Xu J, Wong C. 2008. A computational screen for mouse signaling pathways targeted by microRNA clusters. *Rna* **14**(7): 1276-1283.

Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. 2011. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic acids research* **39**(Database issue): D202-209.

Ye JJ, Cao J. 2014. MicroRNAs in colorectal cancer as markers and targets: Recent advances. *World journal of gastroenterology : WJG* **20**(15): 4288-4299.

Yu J, Wang F, Yang GH, Wang FL, Ma YN, Du ZW, Zhang JW. 2006. Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochem Biophys Res Commun* **349**(1): 59-68.

Zeeberg BR, Reinhold W, Snajder R, Thallinger GG, Weinstein JN, Kohn KW, Pommier Y. 2012. Functional categories associated with clusters of genes that are co-expressed across the NCI-60 cancer cell lines. *PLoS One* **7**(1): e30317.

Zhang B, Pan X, Cobb GP, Anderson TA. 2007. microRNAs as oncogenes and
tumor suppressors. *Dev Biol* **302**(1): 1-12.

VITA

Sangbae Kim was born in Namyangju, Kyonggi, Korea in 1972, and received the degree of Bachelor of Science (B.S.) from Kang-Won National University in Chunchon, South Korea in 1998 and Master of Science (M.A.) degree from Korea University in Seoul, South Korea in 2002. In September 2008, he enrolled in the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences, and joined Dr. Ju-Seog Lee laboratory in 2009.

➢   Permanent address: 502, HanWha apt, Jikumdong, Namyangju, Kyunggi-do, South Korea, 472-080