

國立臺灣大學電機資訊學院生醫電子與資訊學研究所

碩士論文

Graduate Institute of Biomedical Electronics and Bioinformatics

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

微型核糖核酸在乳癌所調控的蛋白質交互作用網路及其功能

MicroRNA-regulated protein-protein interaction networks and
their functions in breast cancer

李佳憲

Chia-Hsien Lee

指導教授： 阮雪芬 博士

歐陽彥正 博士

Advisor: Hsueh-Fen Juan, Ph.D.

Yen-Jen Oyang, Ph.D.

中華民國 101 年 6 月

June, 2012

國立臺灣大學碩士學位論文
口試委員會審定書

微型核糖核酸在乳癌所調控的蛋白質交互作用網路及
其功能

MicroRNA-regulated protein-protein interaction networks
and their functions in breast cancer

本論文係李佳憲君 (R99945042) 在國立臺灣大學生醫電子與資訊學研究所完成之碩士學位論文，於民國一零一年六月二十六日承下列考試委員審查通過及口試及格，特此證明

口試委員：

陳忠信 (指導教授)

歐陽彥正 (指導教授)

黃立國

蔡懷寬

陳倩瑜

所長

賴市龍

誌謝

經過了兩年的時間，在許多人的幫助下，終於完成了這篇論文。雖然有句話說「要感謝的人太多，就謝天吧」，但我認為這樣做對不起那些努力幫助過我的人們。

能夠完成這篇論文，首先必須感謝我的指導老師阮雪芬老師、歐陽彥正老師，以及陽明大學生物醫學資訊研究所的黃宣誠老師。阮老師和黃老師帶領我進入生物資訊的世界，並在整個實驗方向，從資料分析、結果判讀以及最後的圖表呈現，都給予了許多珍貴的建議，並且提供了經濟上的幫助，讓我能夠無後顧之憂地完成這個研究。歐陽老師則給了我許多統計的基礎，並幫助我加強本論文的架構。此外，除了我的指導教授外，也感謝台大生機系的陳倩瑜老師，以及中研院資訊科學研究所的蔡懷寬老師在百忙中抽空審閱我的論文，給我意見，並且參加我的口試，以及給予我論文口試時口頭報告上的建議；以及臺大醫院的郭文宏醫師，他提供了一組有趣的病人的基因與微型核糖核酸表現資料，並參與了資料分析的討論。如果沒有這些師長的幫忙，這篇論文就無法被完成。

我還要感謝我的實驗室夥伴在我從大三進來這個實驗室開始到現在的陪伴以及幫助。在這段期間，要感謝翠琴和心儀學姊一開始帶我做實驗，並且(和阮老師一起)讓我明白自己沒有那麼適合做實驗，進而走向生物資訊的道路。也感謝振慶和卓逸學長在這個研究流程中的實作方面給我的許多建議。在和他們討論的過程中，我學到了許多的資料處理方法，還有許多要注意的細節。此外也感謝實驗室其他所有的同仁在這段時間中給我帶來的歡樂時光。

最後，要感謝這一路上支持我的人們。首先是我的家人：父母親、弟弟、還有我的兩隻可愛的寵物龜。沒有他們的經濟以及精神上的支持，我應該沒有辦法走到這一步。接著，要感謝自由軟體運動的發起人理查·史托曼先生(Richard M. Stallman)。沒有他和他的自由軟體運動，就沒有在本研究中使用到的許多優秀的好軟體(Python, R 等)，也會讓我少掉一項重要有趣的休閒活動並失去 GNU/Linux 的陪伴。最後，要特別感謝花澤香菜(Kana Hanazawa)小姐(日本配音員、歌手)：她甜美的聲音支持著我，讓我得以度過每一個難熬又可怕的挫折。

謹將這個論文獻給支持我或是讓我看見夢想的人們。

中文摘要

微型核糖核酸是一種被發現和許多種的癌症有關連性的內生性基因調節分子。有些微型核糖核酸會調控和癌症功能相關的基因，包括細胞凋亡、細胞新生、抑制癌症或是導致癌症產生。乳癌對於全世界的女性而言是一個主要的健康威脅。有許多的微型核糖核酸被報導與乳癌的癌化及進程有關，本研究主要在探討並試圖尋找新的與乳癌相關的微型核糖核酸。

首先，我們把微型核糖核酸的目標預測資料庫與微型核糖核酸、訊息核糖核酸的表現資訊整合在一起，以找出具信心的微型核糖核酸—目標配對。由於我們把實際在乳癌與正常組織中的表現資訊考慮進去，我們得以降低偽陽性，並且取得與乳癌功能相關的微型核糖核酸—目標配對。接著，我們利用具信心的微型核糖核酸—目標配對，以及在 HPRD 中已知的蛋白質交互作用資料來建立微型核糖核酸與其調控的蛋白質交互作用網路。最後，我們利用功能性分析來找出微型核糖核酸與其調控的蛋白質交互作用網路所擁有的功能。

從我們的結果看來，我們除了找到了 miR-125a、miR-125b、miR-21 以及 miR-497 這些之前已知和乳癌相關的微型核糖核酸，也找到有些以前未曾報告過與乳癌相關的微型核糖核酸，例如 miR-139 和 miR-383；同時，我們也試著在這篇研究中推測它們的功能。我們更利用其他表現量資料跑 ROC 曲線分析，使用 GOBO 做存活率分析，以及文獻搜尋來驗證我們的結果。我們的結果可能可以替未來的乳癌相關的微型核糖核酸的研究提供一些新的線索。

關鍵字：微型核糖核酸, 乳癌, 蛋白質交互作用網路, 功能性分析



ABSTRACT

MicroRNAs, small endogenous RNA regulators, were found to be related to various types of cancer. Some of the miRNAs which target some tumor related genes, including apoptosis and cell proliferation, were confirmed to be tumor suppressors or oncomirs. Breast cancer is a major health threat for women worldwide. Many miRNAs were reported to be associated to the progression and carcinogenesis of breast cancer. In this work, we tried to elucidate and discover novel breast cancer related miRNAs.

First, we find confident miRNA-target pairs by combining miRNA target prediction databases and expression profiles of miRNA and mRNA. By considering actual expression profiles in normal and breast tumor samples, we may reduce false positives while obtaining the miRNA-target pairs relevant to breast cancer. miRNA-regulated PINs were then constructed with the confident pairs and known interaction data in HPRD. Finally, the functions of miRNA-regulated PINs were then elucidated by functional enrichment analysis.

From the result, we found some previously known breast cancer-related miRNAs and functions of the PIN, like miR-125b, miR-125a, miR-21, and miR-497. Some miRNAs without known association to breast cancer were also found, and the putative functions of their PINs were also elucidated. Such novel miRNAs were miR-139 and miR-383. Furthermore, we tried to validate our result by ROC analysis using another expression profile data, GOBO survival analysis, and literature search. Our results may provide new insights for researches of breast cancer associated miRNAs.

Keywords: miRNA, breast cancer, protein interaction network, functional analysis



CONTENTS

口試委員會審定書	i
誌謝	iii
中文摘要	v
ABSTRACT	vii
CONTENTS	ix
LIST OF FIGURES	xi
LIST OF TABLES	xiii
Chapter 1 Introduction.....	1
1.1 Breast Cancer and miRNA	1
1.2 Integrated Analysis of miRNA and mRNA Expression Profiles	3
1.3 Specific aims.....	4
Chapter 2 Materials and Methods.....	5
2.1 Expression datasets.....	5
2.1.1 Public mRNA Microarray Data.....	5
2.1.2 Public RNA-seq miRNA data	6
2.1.3 miRNA array dataset obtained from clinical trial	7
2.2 Databases related to miRNAs.....	8
2.2.1 miRBase	8
2.2.2 TargetScan	9
2.2.3 miRanda	9
2.2.4 PicTar	9
2.3 Construction of miRNA target prediction collection from TargetScan, MiRanda and PicTar	10

2.4	Significance Analysis of Microarrays (SAM)	10
2.5	Protein-protein interaction database	10
2.6	Construction of miRNA-regulated protein interaction network.....	11
2.7	Gene Ontology.....	11
2.8	Hypergeometric Test and GO Enrichment Analysis.....	12
2.9	Activity analysis	13
2.10	ROC curve and ROC analysis	14
2.11	Survival analysis with GOBO	15
Chapter 3	Results	16
3.1	Differentially expressed miRNAs and mRNAs.....	16
3.2	miRNA-regulated PINs	16
3.3	Enriched GO Terms in the miRNA-regulated PINs	16
3.4	Activity of the miRNA-regulated PINs	17
3.5	Validation.....	18
3.5.1	ROC curve and ROC analysis on NTUH miRNA array data	18
3.5.2	Validation of functional analysis by GOBO.....	18
Chapter 4	Discussions	19
4.1	Construction of miRNA-regulated PINs	19
4.2	Functional analysis of miRNA-regulated PINs	21
Chapter 5	Conclusions.....	24
FIGURES		25
TABLES		57
REFERENCE		89

LIST OF FIGURES

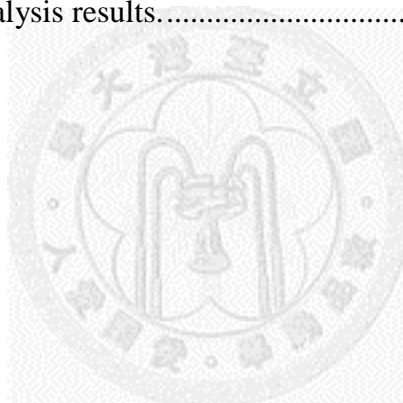
Figure 1. Pipeline for miRNA-regulated PINs construction and functional analysis.	25
Figure 2. Definition of GO level in this work.	26
Figure 3. Schematic figure describing activity analysis.	27
Figure 4. Distribution of miRNA dataset GSE29173.	28
Figure 5. Distribution of gene expression dataset GSE29174.	29
Figure 6. miRNA-regulated PIN (hsa-let-7c).	30
Figure 7. miRNA-regulated PIN. (hsa-miR-21-5p)	31
Figure 8. miRNA-regulate PIN. (hsa-miR-22-3p)	32
Figure 9. miRNA-regulated PINs. (hsa-miR-31-5p and hsa-miR-99a-5p)	33
Figure 10. miRNA-regulated PINs. (hsa-miR-100-5p and hsa-miR-139-5p)	34
Figure 11. miRNA-regulated PIN. (hsa-204-5p)	35
Figure 12. miRNA-regulated PINs. (hsa-miR-214-3p, hsa-miR-215, and hsa-miR-122-5p)	36
Figure 13. miRNA-regulated PIN. (hsa-miR-125b-5p)	37
Figure 14. miRNA-regulated PIN. (hsa-miR-145-5p)	38
Figure 15. miRNA-regulated PIN. (hsa-miR-125a-5p)	39
Figure 16. miRNA-regulated PINs. (hsa-miR-383 and hsa-miR-193b-3p)	40
Figure 17. miRNA-regulated PIN. (hsa-miR-497-5p)	41
Figure 18. miRNA-regulated PIN. (hsa-miR-520d-3p)	42
Figure 19. Distribution of Pearson correlation coefficient of gene pairs in microarray data.	43
Figure 20. ROC curves and expression distribution of miRNAs.	44
Figure 21 GOBO Kaplan-Meier survival curve of selected gene set with significant	

clinical outcome.....53



LIST OF TABLES

Table 1. Significantly differentially expressed miRNAs found in dataset GSE29173....	57
Table 2. Down-regulated genes found in dataset GSE29174.	58
Table 3. Up-regulated genes found in dataset GSE29174.	65
Table 4. Raw data count of each prediction database.	70
Table 5. Summary of constructed miRNA-regulated networks.	71
Table 6. Specific enriched GO terms of each miRNA-regulated PINs.	72
Table 7. P-value of activation status of each miRNA-regulated PINs.	83
Table 8. Result of ROC curve analysis on NTUH miRNA array data.	84
Table 9. GOBO survival analysis results.	85





Chapter 1 Introduction

1.1 Breast Cancer and miRNA

Breast cancer is a global threat to women health. According to a 2008 survey [1], breast cancer was the leading cause of cancer death in women. Currently, we know some possible risk factors, diagnostic methods and some drug and surgery procedure for treatment, but we don't know the detail of progression of breast carcinoma, and the most important, how to cure breast cancer.

From previous researches, we knew some risk factors of breast cancer. Early menarche, late menopause, obesity, late first full pregnancy, and hormone replacement therapy were considered as high risk factors of breast cancer [2]. Breast cancer risk was also reported to be related to fat intake when red meats and high-fat dairy foods were main source of fat [3].

For diagnosis, the most widely used measures are mammography, ultrasonography, MRI (magnetic resonance imaging) and PET (positron emission tomography) [2]. Biopsy may be used to confirm the existence of malignant tissue if suspicious results are produced in the diagnosis stage. Further analysis on the tissue obtained from biopsy may point out the best methods for treatment. After breast tumor is confirmed, the patient can be treated by surgery, radiotherapy, chemotherapy, and targeting therapy, depending on the conditions of the patient.

One of the most widely used classifying methods for breast cancer is to use expression state of three biological markers: estrogen receptors (ER), progesterone receptors (PR), and HER2/neu [4]. ER and PR were hormone receptors which may be over-expressed in breast carcinoma. To treat ER positive patients, it is possible to use anti-estrogen therapy or tamoxifen, which is a selective ER modulator [5]. It was

reported that chemotherapy had better performance in ER negative patients than ER positive ones [6]. Her2/neu is a widely used prognostic and predictive marker in breast cancer. For patients with newly diagnosed breast cancer without receiving adjuvant systemic therapy, overexpression of Her2/neu is associated to bad prognosis [7]. A targeting therapy, trastuzumab, which is a Her2/neu targeting antibody, was developed for breast cancer treatment [8].

MicroRNAs (miRNAs), one of the small non-coding RNAs [9], are short endogenous nucleotides which is able to regulate gene expression. After miRNAs precursor are transcribed from the genome, or generated from spliceosomes, they are exported to cytoplasm and further processed by Dicer complex [10]. The mature miRNA would be then bound to Argonaute protein, forming a miRNA-protein complex known as miRNP or RNAi (RNA interference) enzyme complex [11, 12]. The miRNP complex was reported to be able to down-regulate target genes by translational repression [13] or mRNA cleavage [14].

Like other protein-based regulators, miRNAs were known to have important roles in cancer. Calin *et al.* reported that miR-15 and miR-16 were deleted in leukemia [15], which was believed to be one of the earliest report of the association of miRNA and cancer [16]. After the report, many miRNAs acting as tumor suppressor or oncogenes (which were also known as oncomirs) were found. For example, miR-21 was identified as oncomir at least in hepatocellular cancer [17], breast cancer [18], and kidney cancer [19]. Let-7c was found to be a tumor suppressor in prostate cancer [20]. miR-181a was reported as a tumor suppressor in glioma [21]. MiR-125b [22], and miR-145 [23] were known as tumor suppressors of breast cancer. Mir-125a was found to repress tumor growth in breast cancer [24]. It is thus undeniable that miRNAs play an important role in breast cancer.

1.2 Integrated Analysis of miRNA and mRNA Expression Profiles

Since miRNA functions by regulating its target genes, we may deduce the effect of a miRNA by analyzing its regulated networks. To use such a method to elucidate miRNA functions, targets of miRNAs should be deduced. Currently, predictions in most target prediction database were based on sequence and statistical methods [25]. For example, in TargetScan, seed base pairing, target site context, conservation of target site and miRNA, and site accessibility were considered in the prediction process [26].

Another way to elucidate targets of miRNAs is to integrate expression profiles of miRNA and mRNA. In the work of Huang *et al.* [27], they developed a Bayesian-based algorithms, GenMiR++, to predict possible targets of 104 miRNAs in human. They also verified their results by RT-PCR and microarray experiments. However, the power of other sequence-based target prediction algorithms was not utilized in their work.

It is also possible to combine sequence-based target prediction and expression-based target prediction methods. By integrating expression data into sequence-based predictions, possible false positives can be reduced. Previously, miRNA-mRNA interactions were explored with splitting-averaging Bayesian networks [28]. In that work, expression profiles of miRNA and mRNA from public databases, miRNA target prediction databases, and miRNA sequence information were integrated together to discover miRNA-mRNA interaction network.

Here, we combined expression profiles of miRNA and mRNA, and 3 target prediction databases, TargetScan, PicTar and miRanda, to obtain confident miRNA-mRNA relationships and construct miRNA-regulated protein-protein networks for breast cancer. Furthermore, we explored the functions of the miRNAs by inspecting

the underlying PINs of the miRNA with functional enrichment analysis. This method was already used previously to find functions of gastric cancer related miRNAs in the work of Tseng *et al.* [29]. In that research, a gastric cancer related miRNA, miR-148a, were discovered and validated to be related to tumor proliferation, invasion, migration, and survival rate of the patients. By using similar method, we wish to elucidate breast cancer related miRNA-regulated protein interaction networks (PINs) and their functions.

1.3 Specific aims

Main purposes of this work are described as follows:

1. To construct breast cancer-related miRNA-regulated protein interaction networks.
2. To elucidate the function of miRNA-regulated PINs.
3. To validate the results of the previous step by another expression profiles with clinical annotations (such as metastasis, survival).

Chapter 2 Materials and Methods

Overall workflow design was similar to the previous work of Tseng *et al* [29] (see **Figure 1**). However, we applied the workflow on breast cancer expression profile instead of gastric cancer in the work of Tseng *et al*. Additionally, we used 3 miRNA target prediction databases, TargetScan, PicTar, and miRanda here, while only TargetScan was used before.

We first elucidated differentiated miRNAs and mRNAs, and combined the results with target prediction databases. By these means we can construct a miRNA-target networks consisted only genes directly regulated by the miRNA. We defined such network “miRNA-L0 network”. To extend our view, we combined interaction data from HPRD with our miRNA-L0 network and constructed miRNA-regulated PINs. We then elucidated putative functions of the miRNA-regulated PINs by GO enrichment analysis and validated our results by literature search and GOBO.

2.1 Expression datasets

2.1.1 Public mRNA Microarray Data

The mRNA expression profile used in this work for miRNA-regulated protein-protein interaction network construction is fetched from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>, ID: GSE29174). This dataset was produced by Farazi *et al* [30]. There are totally 161 clinical samples collected from breast cancer patients by biopsy in this dataset: 110 IDC, 11 normal, 17 DCIS, 1 mucinous A, 8 atypical medullary, 4 apocrine, 8 metaplastic, and 2 adenoid, as classified by Farazi *et al*. We used the 110 IDC samples as tumor group and 11 normal samples as normal group in this research. The array platform used in the dataset, NKI-CMF Homo sapiens 35k oligo array, contains 37,632 probes.

The array data was normalized with R using limma package (version 3.10.0). First, we used quantile normalization function provided by limma package. Then, we used following formula to scale the array.

$$SNE(s, g) = e^{\frac{NE(s, g)}{\max(NE)}}$$

where SNE is “scaled normalized expression”, NE is “normalized expression”. The parameter ‘s’ and ‘g’ represents “sample” and “gene” respectively.

After normalization, we transformed probe ID into gene ID. For those genes represented by multiple probes, we picked the probe with highest absolute expression value as expression of the gene. (This method was inspired by Miller *et al.* [31]) After these preprocessing steps, 15,560 genes were retained and would be used in the analysis step.

2.1.2 Public RNA-seq miRNA data

The miRNA expression profile used in this work was fetched from supplementary table S4A of the work of Farazi *et al* [30]. The table is actually the refined (by Farazi *et al.*) version of GEO dataset GSE29173. Since the data in GEO is stored as sequence-read pair, for example, sequence “CTTGACCGCTCTGACCA” has 279 reads. It would be easier to use the refined table provided by Farazi *et al.* There were 189 samples in this dataset, 6 of them were from the cell line, and another 183 samples are from patient tissues. In the 183 clinical samples collected from breast cancer patients by biopsy in this dataset, there were 128 IDC, 11 normal, 18 DCIS, 1 mucinous A, 8 atypical medullary, 4 apocrine, 9 metaplastic, and 2 adenoid, as classified by Farazi *et al.* We used the 128 IDC samples as tumor group and 11 normal samples as normal group in this research. “Predominant miR sequence” in supplementary table S4F of the work of Farazi *et al.* is used to determine miRBase accession number of the miRNAs in the

dataset.

The normalization steps used here is the same as mRNA array data described in section 2.1.1. We removed miRNAs which could not be mapped to a miRNA accession number in miRBase after normalization. There were 816 miRNAs left after all the preprocessing steps were completed.

2.1.3 miRNA array dataset obtained from clinical trial

A miRNA expression dataset generated by miRNA microarray were obtained from Dr. Wen-Hong Kuo, National Taiwan University Hospital. There were totally 15 normal samples and 101 tumor samples used in this work. All the human tissue samples collected from breast cancer patients have been approved and human subject confidentiality has been protected by the Institute Review Board (IRB, 20071211R). Collected samples were sent to Welgene for RNA extraction and miRNA array experiments, describes below.

For RNA extraction, total RNA was extracted from tissues collected from the patients by Trizol[®] Reagent (Invitrogen, USA) according to the instruction manual. RNA purified was quantified at OD260nm by using a ND-1000 spectrophotometer (Nanodrop Technology, USA) and qualitated by using a Bioanalyzer 2100 (Agilent Technology, USA) with RNA 6000 nano labchip kit (Agilent Technologies, USA).

After RNA extraction, 100 ng of total RNAs were dephosphorylated and labeled with pCp-Cy3 by Agilent miRNA Complete Labeling and Hyb Kit (Agilent Technologies, USA, microRNA Spike- In Apply). 2X hybridization buffer (Agilent Technologies, USA) was added to the labeled mixture to a final volume of 45 μ l. The mixture was heated for 5 min at 100°C and immediately cooled to 0°C. Each 45 μ l sample was hybridized onto an Agilent human miRNA Microarray Release 12.0, 8x15K

(Agilent Technologies, USA) at 55°C for 20 hours. After hybridization, slides were washed for 5 min in Gene Expression Wash Buffer 1 at room temperature, then for 5 min in Gene Expression Wash Buffer 2 at 37°C. Slides were scanned on an Agilent microarray scanner (Agilent Technologies, model G2505C) at 100% and 5% sensitivity settings. Feature Extraction (Agilent Technologies) software version 10.7.3.1 was used for image analysis.

To preprocess the dataset, we first extracted the expression value from original data provided by Welgene. In the data provided by Welgene, there were 2 columns, called gTotalGeneSignal (expression value) and gIsGeneDetected (is the gene actually being detected?). We utilized them by setting gTotalGeneSignal to 0.1 (minimum value of all expression data) while gIsGeneDetected is 0 (i.e. not detected). After the initial table were extracted by the steps above, we preprocess (i.e. normalizing and scaling) the data by the steps described in section 2.1.1. This dataset would be called NTUH miRNA array data from now on.

2.2 Databases related to miRNAs

2.2.1 miRBase

In this work, miRBase [32], containing information of miRNA names, accession number, sequence, and related citations, was used to provide a reference ID (miRBase accession number). If the miRNA identity used in a dataset is “miRNA name”, it should be converted, according to mature.fa data file, which contains name, accession number, and sequence of a miRNA, of miRBase data, to miRBase accession number before using it in the analysis pipeline. After all analysis was done, miRBase can also be used to convert miRBase accession number back to human-readable miRNA name.

2.2.2 TargetScan

TargetScan v6.0 [26, 33, 34] was one of the three miRNA target prediction databases used for miRNA target prediction collection in this work. There are totally 4,305,160 miRNA-target gene pairs in this database.

Two types of conservation were defined in TargetScan: conserved miRNA and conserved sites. To determine if a miRNA is conserved or not, we used “miR Family” table of TargetScan database. To determine if a target site is conserved, we used “Conserved site context+ scores” and “Nonconserved site context+ scores”.

2.2.3 miRanda

miRanda [35] (the version used here was released in August, 2010) was one of the three miRNA target prediction databases used for miRNA target prediction collection here. For human, there are totally 7,333,699 miRNA-target gene pairs in this database. The author separated the database into 4 parts according to the conservation of the miRNA and the prediction score (mirSVR score): good mirSVR score with conserved miRNA, good mirSVR score with non-conserved miRNA, non-good mirSVR score with conserved miRNA, and non-good mirSVR score with non-conserved miRNA.

2.2.4 PicTar

PicTar [36, 37] was one of the three miRNA target prediction databases used for miRNA target prediction collection here. PicTar data table, which was built upon reference genome hg17 and arranged as a data track in UCSC genome browser, was downloaded from UCSC genome browser, and of PicTar 5sp and 4sp, PicTar 5sp were used in this work. There are totally 56,058 miRNA-target gene pairs in PicTar 5sp. The “sp” in PicTar 5sp means “species”. In PicTar 5sp, conservation among five vertebrates, human, mouse, rat, dog, and chicken, were considered when building the database. We

used only ‘name’ column in downloaded PicTar 5sp data, since it contains NCBI accession number of target mRNA and miRNA name.

2.3 Construction of miRNA target prediction collection from TargetScan, MiRanda and PicTar

First, we applied different filters for each database. For TargetScan, we used miRNA-target gene pairs whose target site and miRNA were both conserved. For MiRanda, we used miRNA-target gene pairs marked as “high score and conserved miRNA” (which was classified by miRanda authors). For PicTar, we used all data included in the 5sp dataset. After we processed each of the target prediction databases, we combined them by selecting the miRNA-target gene pairs which were in 2 of 3 target prediction databases and removing the rest of the pairs.

2.4 Significance Analysis of Microarrays (SAM)

To get a list of differentially expressed genes and miRNAs between normal and tumor group, we used significance analysis of microarrays (SAM) [38] implemented in R package samr (version 2.0). We set $FDR \leq 0.0001\%$, fold change ≥ 2.5 for miRNA, and fold change ≥ 1.9 for genes as thresholds to reduce false positives. If one gene / miRNA were expressed higher in normal group than in tumor group, we defined that gene / miRNA as a down-regulated gene / miRNA, and *vice versa*.

2.5 Protein-protein interaction database

We used Human Protein Reference Database (HPRD) release 9 [39], which contains known protein-protein interaction data obtained by experimental methods and related citations, in the miRNA-regulated protein-protein interaction network construction process. There are 39,240 raw data, 38,923 irredundant interactions in this

database.

2.6 Construction of miRNA-regulated protein interaction network

To construct miRNA-regulated protein interaction networks (miRNA-regulated PINs), we combined expression profiles of miRNA and mRNA, and miRNA-target gene pairs provided by miRNA target prediction collection (method of collection construction was described in section 2.3). First, we picked significant miRNAs and genes with reverse trends and made them form miRNA-gene pairs (for example, down-regulated gene would be paired with up-regulated miRNA). Then, we removed the pairs which cannot be found in the miRNA target prediction collection. The miRNA-gene pairs left were defined as confident miRNA-gene pairs. By these confident miRNA-gene pairs, we could construct miRNA-regulated L0 PINs. L0 genes were thus defined by genes connected directly to the miRNAs.

Since number of direct targets of each miRNA was often small, we extended the miRNA-regulated L0 PINs by incorporating HPRD protein-protein interaction data: the interacting partners of L0 proteins were added into each miRNA-regulated PIN. Genes who were connected with L0 genes only were then defined as L1 genes.

2.7 Gene Ontology

We chose gene ontology (GO) [40] as the functional annotation used in this work. In GO, functions of a gene were described by a set of pre-defined, hierarchical terms. To obtain the GO annotations for proteins in miRNA-regulated PINs, we downloaded NCBI gene2go database, which contains genes (in gene ID) and their associated GO terms, from its FTP site (downloaded at Jan. 16, 2012). For human (*Homo sapiens*, NCBI taxonomy ID 9606), there were 17,967 unique genes, 12,314 unique GOs,

203,844 irredundant annotations in this database.

2.8 Hypergeometric Test and GO Enrichment Analysis

Sometimes we wish to determine if the ratio of individuals associated with some specific attributes in a designated group was higher than background (i.e. the population) or not. In other words, we want to test if the attributes were enriched in a specific set of individuals. In such situation, we can use chi-square test, Fisher's exact test, or hypergeometric test.

To determine if a GO term was enriched (i.e. overrepresented) in a miRNA-regulated PIN, we used hypergeometric test. P-value of hypergeometric test can be calculated by the following formula [41]:

$$\sum_{i=m}^{\min(n,m)} \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

where N is number of genes in the genome, m is number of genes annotated with specific GO term in the genome, n is number of genes in the miRNA-regulated PIN, k is number of genes annotated with specific GO term in the miRNA-regulated PIN.

In our enrichment analysis, we excluded miRNA-regulated PINs whose number of genes in the network were smaller than 5. We also excluded GO terms whose level were smaller than 5 to avoid general GO terms and use more specific GO terms. Level of GO terms was defined as the minimum distance between the term and its root term. (See **Figure 2** for an example.)

Since we tested multiple GO terms on each miRNA-regulated PIN, the significance of the test should be adjusted. Here, we used the false discovery rate method developed by Benjamini *et al.* [42], which was also implemented and used in

BiNGO [43] for multiple test adjustment, to calculate adjusted p-value for each GO enrichment test. Adjusted p-value can be calculated as follows:

$$p_i^{adj} = \min_{k=i-m} (\min(1, \frac{m}{k} p_k \sum_{g=1}^m \frac{1}{g})) \text{ for } i \text{ in } 1 \text{ to } m$$

where m is number of p-values which will be adjusted. We used adjusted p-value < 0.0001 as our threshold. The adjustment was run for each network: for each network, we run the GO enrichment analysis, collected the calculated p-values, and adjusted them with the formula above.

2.9 Activity analysis

It makes sense that when a miRNA is down-regulated in cancer, proteins in the PIN regulated by that miRNA should be expressed higher and functional, and vice versa. We define such PIN as activated. The concept was shown in **Figure 3**. We used Pearson correlation coefficient (PCC) as an index for co-expression. We can calculate PCC of expression between 2 genes with the following formula:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

where i represents each sample, n is number of samples, X_i and Y_i are expression of each gene in sample i, and r is Pearson correlation coefficient of the 2 genes.

To determine if a miRNA-regulated PIN was activated, we considered the proportion of co-expressed PPIs in the PIN. A co-expressed PPI (CePPI) was defined as the PPI whose interactants were highly co-expressed. To define thresholds for high and low co-expression discrimination, we first plot histograms, comparing distribution of PCC of the pairs formed by all genes in the gene microarray (All PCC) and of the PPI pairs in HPRD (PPI PCC). By comparing such graph, we can obtain the PCC, defined

as threshold A, when the proportion of PPI PCC was higher than All PCC. After that, we obtained threshold B by using the minimum value of top 5% of sorted All PCC, which was sorted in descending order. Finally we defined the final threshold as threshold B when threshold B was larger than threshold A. Threshold for normal and tumor samples was determined separately.

To determine if CePPI was enriched in a miRNA-regulated PIN, we use hypergeometric test, which was described in section 2.8.

2.10 ROC curve and ROC analysis

To see if the miRNAs we found can be served as a classification marker for discriminating between normal and tumor samples, we applied ROC analysis on the miRNA array data obtained from National Taiwan University Hospital (described in section 2.1.3). ROC analysis is usually used to evaluate the efficiency of a classifier or a biological marker.

R package ROCR [44] (version 1.0.4) was used to plot the ROC curve and calculate AUC (area under curve). After we obtained AUC, methods described in the work of Hanlye and McNeil was used to calculate standard error of AUC [45], which is listed below.

$$SE(\theta = AUC) = \sqrt{\frac{\theta(1-\theta) + (n_A - 1)(Q_1 - \theta^2) + (n_N - 1)(Q_2 - \theta^2)}{n_A n_N}}$$

where:

$$Q_1 = \frac{\theta}{2 - \theta}$$

$$Q_2 = \frac{2\theta^2}{1 + \theta}$$

To test if selected markers have diagnostic value, we applied a statistical test. In that test, the null hypothesis was that the AUC was smaller than or equal to as 0.5,

which means the marker was useless, since its performance was not better than random guess, and the alternative hypothesis was that the AUC was larger than 0.5. After calculating Z-score of AUC obtained from ROCR with the following formula, we can determine p-value of our hypothesis by the Z-score:

$$Z = \frac{AUC - 0.5}{SE(\theta = 0.5)}$$

2.11 Survival analysis with GOBO

To further validate if the PINs we found were related to cancer, we used survival analysis implemented in GOBO [46], which provides large amount of breast cancer gene expression profile collected from public databases with clinical outcome data. GOBO is available at <http://co.bmc.lu.se/gobo>. Kaplan-Meier survival analysis can be done by using a gene set as input. To select gene set for GOBO analysis, we choose genes which were in specific miRNA-regulated network and annotated with survival-related GO terms. Survival-related terms were defined as GO terms which include “proliferation”, “death”, “apoptosis”, “signaling”, “actin”, or “microtubule” in their GO term description text. Parameter “number of groups (quantiles)” was set to 2, “censoring (years)” was set to 10, “end-point” was set to “overall survival”, and no multivariate analysis were selected. Results of survival analysis would be included in N_GOBO_GeneSetAnalysis_KaplanMeier_Pvalues.txt (where N is an arbitrary ID), which is one of downloadable data files after running GOBO. In this work, we collected the p-values of survival analysis to determine if the gene set was related to survival in another clinical data.

Chapter 3 Results

3.1 Differentially expressed miRNAs and mRNAs

To construct miRNA-regulated PINs, we should extract differentially expressed miRNAs and genes from our dataset after we properly processed our expression profiles. We applied quantile normalization and scaling (as described in section 2.1.1) to our datasets. The distribution of normalized data can be seen in **Figure 4** and **Figure 5**. From our selected miRNA dataset, GSE29173, we found 89 down-regulated miRNAs (93 before filtered with fold change filter) and only 1 up-regulated miRNAs (**Table 1**). In gene expression dataset GSE29174, we found totally 1,268 down-regulated genes and 587 up-regulated genes before applying fold change filter. There were 726 down-regulated genes (**Table 2**) and 437 up-regulated genes (**Table 3**) after significantly and differentially expressed genes were filtered with the fold change (fold change > 2) filter.

3.2 miRNA-regulated PINs

After we appropriately processed miRNA target prediction databases and elucidated differentially expressed miRNAs and genes, miRNA-regulated PINs could then be constructed. We identified and constructed the partial network, containing miRNA and its direct target, with the differentially expressed miRNAs and genes by the methods described in section 2.3. We then extended the network by appending known interactions in HPRD database. Finally, 18 miRNA-regulated PINs were constructed by the steps described above (see **Figure 6** to **Figure 18** and **Table 5**).

3.3 Enriched GO Terms in the miRNA-regulated PINs

To elucidate functions of a miRNA with a regulated PIN, GO enrichment analysis

was applied to the miRNA-regulated PINs. We did not consider the miRNAs with 5 or less genes in its regulated PINs, and some of the miRNA-regulated PINs have no enriched functions in this threshold ($FDR < 0.0001$). To exclude GO terms that were too general, we only included GO terms which level was larger than 5.

We listed the result related to cancer in **Table 4**. We define a GO term cancer-related if a GO term contained “cell proliferation”, “cell death”, “apoptosis”, “signaling”, “microtubule”, and “actin”. One may notice that 7 miRNAs have enriched GO terms related to apoptosis, cell death, and cell proliferation: they are miR-520d-3p, miR-497-5p, miR-125b-5p, miR-21-5p, miR-31-5p, let-7c, and miR-125-5p. We may also find that some miRNA-regulated PINs may have functions other than cell survival. For example, nerve growth factor receptor pathway was enriched in miR-regulated PINs of miR-520d-3p, miR-497-5p, miR-125a-5p, miR-125b-5p, and 31-5p, and epidermal growth factor receptor pathway was enriched in miR-regulated PINs of miR-520d-3p, miR-21-5p, and miR-497-5p.

3.4 Activity of the miRNA-regulated PINs

To do enrichment analysis for CePPI, we should first determine the threshold of PCC for CePPI. The distribution histograms were in **Figure 19**. From the figure, when PCC was larger than 0.4~0.5, the proportion of PPI PCC was larger than All Paired PCC, and when PCC was larger than 0.1~0.2, the proportion of PPI PCC was larger than All Paired PCC. To actually determine the threshold, we need to elucidate the minimum of top 5% PCCs. From the distribution of All Paired PCC, the minimum of top 5% PCCs was 0.69522 for normal samples, and 0.2788 for tumor samples. Since 0.69522 is larger than 0.5 for normal samples, and 0.2788 is larger than 0.2 for tumor samples, we set 0.69522 as threshold for normal CePPI, and 0.2788 for tumor CePPI.

After thresholds were determined, we conduct the CePPI enrichment analysis. The p-values of activation status were shown in **Table 7**. From **Table 7**, we found that 7/18 of the miRNA-regulated PINs was activated in normal, while 3/18 of the PINs was activated in tumor.

3.5 Validation

3.5.1 ROC curve and ROC analysis on NTUH miRNA array data

We used ROC analysis to validate the classification efficiency when our miRNAs were selected as a biomarker (i.e. the diagnostic value of the miRNA). The result of ROC analysis, including p-value testing if $AUC > 0.5$, was listed in **Table 8** and **Figure 20**. From **Table 8** and **Figure 20**, we found that most of the miRNAs had good performance when used as diagnostic marker.

3.5.2 Validation of functional analysis by GOBO

To verify the functions found in GO enrichment analysis, we use survival analysis on the genes set which is annotated with GO terms related to cancer, such as apoptosis and cell proliferation, by GOBO. The results were in **Table 9** (full list) and **Figure 21** (significant part only). As one may notice from the result, most of the gene sets were not related to overall survival significantly.

Chapter 4 Discussions

4.1 Construction of miRNA-regulated PINs

Before constructing miRNA-regulated PINs, we should first find out significantly expressed genes and miRNAs. From the results of SAM analysis (**Table 1**), we could see some well-known breast cancer related miRNAs. For example, hsa-miR-214-3p [47] and hsa-miR-335-5p [48] were known to be down-regulated in previous literatures. Hsa-let-7c, one of let-7 families [49], was found to be down-regulated in previous work. We also found that miR-21-5p, the sole up-regulated miRNAs in our list, was also found to be up-regulated previously [18, 50]. However, most of miRNAs in our down-regulated list were not known to be changed in previous literature. Those miRNAs might be false positives, however, we could not rule out the possibility that they were novel breast cancer-related miRNAs.

The construction step relied on not only differentially expressed miRNAs and genes but also information in the target prediction databases. Also, the preprocessing step on the target prediction database would play a role in our result. Our current preprocessing parameters, using only the predictions presented in at least 2 of 3 databases, were to reduce some possible false positives in the target prediction databases. This setting would also remove some novel miRNAs, since they were too new to be included in the prediction dataset, thus not suitable for functionary discovery of novel miRNAs. However, this problem could be settled by using only the target prediction database with novel miRNAs instead of requiring that a prediction should exist in at least 2 databases.

After construction of the 18 PINs was completed (**Figure 6** to **Figure 18**), we found that size of the PINs were not similar: some of the miRNAs seems to regulate

large size of PINs, while other miRNAs affect only small number of genes. Small miRNA-regulated PINs may be caused by the strict q-value threshold set on SAM analysis, by the processing steps performed on the target prediction databases discussed before, and possibly by lack of protein-protein interaction data of some proteins in HPRD. Although HPRD was known to be the most comprehensive source of protein-protein interaction data [51], there were some proteins not considered and researched by other investigators, and interaction data for those proteins would not be in HPRD. However, it may be true that some of the miRNA-regulated PINs did be small in breast cancer, since the construction of the PINs were based on differentially expressed miRNAs and genes between normal tissues and tumor samples, and those miRNAs with small PINs may not be as important as others with larger PINs.

We further used activity analysis to check if the miRNA-regulated PINs were activated in certain conditions. In the process of threshold determination, we found that the shape of PCC distributions were different between normal and tumor samples. This may be related to the difference of number of samples (which can affect shape of PCC distribution): we had 11 normal samples, 110 tumor samples in our microarray data (GSE29174). For those miRNAs which is down-regulated in tumor tissue, the PIN it regulates would be activated, due to the repression of the miRNAs became lower, and vice versa. However, only 3 (total 17) PINs of down-regulated miRNAs were activated in tumor. In addition, 7 (total 17) PINs of down-regulated miRNAs were activated in normal condition. This result may suggest participation of other biomolecules, such as transcription factors, which could affect the expression of the PINs, thus we could not observe PIN activation in tumor condition for down-regulated miRNAs.

4.2 Functional analysis of miRNA-regulated PINs

To elucidate functions of miRNA-regulated PINs, we conducted GO enrichment analysis on the PINs (for a list of cancer-associated enriched GO terms, see **Table 6**). The first thing we noticed was the miRNA-regulated PINs enriched with apoptosis and / or cell proliferation related GO terms. Such miRNAs were miR-520d-3p, miR-497-5p, miR-125b-5p, miR-21-5p, miR-31-5p, let-7c, and miR-125-5p. Most of the miRNAs were discussed and researched before and known to be related to breast cancer. mir-125b-5p, which was down-regulated in tumor tissue, was previously known as a tumor suppressor [22], and such finding was consistent to our functional enrichment results. mir-125a-5p-regulated PIN was found to have anti-apoptosis activity and the ability to regulate epithelial cell proliferation, and the miRNA was known to repress cell growth [52].

We also found that some breast cancer specific functions were enriched in our results. For example, in miR-497-5p-regulated PIN, function “androgen receptor signaling pathway” was enriched. Although it is not clear whether androgens were related to breast cancer, androgen receptors were known to be up-regulated in breast cancer and relate to node invasiveness [53].

To further verify the findings of enriched cancer-related functions, we used GOBO for survival analysis. Our hypothesis is that expression of genes annotated with enriched cancer-related terms may be related to survival outcome of the patients. Except some cell death / proliferation related terms, we knew that some pathways or functions are also related to clinical outcomes. For example, cell proliferation-related GO terms have high probability to affect survival of cancer tissue. The patient may become worse if cancer tissue survives. In addition, some signaling pathways were known to enhance

invasiveness, migration abilities, or even decrease the survivability of the patient. For example, BMP signaling pathway was known to confer various tumor cells with migration and invasion ability [54]. NGFR was found to be associated with overall survival of breast cancer [54]. Furthermore, Toll-like receptor 4 was known to promote adhesion and invasive migration in breast cancer [55]. Finally, cytoskeleton is an important part for every kind of cells in cell mobility. Actin filament was known to be participated in the process of invasive migration of cancer cells [56]. Since we had some of these functions in our enriched terms, we wished to test if the expression of gene set annotated to the cancer related enriched terms in the PIN would be related to clinical outcome of patients.

Our results showed in **Table 9** and **Figure 21**. As one can see, only some of the enriched terms were associated significantly with clinical outcome. It may be that the changes of these key genes occurred on protein level, such as protein expression or post-translational modification, so association of such genes to clinical outcomes cannot be explored by tools like GOBO. It may also because the miRNA did not regulate whole pathway, or the miRNA did not target the key part of the pathway directly, thus we cannot find the clinical outcome of gene sets of the enriched GO terms in such condition. However, we did find some functions associated with clinical outcomes. For example, enriched term “regulation of epithelial cell proliferation” of both miR-125a-5p and miR-125b-5p were found to be associated with overall 10 years survival of patients. This result acted as evidence that further supported GO enrichment analysis and the discussion before.

Since the miRNAs of the miRNA-regulated PINs were differentially expressed between normal and tumor tissue, and we found some cancer related functions in our functional enrichment analysis, the miRNAs may be used as diagnostic marker for

breast cancer. To verify this, we applied ROC curve analysis on the miRNA expression profile which was obtained from NTU Hospital and was not used in construction process of miRNA-regulated PINs. Notably, our results showed that miR-497-5p, miR-125b-5p, along with some other miRNAs of miRNA-regulated PINs, had good performance when used as diagnostic marker. The results were not beyond our expectation, since miR-497-5p and miR-125b-5p were found to have cancer-related functions in this work and in literature (discussed before).



Chapter 5 Conclusions

In this work, we obtained confident miRNA-target pairs by combining differentially expressed miRNAs and mRNAs elucidated from expression profiles and popular target prediction databases. By incorporating known protein interaction data in HPRD and confident miRNA-target pairs, breast cancer specific miRNA-regulated networks were constructed. After network construction, putative functions of the miRNA-regulated networks were discovered by GO enrichment analysis. From the result, we found previously known miRNAs which were enriched with cancer-related terms, such as “apoptosis” and “cell proliferation”. Novel breast cancer related miRNAs and putative functions of their PINs were also discovered in our work. We also validated some of our results by ROC curve analysis, survival analysis (GOBO), and literature search. This work may contribute to further research of breast cancer and related miRNAs, since some novel miRNAs associated with breast cancer were found. Furthermore, it is also possible to apply the method used in this work to other cancers for novel cancer-associated miRNA discovery and functional elucidation.

FIGURES

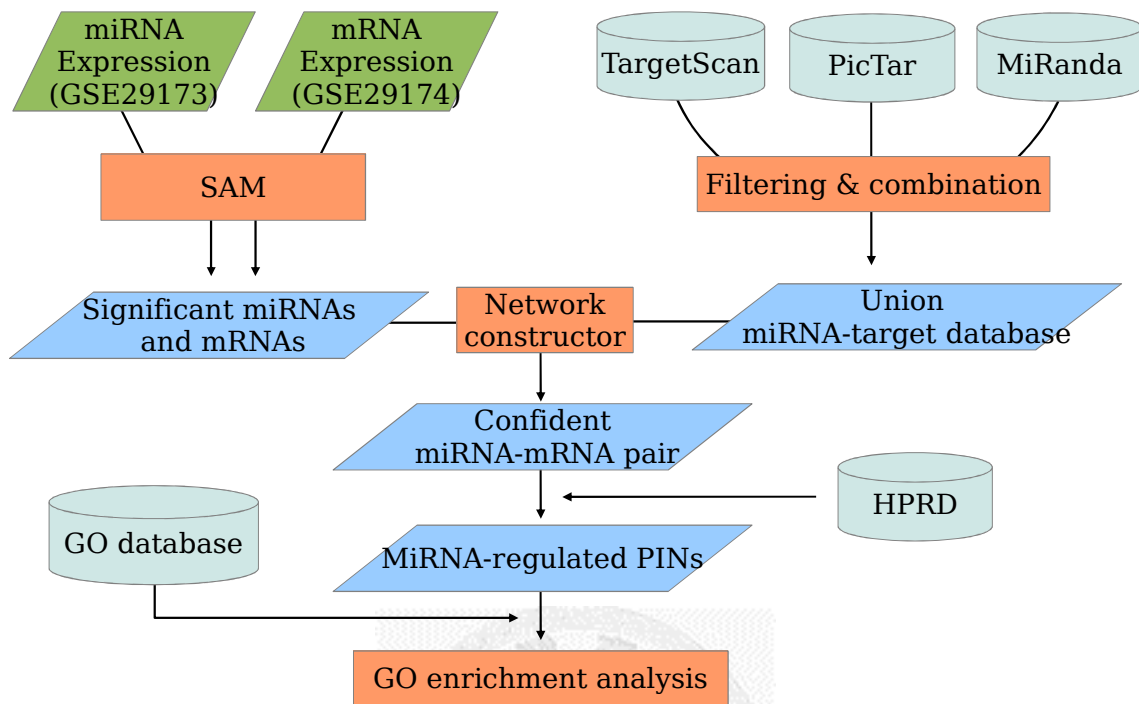


Figure 1. Pipeline for miRNA-regulated PINs construction and functional analysis.

Green: external dataset; **cyan:** external database; **blue:** dataset generated in the analysis steps; **orange:** analyzing steps which do the calculations or preprocessing.

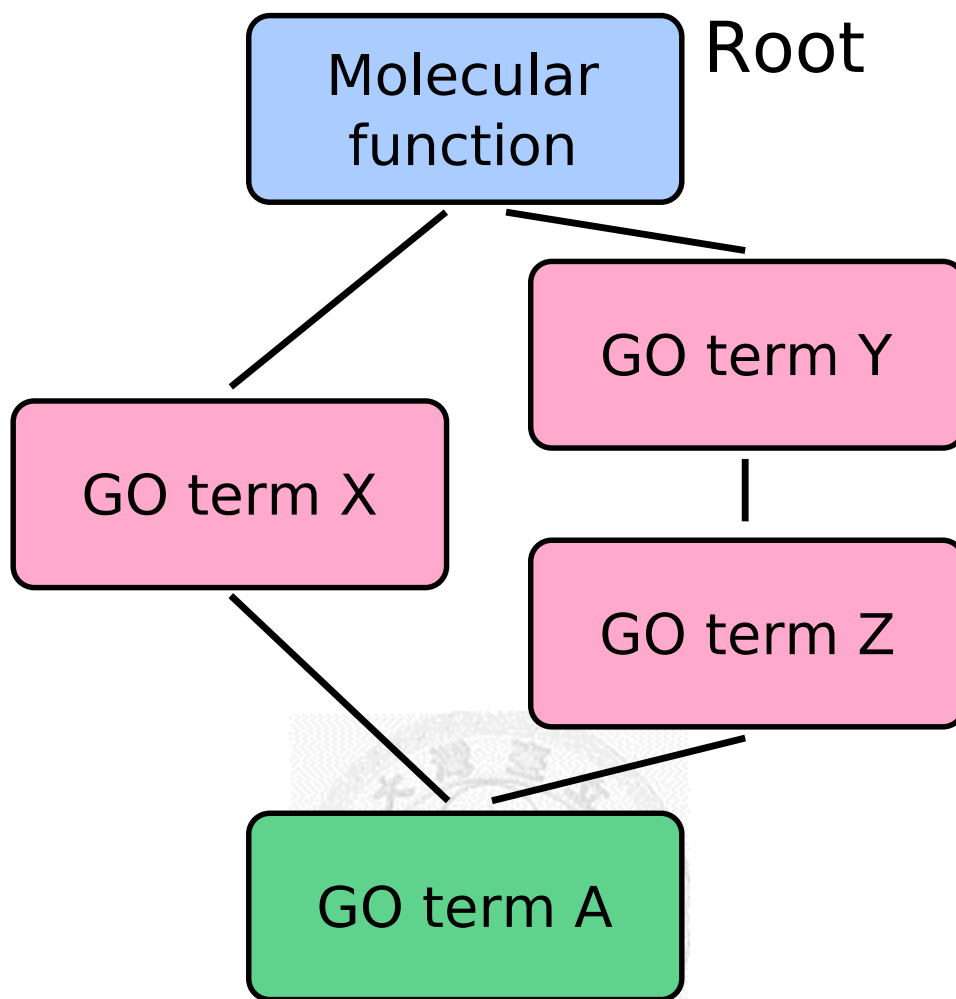


Figure 2. Definition of GO level in this work.

GO level was defined as the minimum distance between the term and its root node. Thus, GO level of GO term Y and term X is 1; GO level of GO term is Z; GO level of GO term A is also 2, since 2 is the minimum distance of the 2 possible paths to its root node.

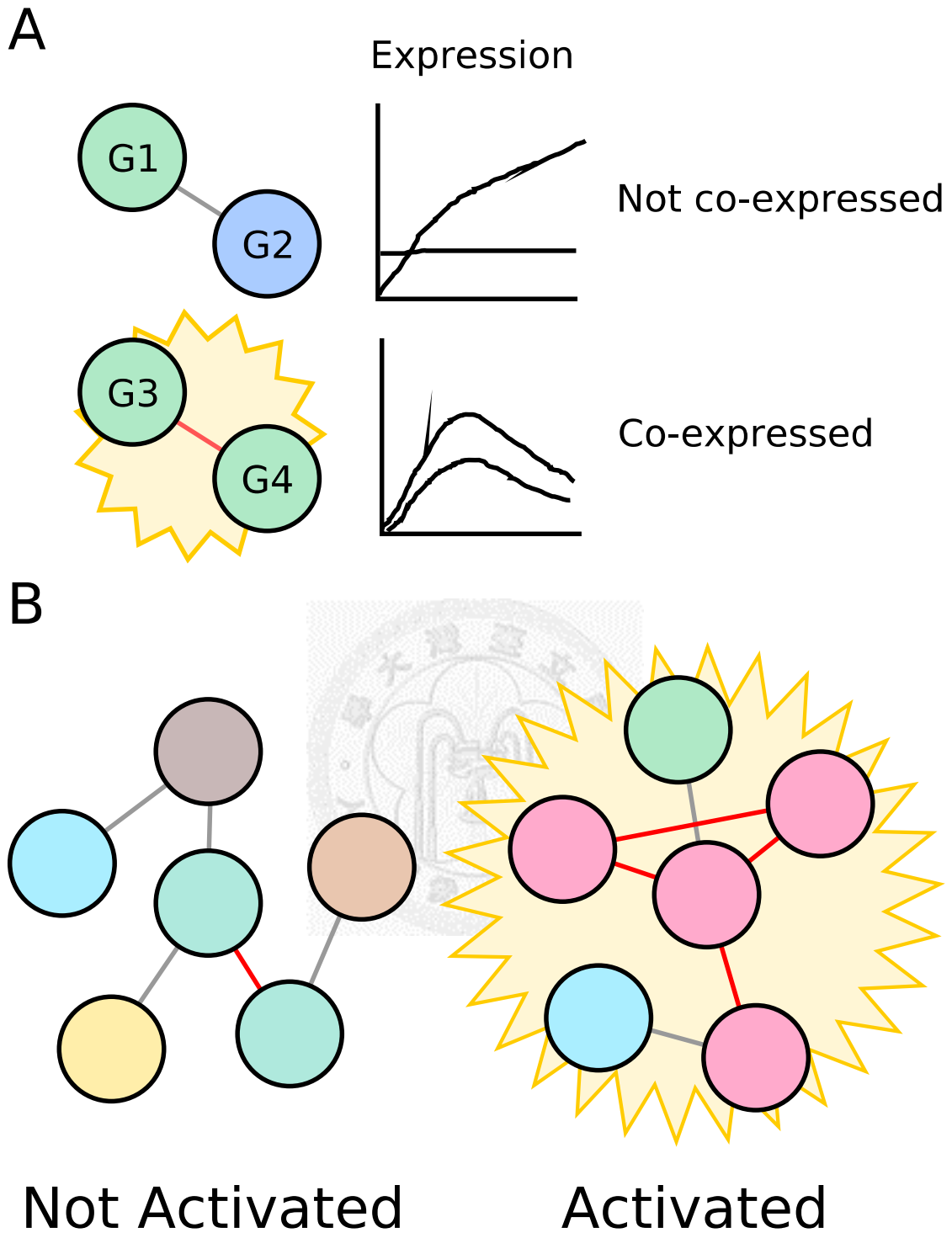
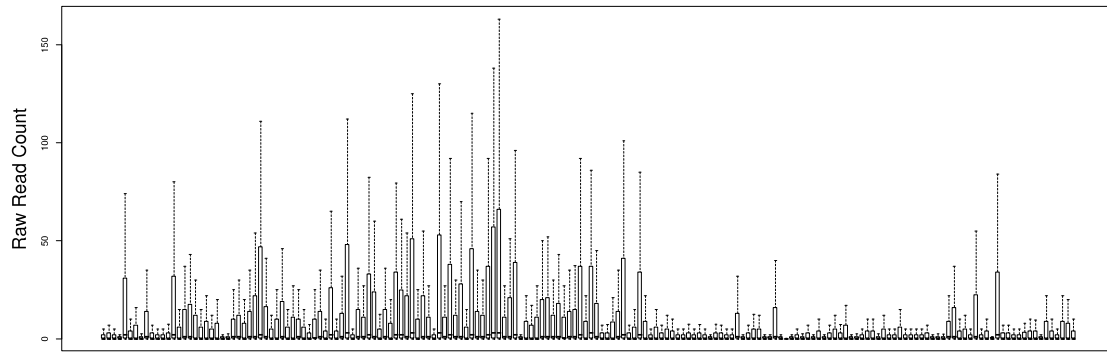


Figure 3. Schematic figure describing activity analysis.

Red edge stands for co-expressed PPI. **A:** Co-expression means that expression of two interacting gene are positively correlated. **B:** When the co-expressed interactions are enriched in a network (right part), we define it to be activated.

A



B

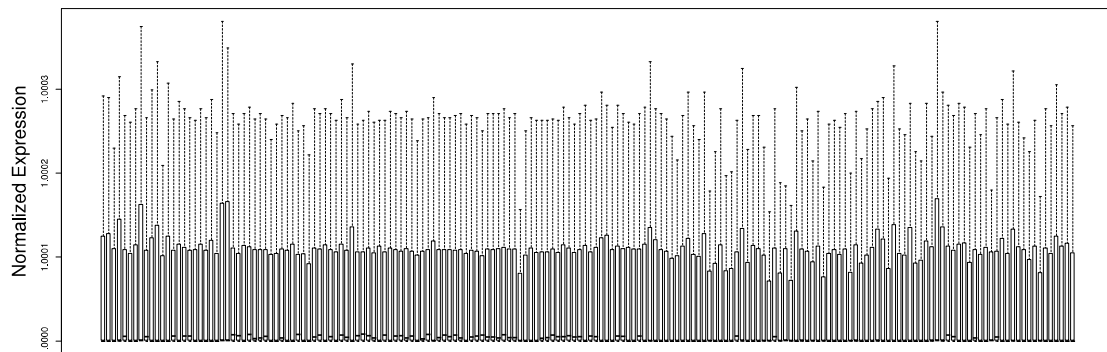
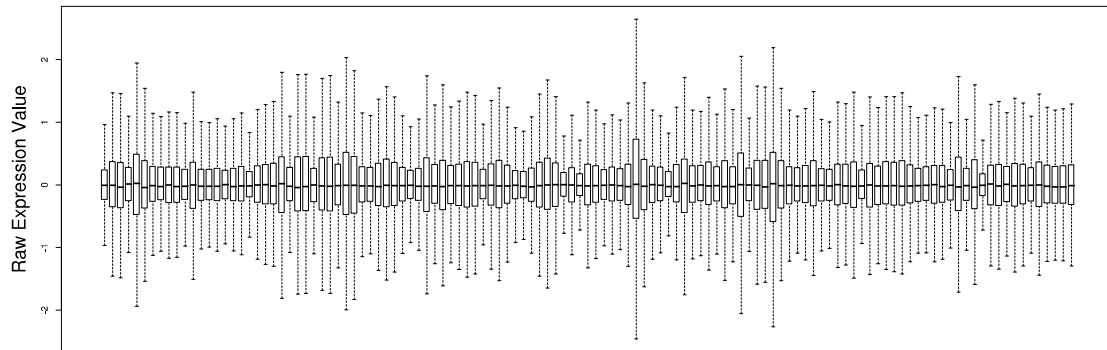


Figure 4. Distribution of miRNA dataset GSE29173.

Each bar represents data distribution of a single sample. **A**: Distribution, presented in boxplot, of the original dataset. **B**: Distribution of the dataset after normalization steps.

A



B

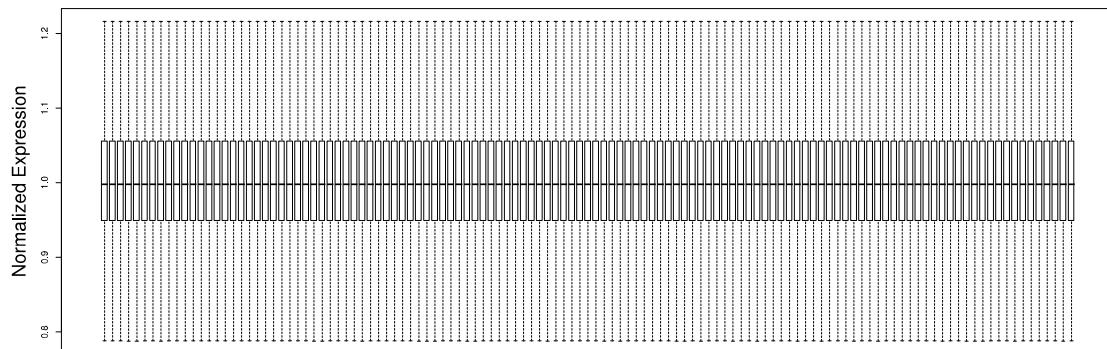


Figure 5. Distribution of gene expression dataset GSE29174.

Each bar represents data distribution of a single sample. **A**: Distribution, presented in boxplot, of the original dataset. **B**: Distribution of the dataset after normalization steps.

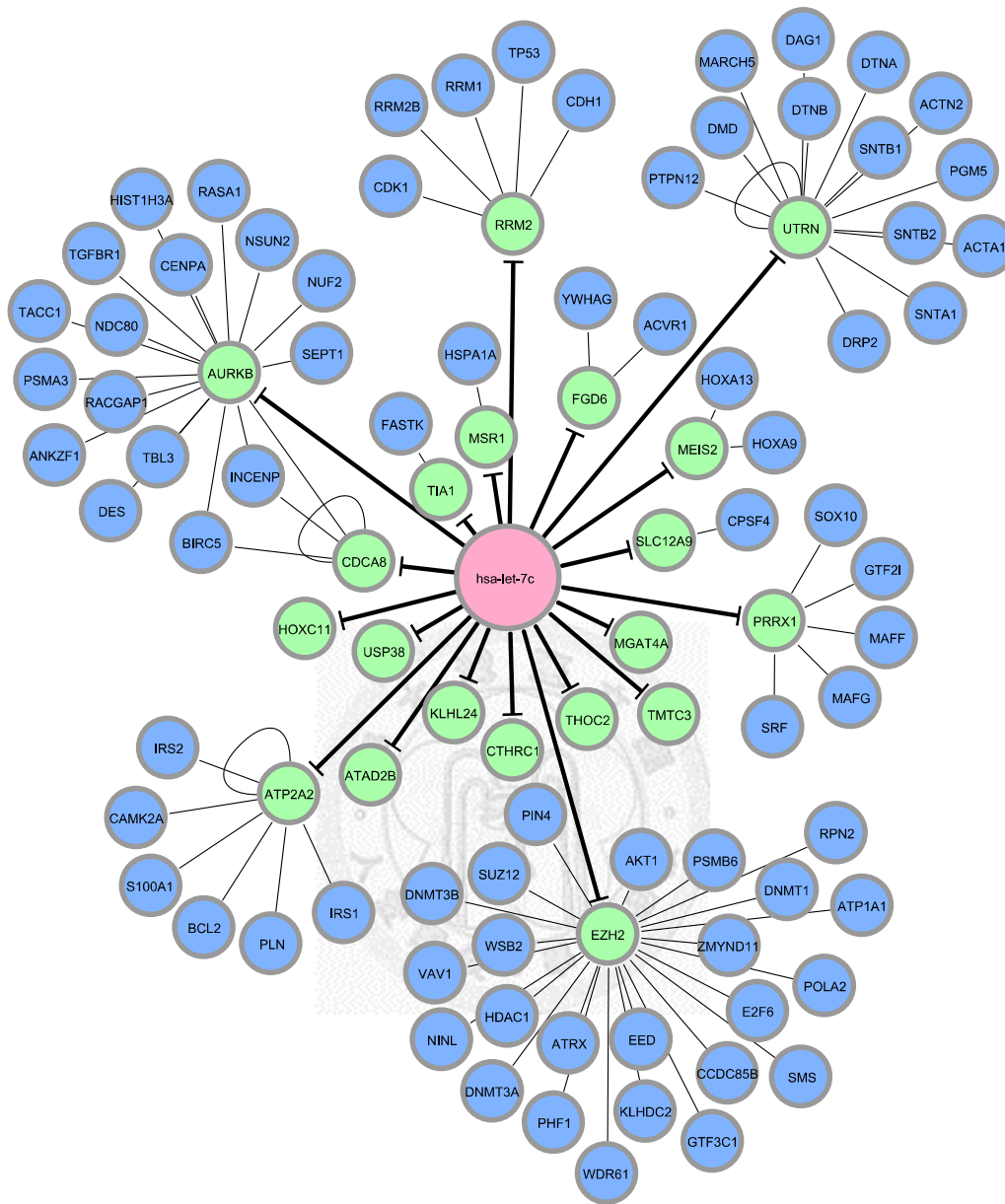


Figure 6. miRNA-regulated PIN (hsa-let-7c).

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

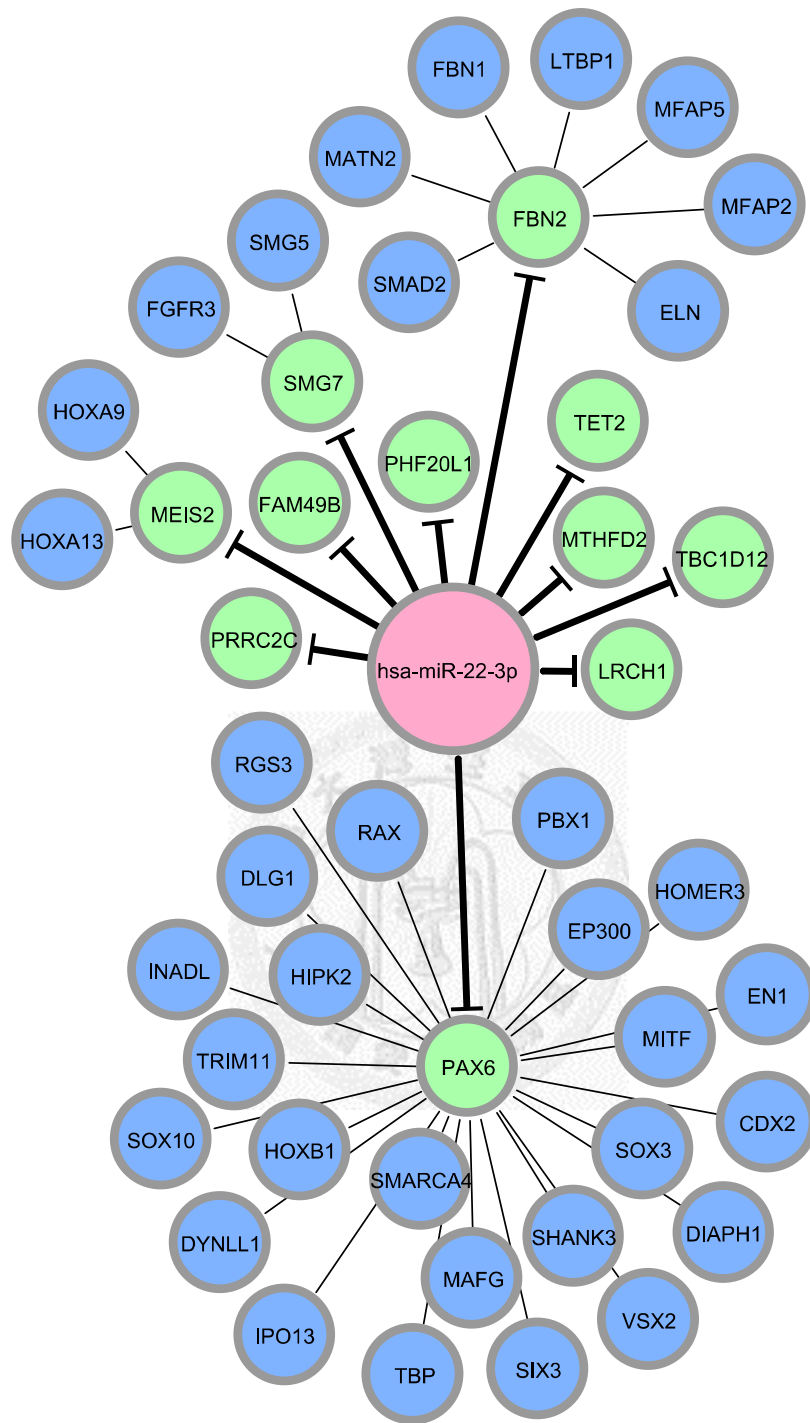


Figure 8. miRNA-regulate PIN. (hsa-miR-22-3p)

Red: miRNA; **green**: direct target of the miRNA; **blue**: proteins interacting with direct target of the miRNA.

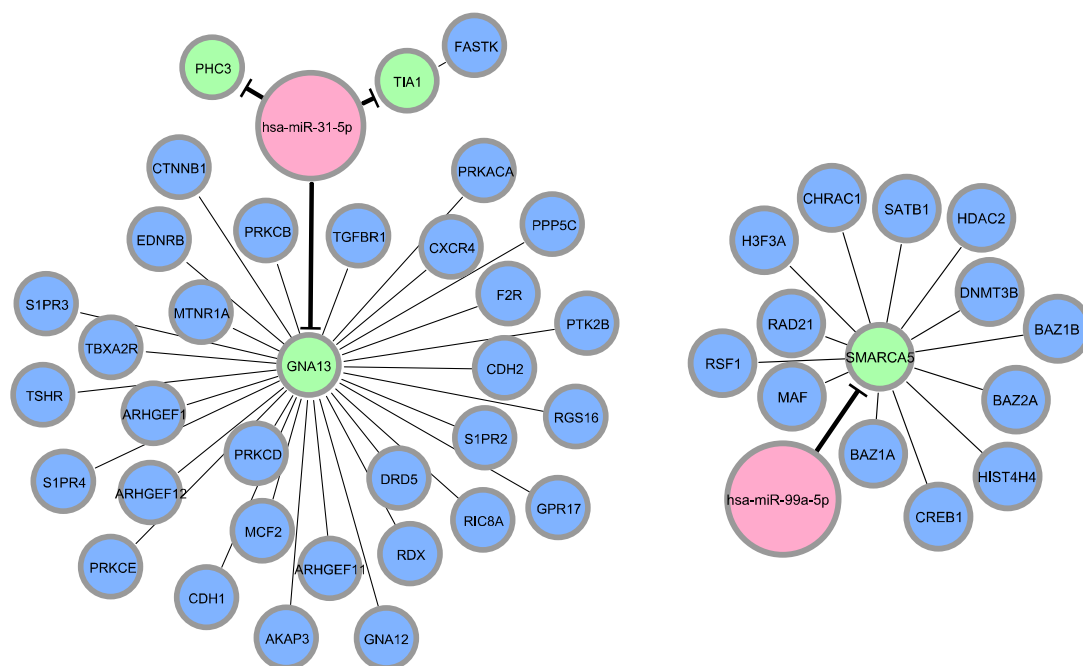
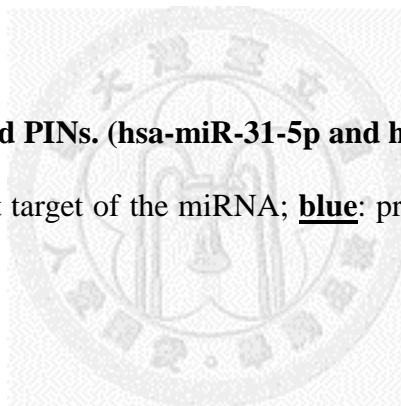


Figure 9. miRNA-regulated PINs. (hsa-miR-31-5p and hsa-miR-99a-5p)

Red: miRNA; **green**: direct target of the miRNA; **blue**: proteins interacting with direct target of the miRNA.



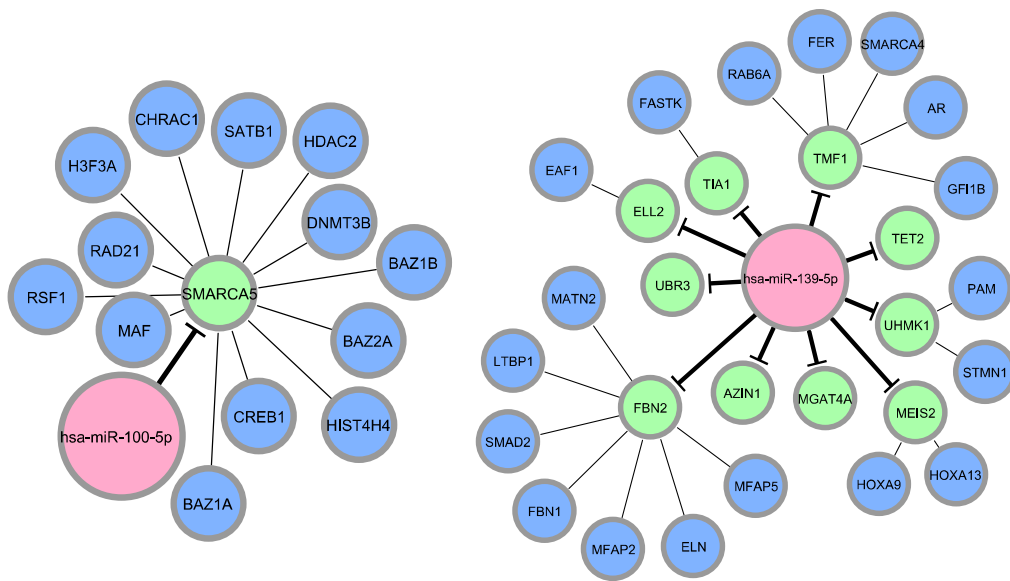
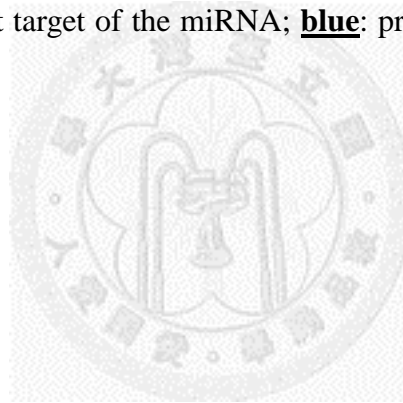


Figure 10. miRNA-regulated PINs. (hsa-miR-100-5p and hsa-miR-139-5p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.



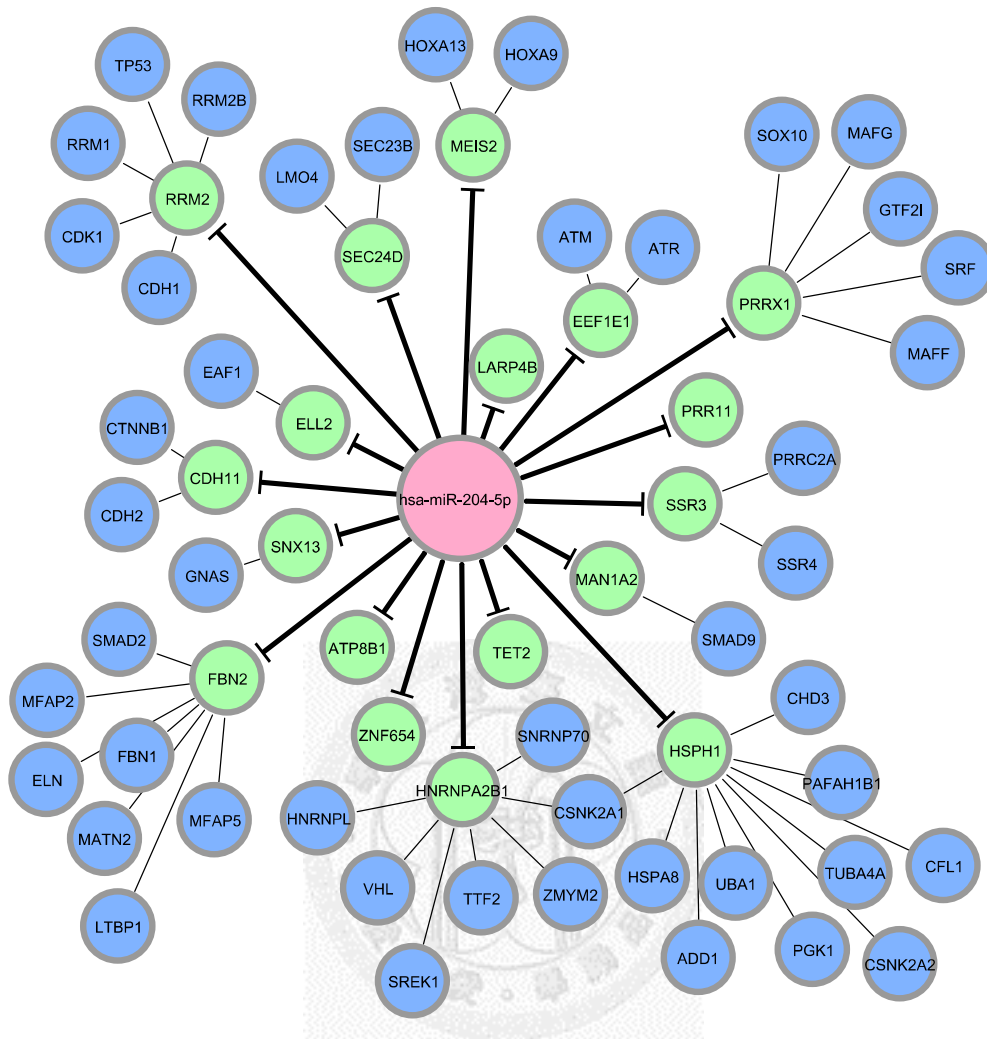


Figure 11. miRNA-regulated PIN. (hsa-204-5p)

Red: miRNA; **green**: direct target of the miRNA; **blue**: proteins interacting with direct target of the miRNA.

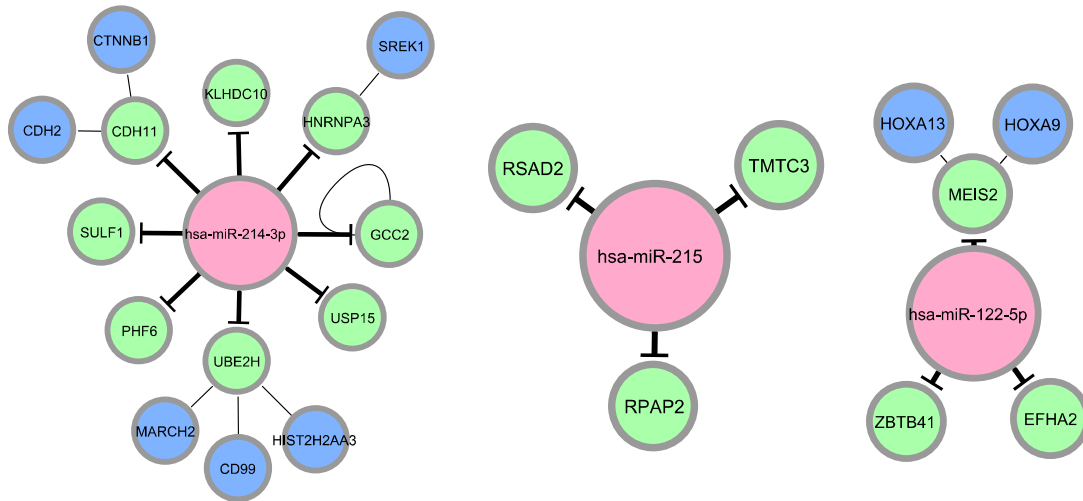
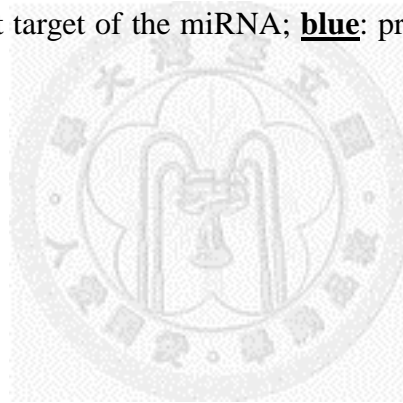


Figure 12. miRNA-regulated PINs. (hsa-miR-214-3p, hsa-miR-215, and hsa-miR-122-5p)

Red: miRNA; **green**: direct target of the miRNA; **blue**: proteins interacting with direct target of the miRNA.



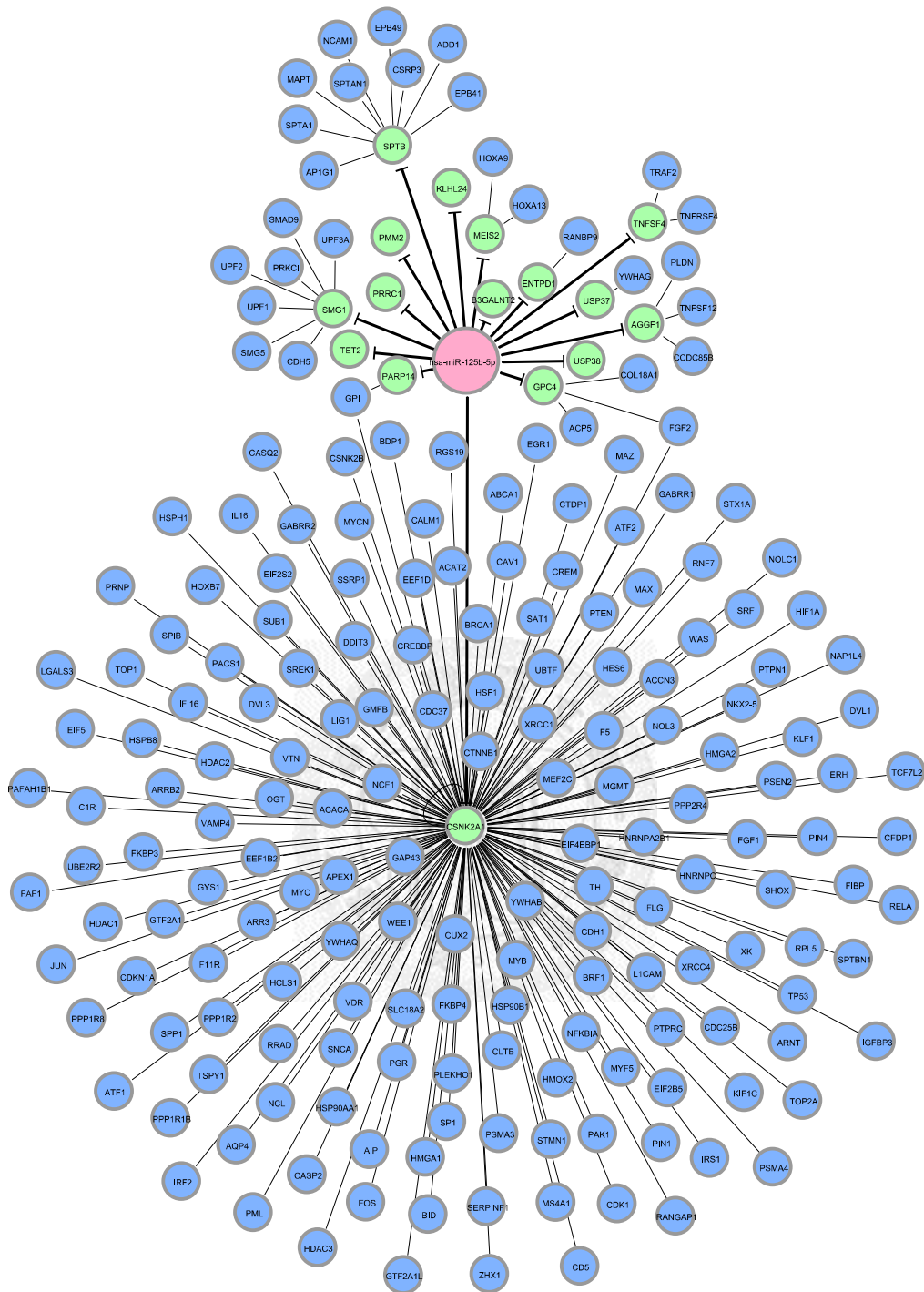


Figure 13. miRNA-regulated PIN. (hsa-miR-125b-5p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

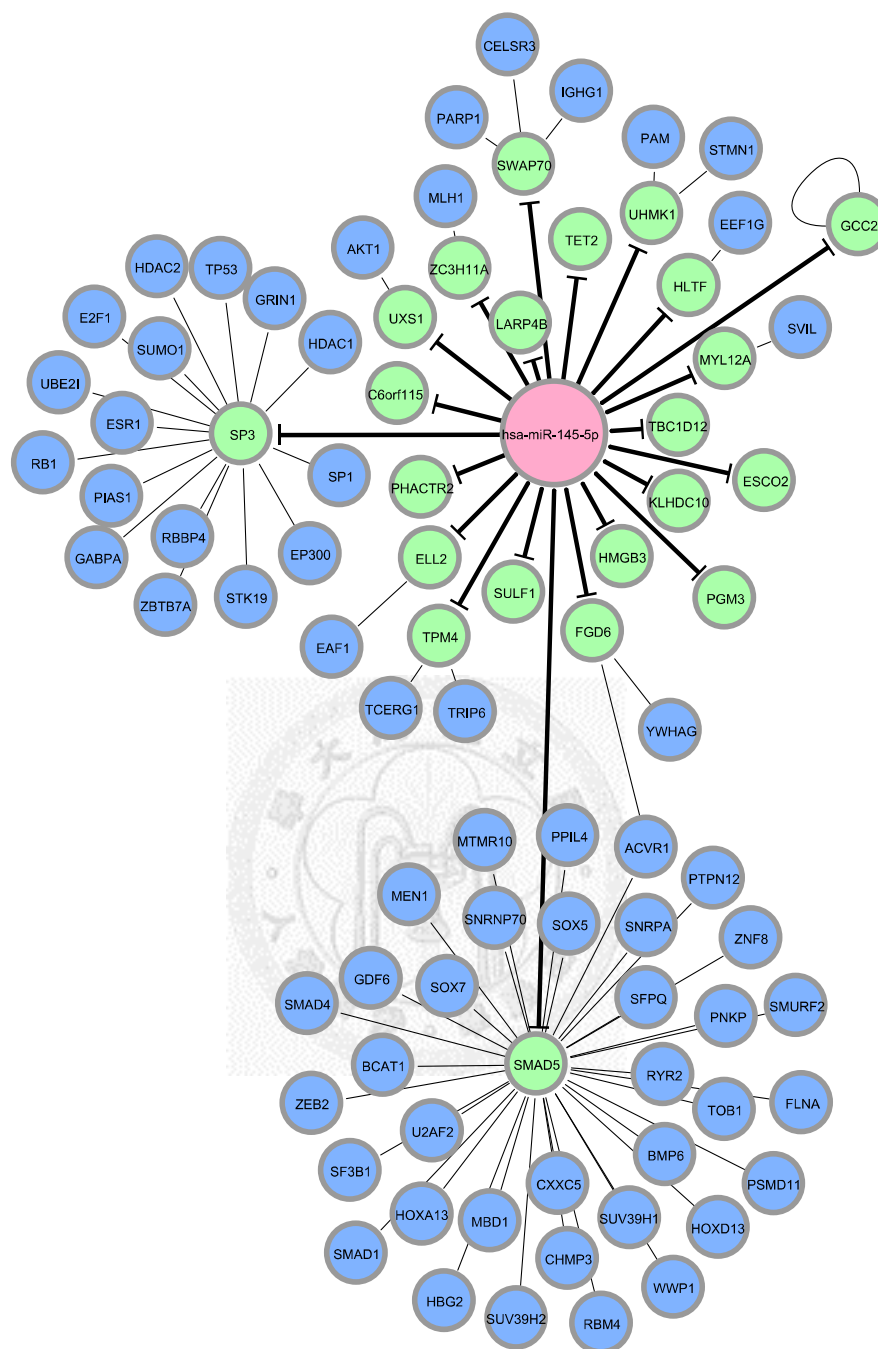


Figure 14. miRNA-regulated PIN. (hsa-miR-145-5p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

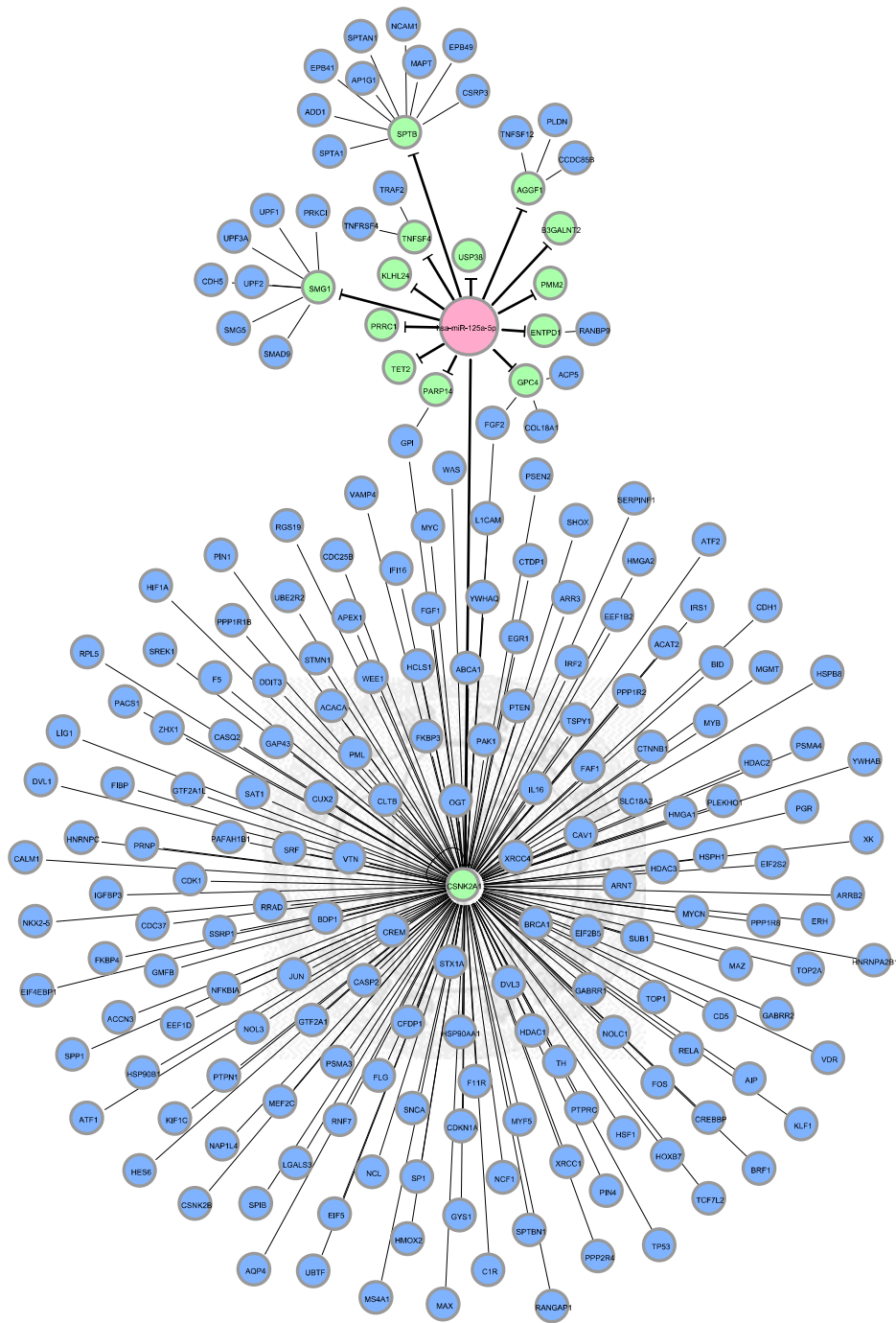


Figure 15. miRNA-regulated PIN. (hsa-miR-125a-5p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

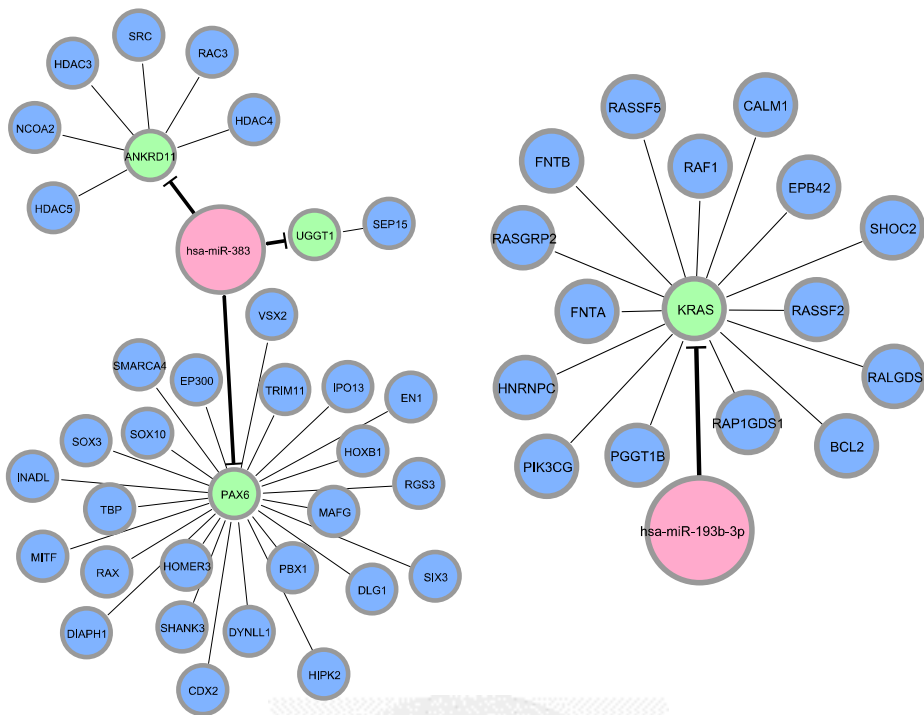


Figure 16. miRNA-regulated PINs. (hsa-miR-383 and hsa-miR-193b-3p)

Red: miRNA; **green**: direct target of the miRNA; **blue**: proteins interacting with direct target of the miRNA.

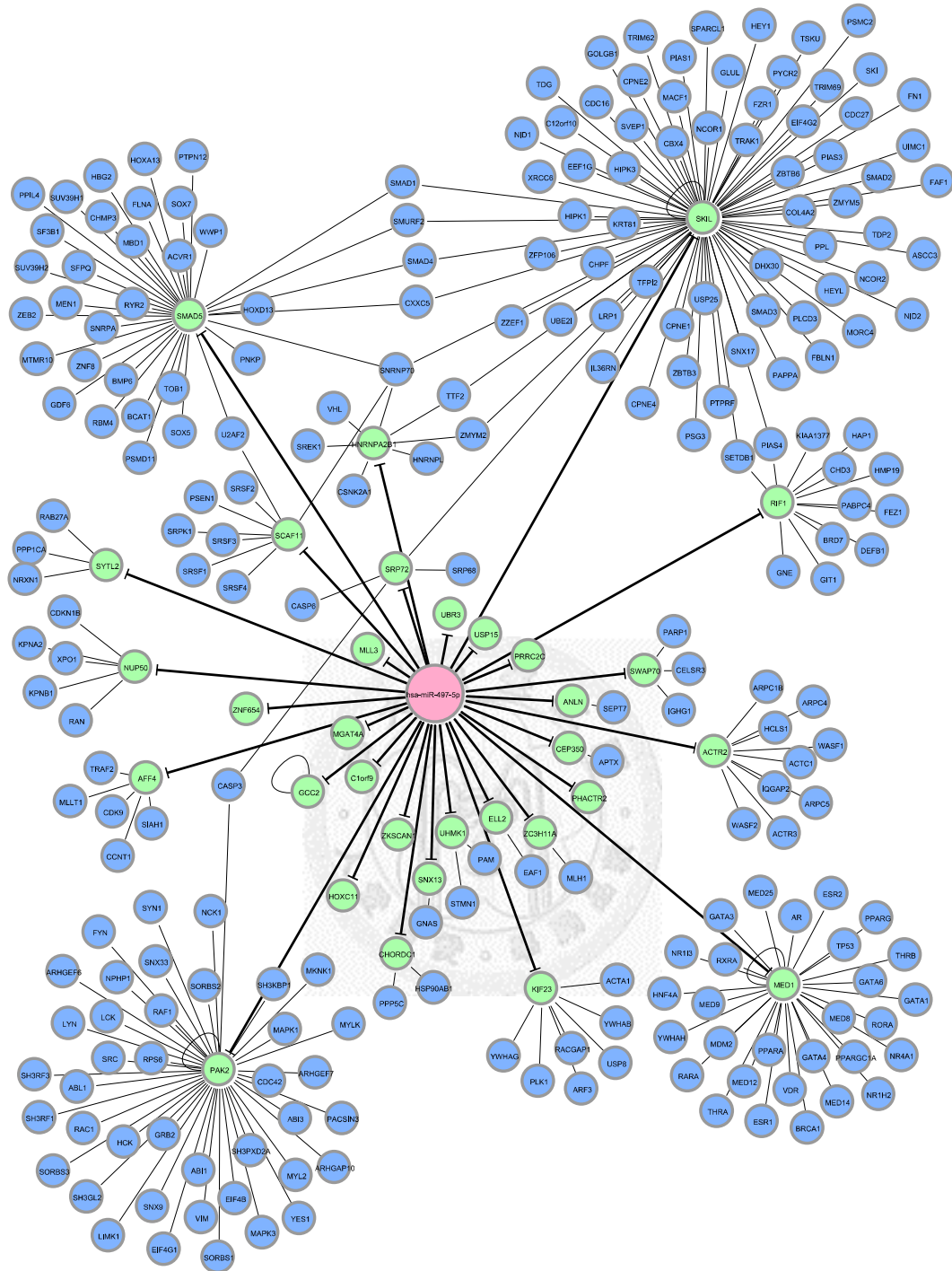


Figure 17. miRNA-regulated PIN. (hsa-miR-497-5p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

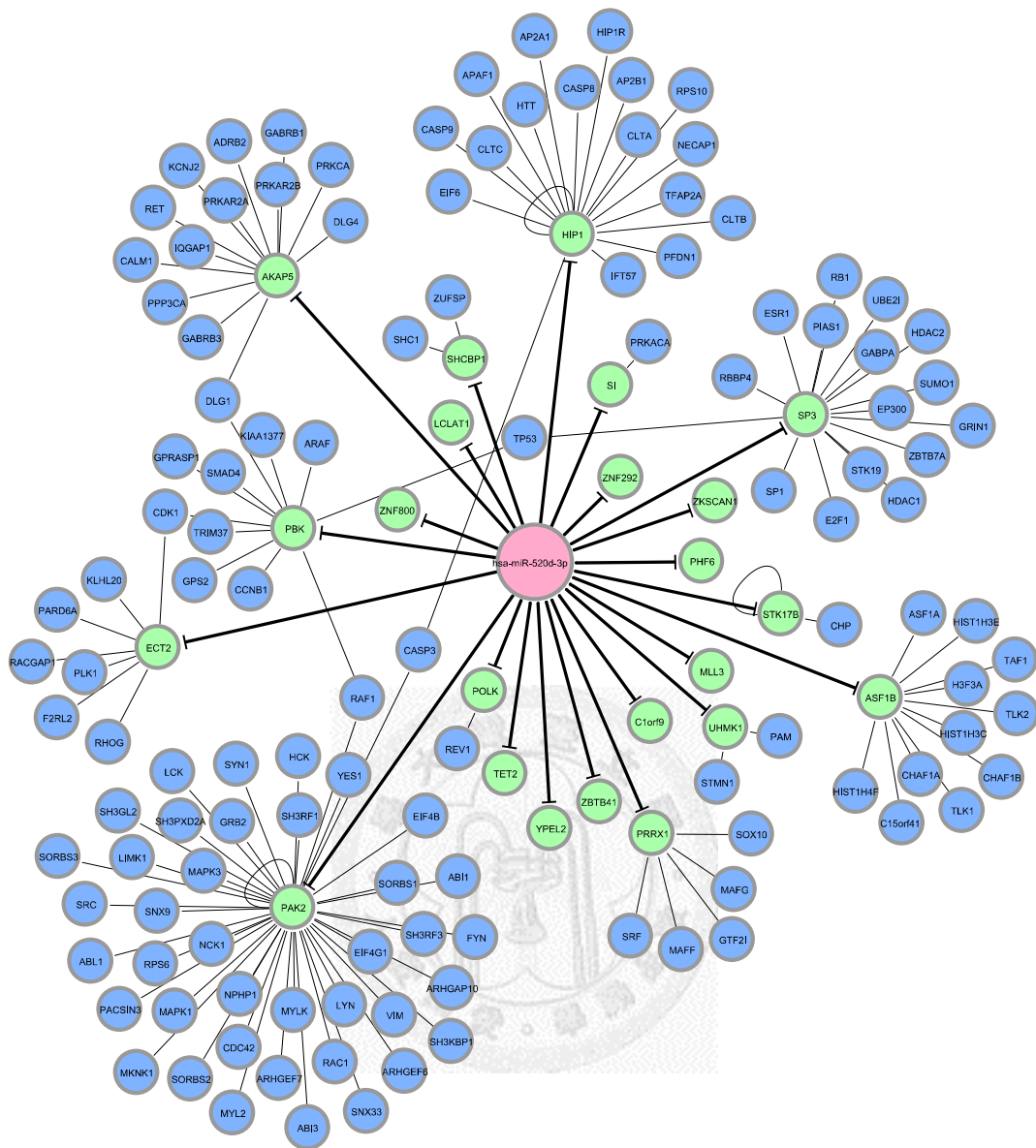


Figure 18. miRNA-regulated PIN. (hsa-miR-520d-3p)

Red: miRNA; **green:** direct target of the miRNA; **blue:** proteins interacting with direct target of the miRNA.

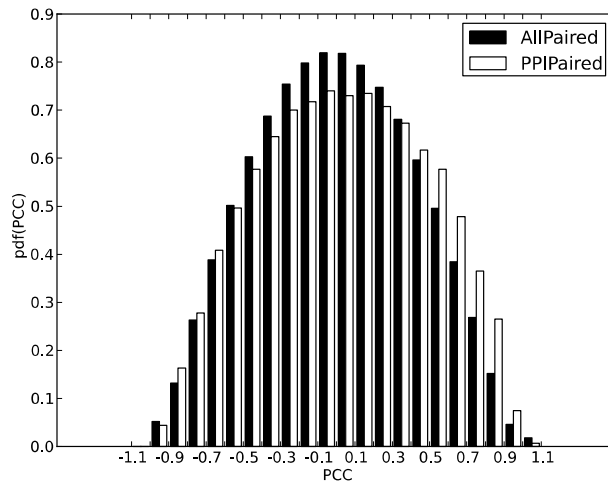
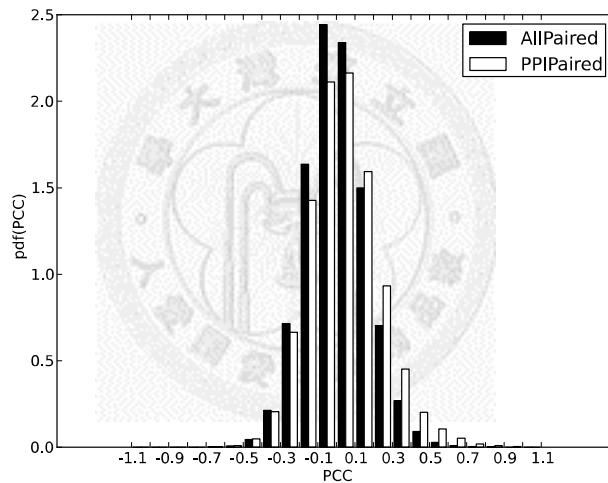
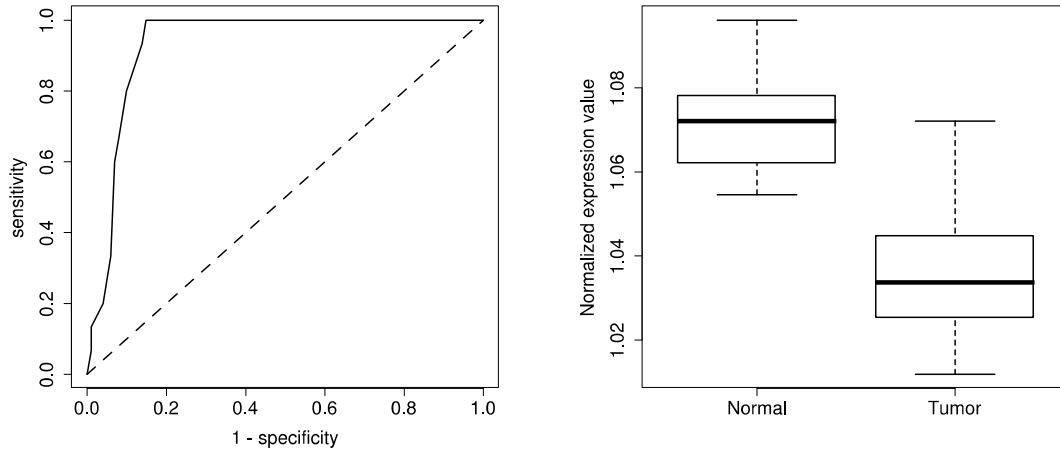
A**B**

Figure 19. Distribution of Pearson correlation coefficient of gene pairs in microarray data.

Here, PCC is the expression correlation of two genes. In normal sample, from when $PCC > 0.4 \sim 0.5$, the proportion of PIPaired is greater than AllPaired. In tumor samples, from when $PCC > 0.1 \sim 0.2$, the proportion of PIPaired is greater than AllPaired. **A**: Distribution of PCCs calculated by normal samples. **B**: Distribution of PCCs calculated by tumor samples. **AllPaired**: Distribution of PCCs of gene pairs formed by any two genes in the gene microarray. **PIPaired**: Distribution of PCCs of HPRD PPIs.

A. hsa-let-7c. MIMAT0000064. ***



B. hsa-miR-21-5p. MIMAT0000076. ***

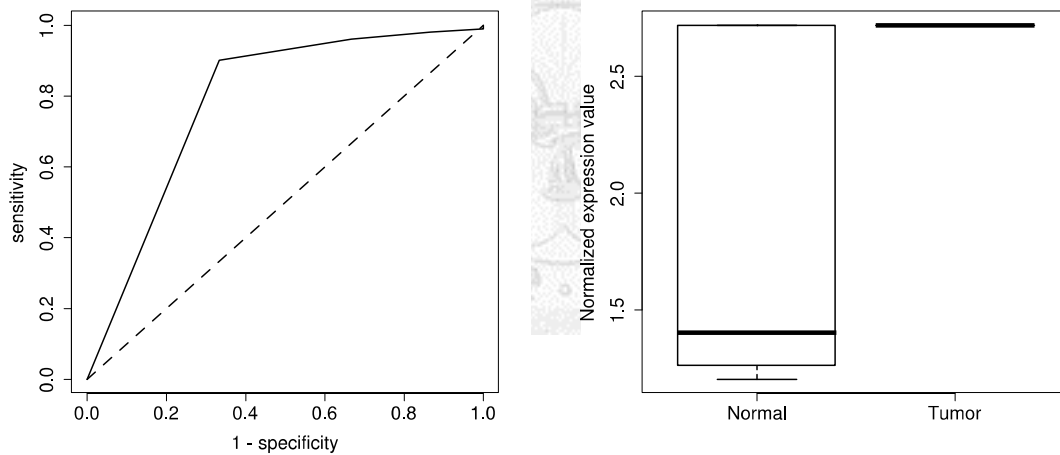
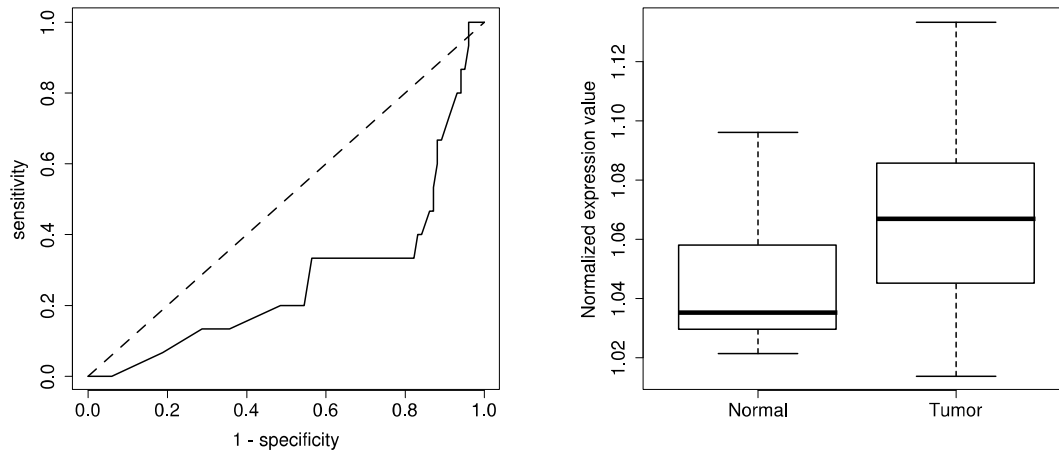


Figure 20. ROC curves and expression distribution of miRNAs.

Expression value from NTUH miRNA dataset was used for validating our results. **Solid line**: actual prediction performance of the miRNA as diagnostic marker; **dashed line**: AUC = 0.5 (no predictive power). *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

C. hsa-miR-22-3p. MIMAT0000077.



D. hsa-miR-31-5p. MIMAT0000089. ***

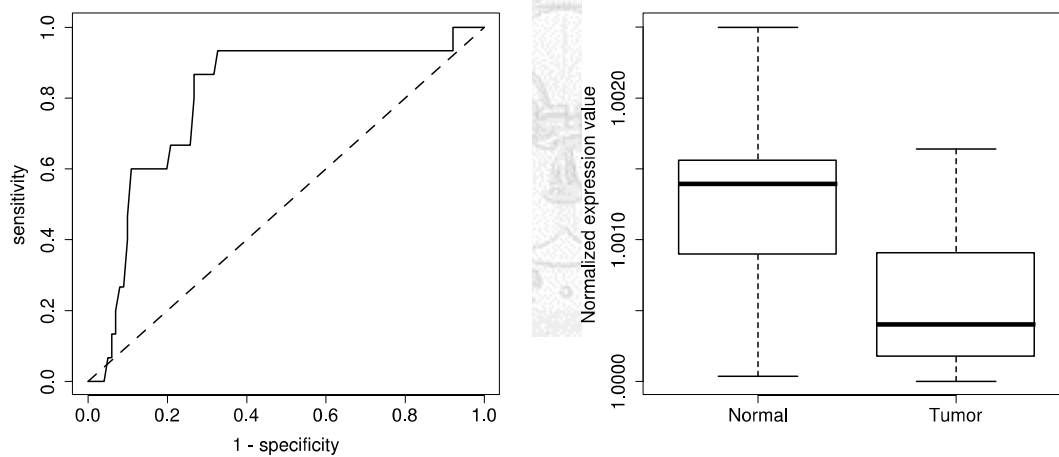
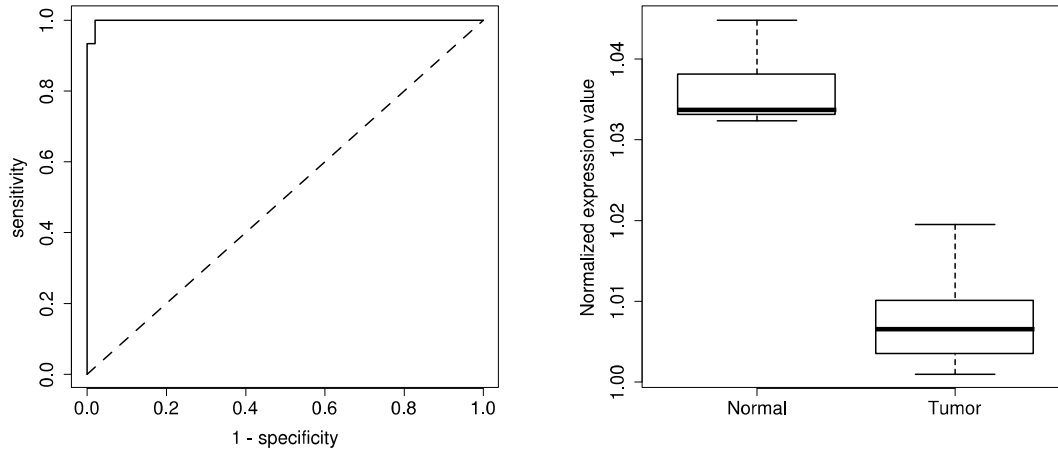


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line:** AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******: p-value < 0.001.

E. hsa-miR-99a-5p. MIMAT0000097. ***



F. hsa-miR-100-5p. MIMAT0000098. ***

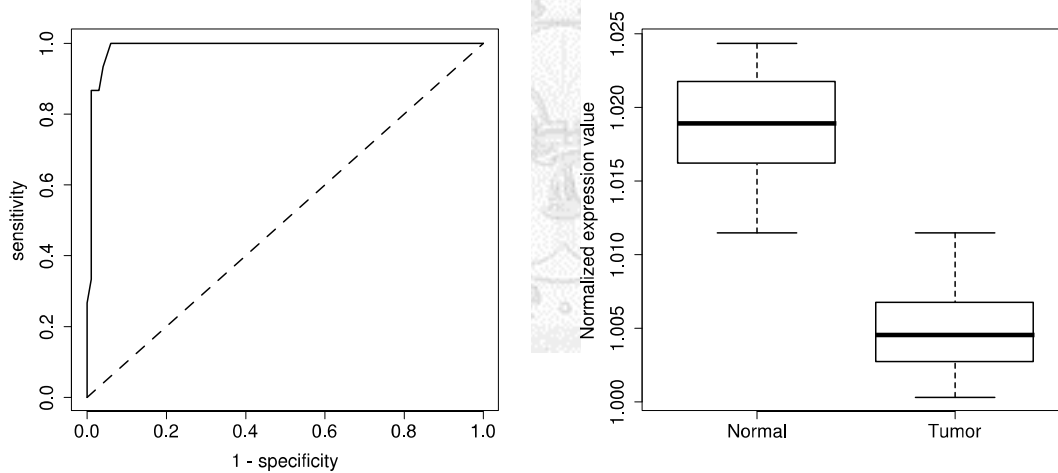
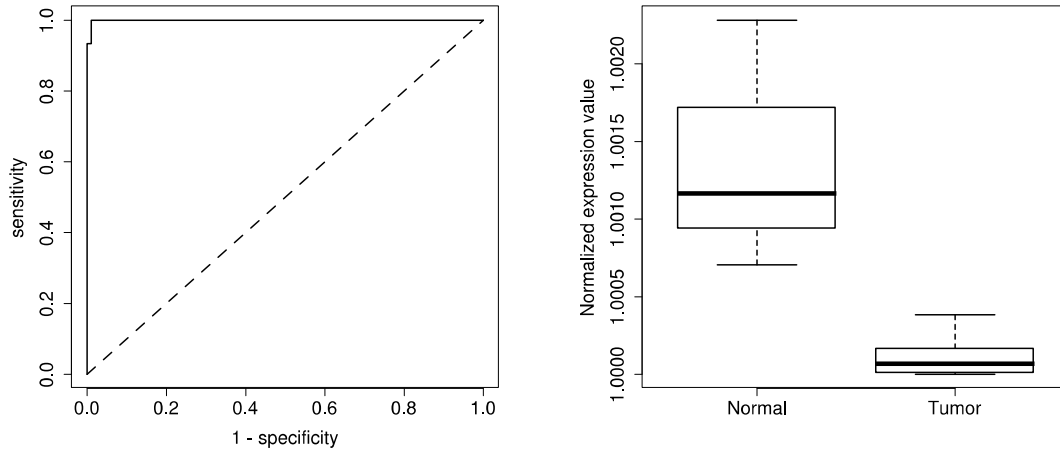


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line:** AUC = 0.5 (no predictive power). *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

G. hsa-miR-139-5p. MIMAT0000250. ***



H. hsa-miR-204-5p. MIMAT0000265. ***

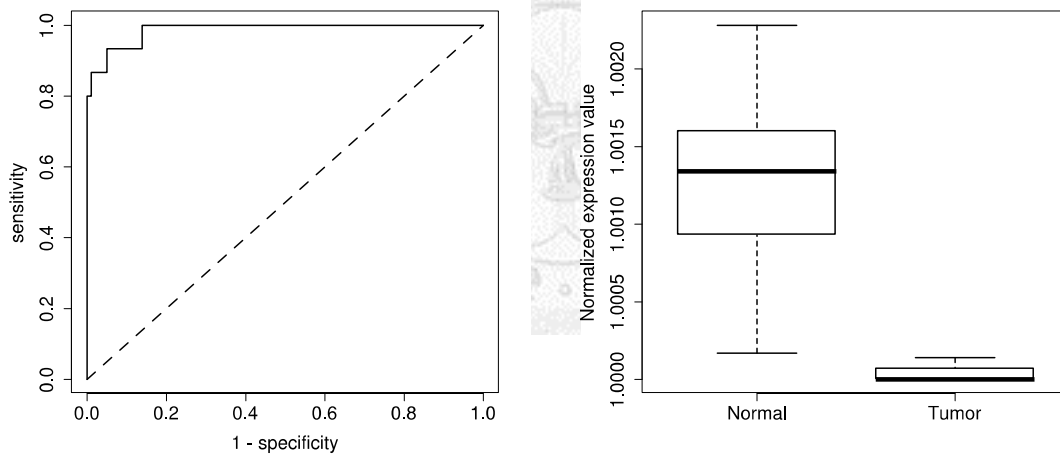
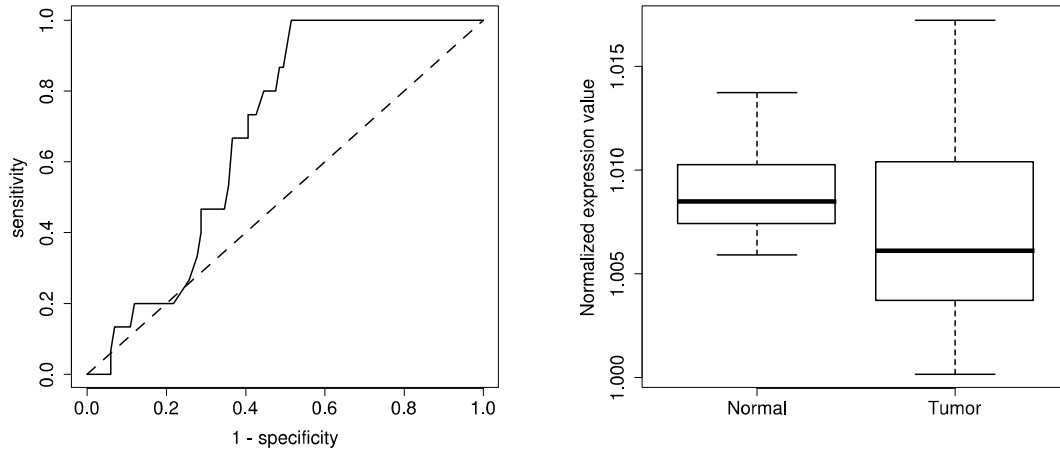


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line:** AUC = 0.5 (no predictive power). *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

I. hsa-miR-214-3p. MIMAT0000271. *



J. hsa-miR-215. MIMAT0000272.

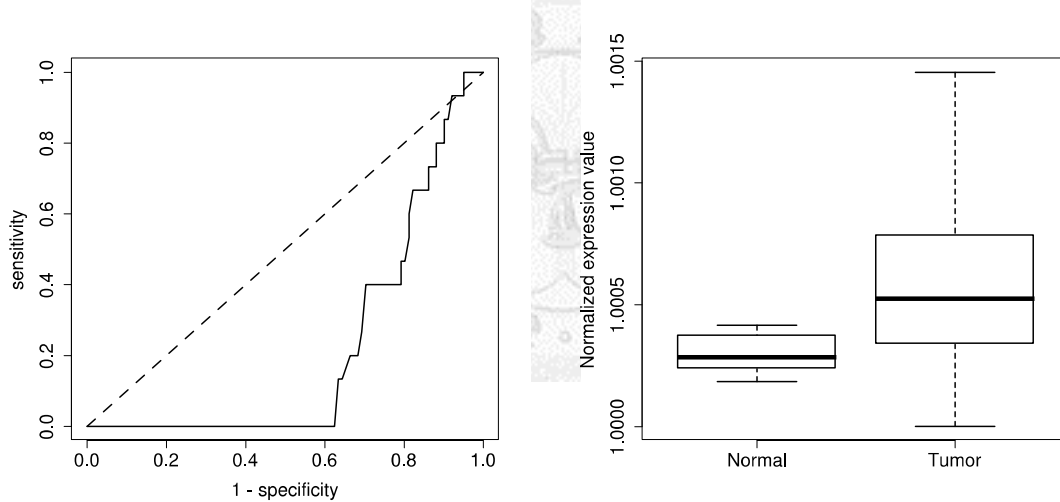
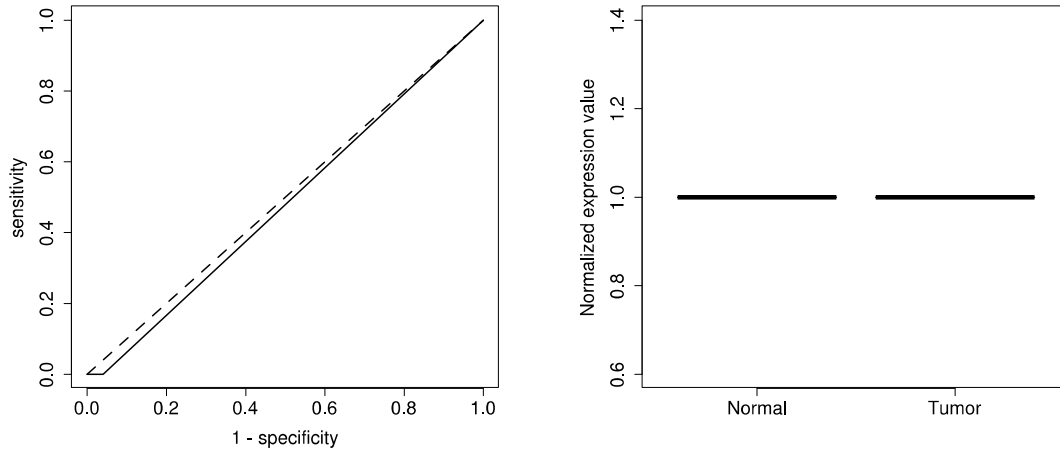


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line**: AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******: p-value < 0.001.

K. hsa-miR-122-5p. MIMAT0000421.



L. hsa-miR-125b-5p. MIMAT0000423. ***

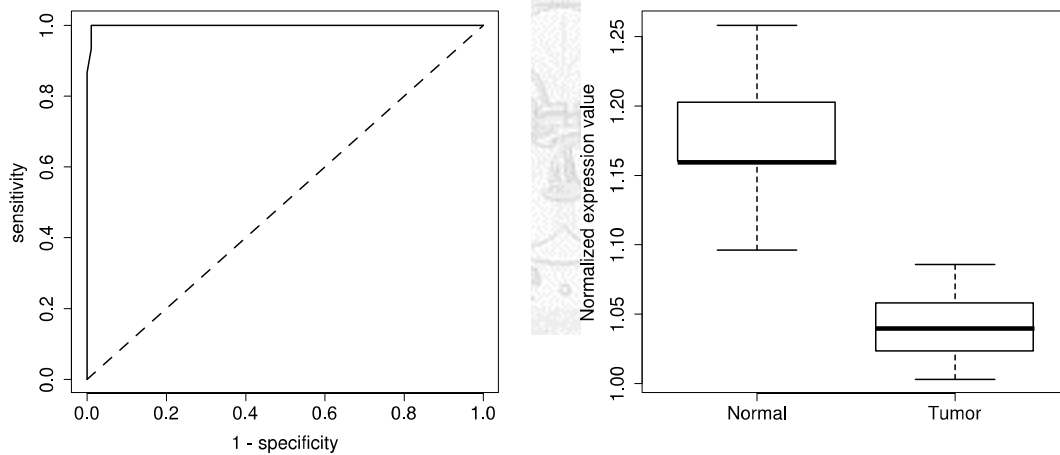
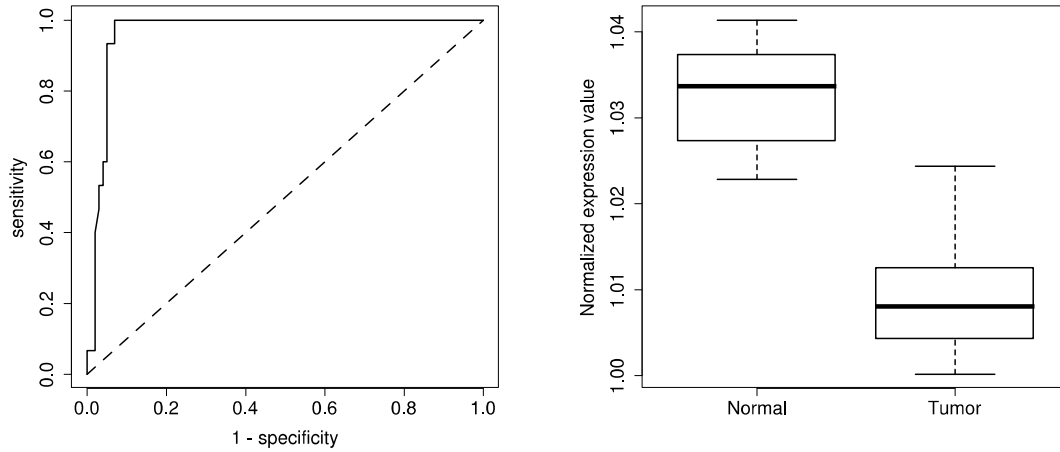


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line**: AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******: p-value < 0.001.

M. hsa-miR-145-5p. MIMAT0000437. ***



N. hsa-miR-125a-5p. MIMAT0000443.

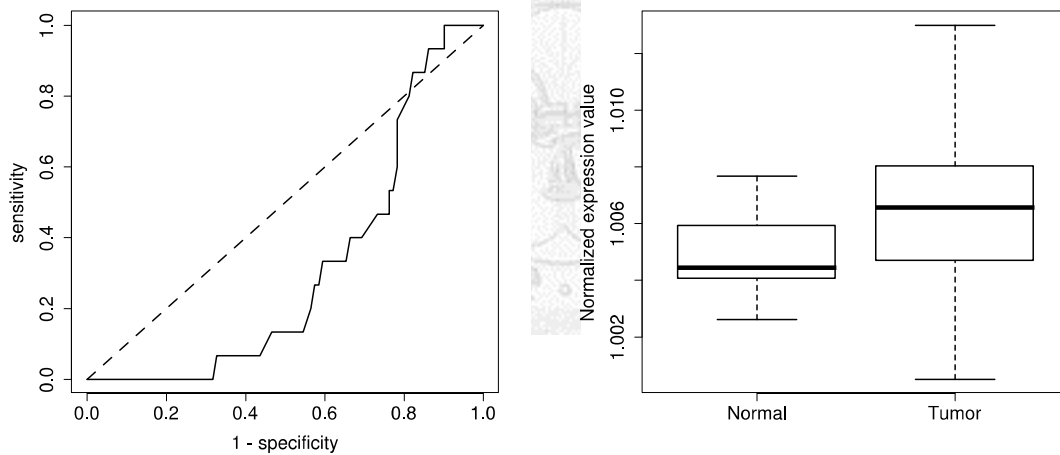
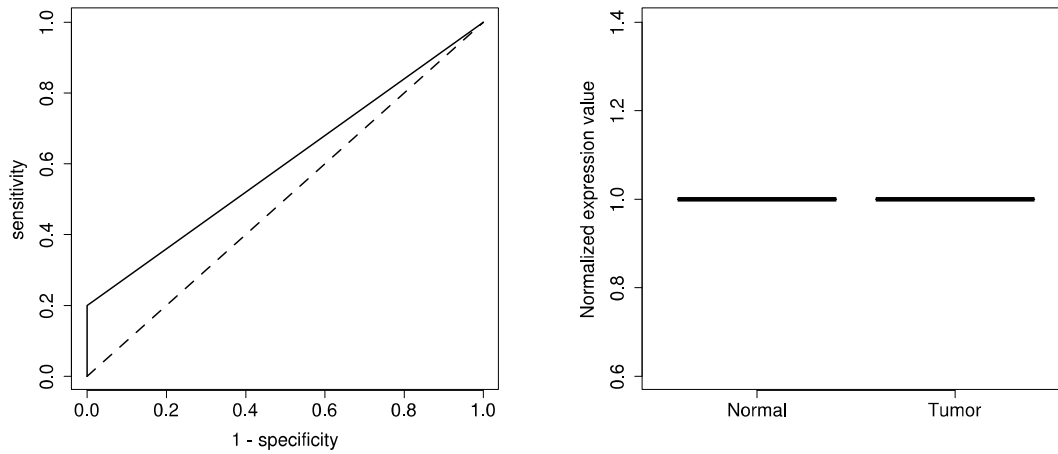


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line**: AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******: p-value < 0.001.

O. hsa-miR-383. MIMAT0000738.



P. hsa-miR-193b-3p. MIMAT0002819.

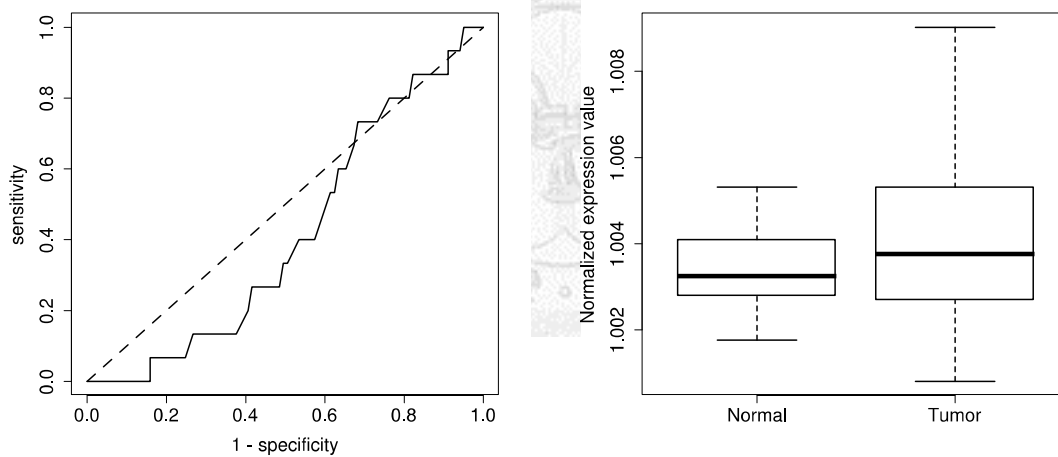


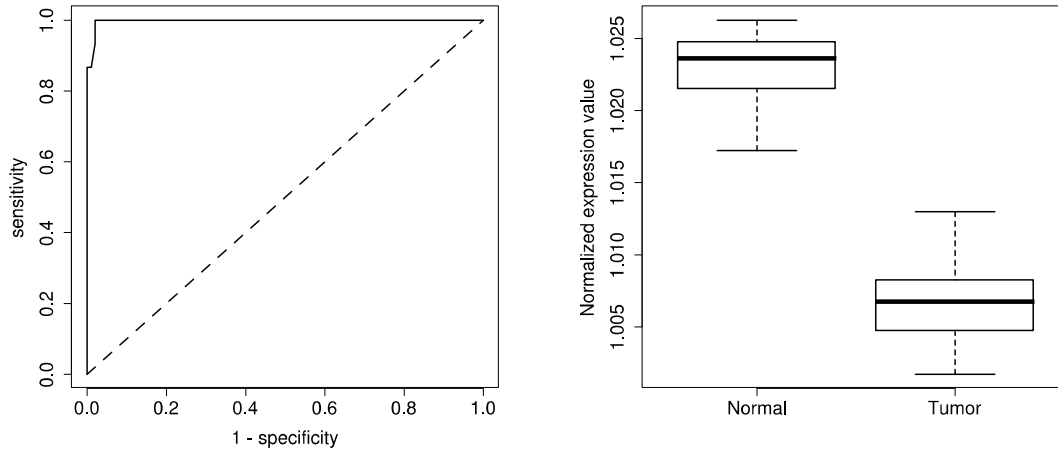
Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed**

line: AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******:

p-value < 0.001.

Q. hsa-miR-497-5p. MIMAT0002820. ***



R. hsa-miR-520d-3p. MIMAT0002856.

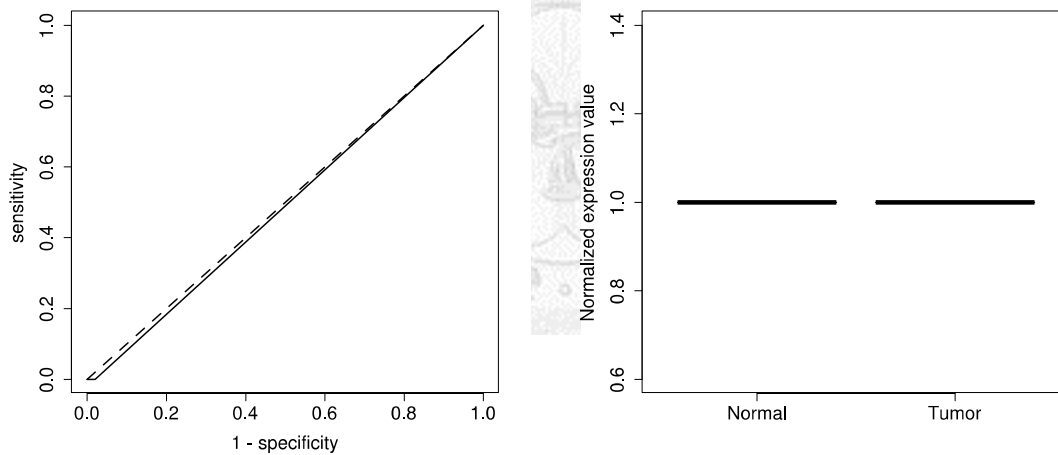
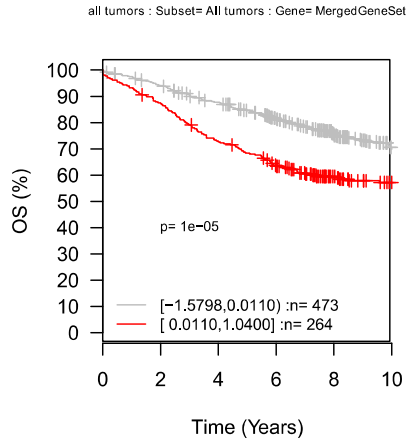


Figure 20. ROC curves and expression distribution of miRNAs. (continue)

Solid line: actual prediction performance of the miRNA as diagnostic marker; **dashed line**: AUC = 0.5 (no predictive power). *****: p-value < 0.05; ******: p-value < 0.01; *******: p-value < 0.001.

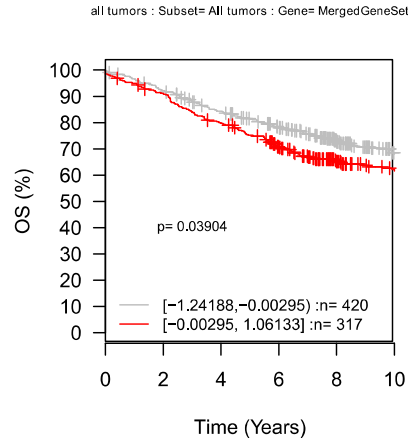
MIMAT0000064 (hsa-let-7c)

GO:0015630, Microtubule cytoskeleton



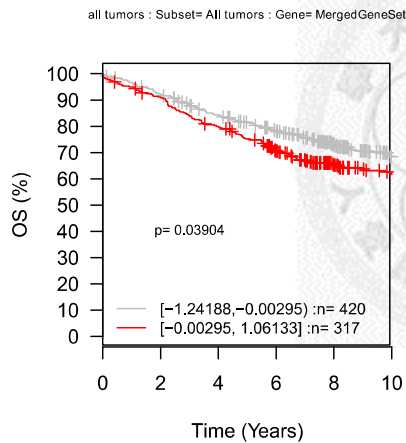
MIMAT0000064 (hsa-let-7c)

GO:0043069, Negative regulation of programmed cell death



MIMAT0000064 (hsa-let-7c)

GO:0060548, Negative regulation of cell death



MIMAT0000089 (hsa-miR-31-5p)

GO:0006917, Induction of apoptosis

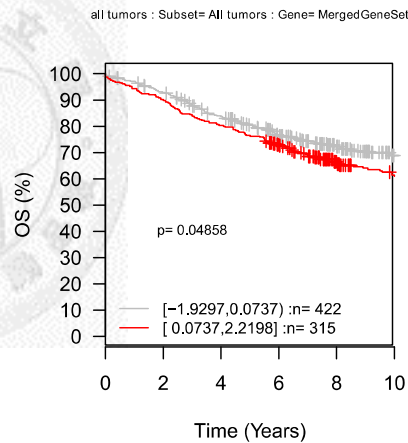
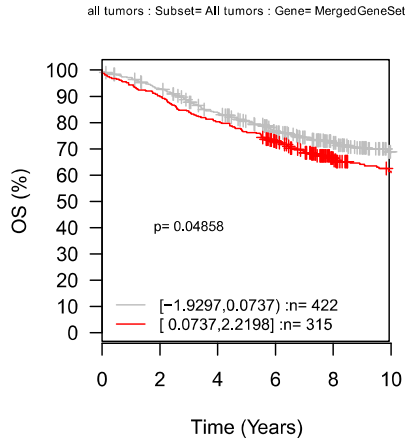


Figure 21 GOBO Kaplan-Meier survival curve of selected gene set with significant clinical outcome.

The gene set was formed by selecting genes annotated with designated GO term in the miRNA-regulated PIN. Full GOBO Kaplan-Meier survival p-value were calculated and listed in **Table 9**. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

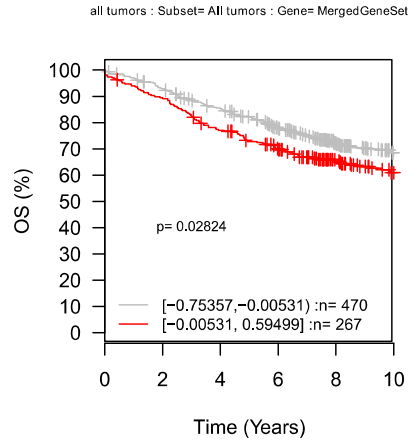
MIMAT0000089 (hsa-miR-31-5p)

GO:0012502, Induction of programmed cell death



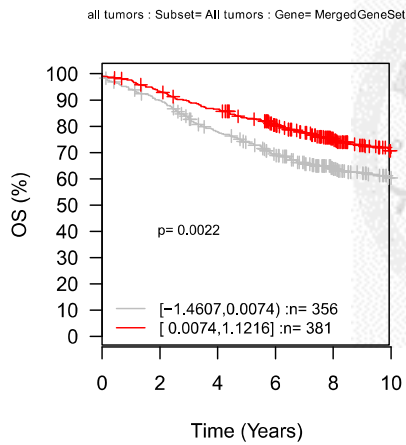
MIMAT0000423 (hsa-miR-125b-5p)

GO:0015630, Microtubule cytoskeleton



MIMAT0000423 (hsa-miR-125b-5p)

GO:0050678, Regulation of epithelial cell proliferation



MIMAT0000443 (hsa-miR-125a-5p)

GO:0015630, Microtubule cytoskeleton

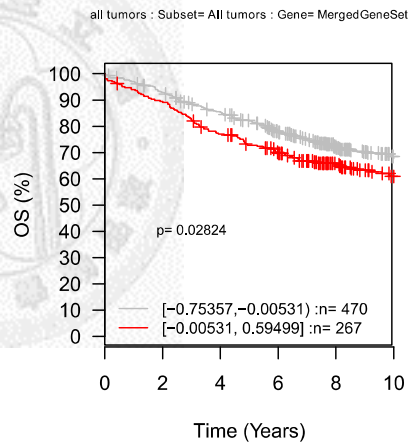
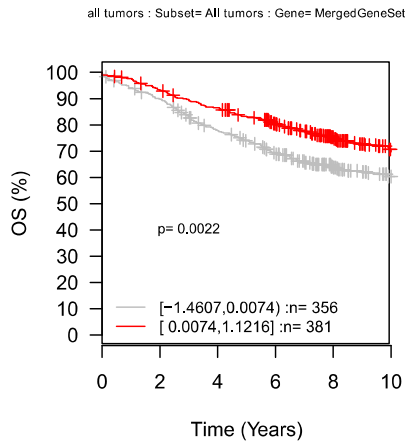


Figure 21 GOBO Kaplan-Meier survival curve of selected gene set (continue)

*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

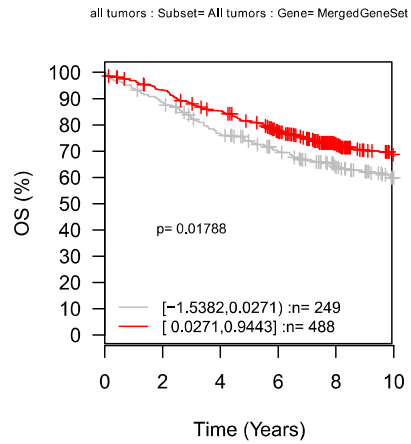
MIMAT0000443 (hsa-miR-125a-5p)

GO:0050678, Regulation of epithelial cell proliferation



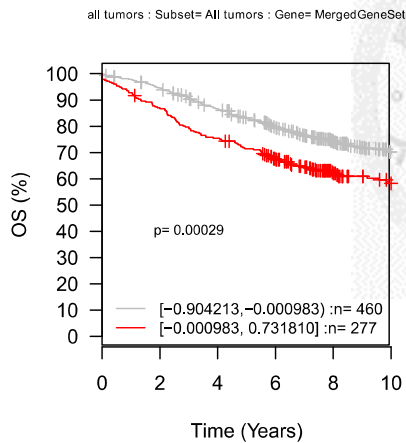
MIMAT0002820 (hsa-miR-497-5p)

GO:0007179, Transforming growth factor beta receptor signaling pathway



MIMAT0002856 (hsa-miR-520d-3p)

GO:0015630, Microtubule cytoskeleton



MIMAT0002856 (hsa-miR-520d-3p)

GO:0048011, Nerve growth factor receptor signaling pathway

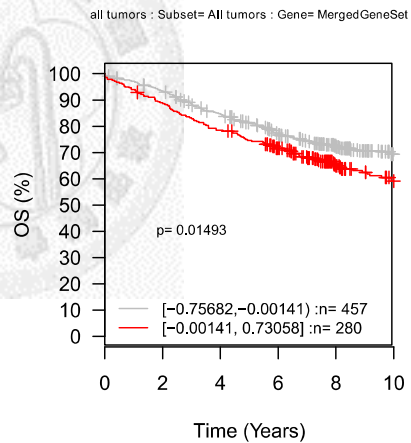


Figure 21 GOBO Kaplan-Meier survival curve of selected gene set (continue)

*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

MIMAT0002856 (hsa-miR-520d-3p)

GO:0051988, Regulation of attachment of spindle
microtubules to kinetochore

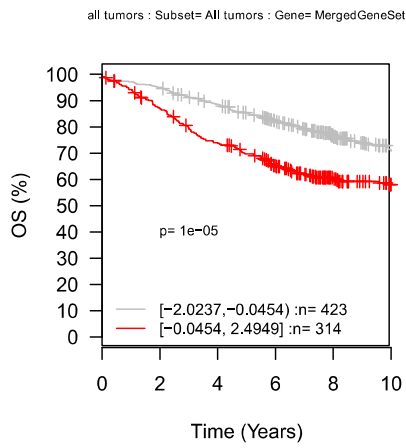
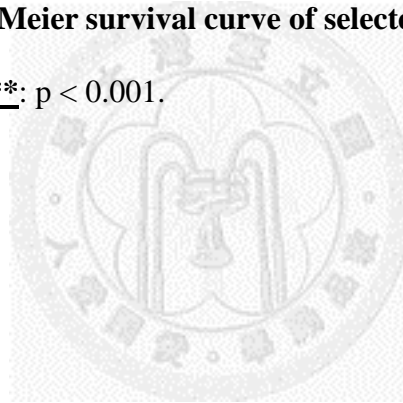


Figure 21 GOBO Kaplan-Meier survival curve of selected gene set (continue)

*****: $p < 0.05$; ******: $p < 0.01$; *******: $p < 0.001$.



TABLES

Table 1. Significantly differentially expressed miRNAs found in dataset GSE29173.

There are 89 down-regulated miRNAs and 1 up-regulated miRNA in this list. Q-value reported by SAM were 0 for all miRNAs in this list.

miRBase Accession	miRNA Name	Fold Change	miRBase Accession	miRNA Name	Fold Change
MIMAT0004761	hsa-miR-483-5p	0.01	MIMAT0000077	hsa-miR-22-3p	0.21
MIMAT0004552	hsa-miR-139-3p	0.01	MIMAT0000089	hsa-miR-31-5p	0.21
MIMAT0000738	hsa-miR-383	0.02	MIMAT0004612	hsa-miR-186-3p	0.21
MIMAT0002856	hsa-miR-520d-3p	0.02	MIMAT0004592	hsa-miR-125b-1-3p	0.22
MIMAT0002811	hsa-miR-202-3p	0.03	MIMAT0001639	hsa-miR-409-3p	0.22
MIMAT0002177	hsa-miR-486-5p	0.04	MIMAT0015032	hsa-miR-3158-3p	0.22
MIMAT0022721	hsa-miR-1247-3p	0.05	MIMAT0004496	hsa-miR-23a-5p	0.22
MIMAT0002175	hsa-miR-485-5p	0.06	MIMAT0000690	hsa-miR-296-5p	0.22
MIMAT0000265	hsa-miR-204-5p	0.07	MIMAT0000731	hsa-miR-378a-5p	0.23
MIMAT0000752	hsa-miR-328	0.07	MIMAT0000448	hsa-miR-136-5p	0.23
MIMAT0000421	hsa-miR-122-5p	0.07	MIMAT0004796	hsa-miR-576-3p	0.23
MIMAT0000447	hsa-miR-134	0.08	MIMAT0010133	hsa-miR-2110	0.23
MIMAT0000722	hsa-miR-370	0.09	MIMAT0004951	hsa-miR-887	0.23
MIMAT0004513	hsa-miR-101-5p	0.09	MIMAT0003239	hsa-miR-574-3p	0.25
MIMAT0000446	hsa-miR-127-3p	0.10	MIMAT0005901	hsa-miR-1249	0.25
MIMAT0000097	hsa-miR-99a-5p	0.10	MIMAT0000510	hsa-miR-320a	0.26
MIMAT0004566	hsa-miR-218-2-3p	0.10	MIMAT0002172	hsa-miR-376b	0.26
MIMAT0000729	hsa-miR-376a-3p	0.11	MIMAT0000250	hsa-miR-139-5p	0.27
MIMAT0009197	hsa-miR-205-3p	0.11	MIMAT0005825	hsa-miR-1180	0.27
MIMAT0004615	hsa-miR-195-3p	0.11	MIMAT0000437	hsa-miR-145-5p	0.28
MIMAT0005899	hsa-miR-1247-5p	0.11	MIMAT0004601	hsa-miR-145-3p	0.28
MIMAT0000720	hsa-miR-376c	0.12	MIMAT0003322	hsa-miR-652-3p	0.28
MIMAT0000762	hsa-miR-324-3p	0.12	MIMAT0000756	hsa-miR-326	0.28
MIMAT0004679	hsa-miR-296-3p	0.12	MIMAT0000098	hsa-miR-100-5p	0.29
MIMAT0004614	hsa-miR-193a-5p	0.12	MIMAT0003296	hsa-miR-627	0.29
MIMAT0003880	hsa-miR-671-5p	0.12	MIMAT0002820	hsa-miR-497-5p	0.31
MIMAT0004795	hsa-miR-574-5p	0.12	MIMAT0004507	hsa-miR-92a-1-5p	0.31
MIMAT0004599	hsa-miR-143-5p	0.13	MIMAT0000271	hsa-miR-214-3p	0.32
MIMAT0000423	hsa-miR-125b-5p	0.13	MIMAT0004702	hsa-miR-339-3p	0.33
MIMAT0004957	hsa-miR-760	0.13	MIMAT0004611	hsa-miR-185-3p	0.33
MIMAT0004911	hsa-miR-874	0.14	MIMAT0000064	hsa-let-7c	0.34
MIMAT0004603	hsa-miR-125b-2-3p	0.15	MIMAT0004673	hsa-miR-29c-5p	0.35
MIMAT0004952	hsa-miR-665	0.15	MIMAT0000733	hsa-miR-379-5p	0.35
MIMAT0018205	hsa-miR-3928	0.15	MIMAT0004594	hsa-miR-132-5p	0.35
MIMAT0004767	hsa-miR-193b-5p	0.15	MIMAT0000765	hsa-miR-335-5p	0.35
MIMAT0002861	hsa-miR-518e-3p	0.15	MIMAT0002819	hsa-miR-193b-3p	0.36
MIMAT0004604	hsa-miR-127-5p	0.16	MIMAT0000088	hsa-miR-30a-3p	0.36
MIMAT0002807	hsa-miR-491-5p	0.16	MIMAT0005951	hsa-miR-1307-3p	0.36
MIMAT0004689	hsa-miR-377-5p	0.16	MIMAT0004597	hsa-miR-140-3p	0.37
MIMAT0004762	hsa-miR-486-3p	0.16	MIMAT0004556	hsa-miR-10b-3p	0.37
MIMAT0000732	hsa-miR-378a-3p	0.17	MIMAT0000272	hsa-miR-215	0.37
MIMAT0017981	hsa-miR-3605-5p	0.18	MIMAT0004511	hsa-miR-99a-3p	0.37
MIMAT0004605	hsa-miR-129-2-3p	0.19	MIMAT0000443	hsa-miR-125a-5p	0.38
MIMAT0006789	hsa-miR-1468	0.20	MIMAT0004482	hsa-let-7b-3p	0.38
MIMAT0000737	hsa-miR-382-5p	0.21	MIMAT0000076	hsa-miR-21-5p	6.58

Table 2. Down-regulated genes found in dataset GSE29174.

There are 726 down-regulated genes in this list. Q-values reported by SAM were 0 for all genes in this list.

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
2949	GSTM5	0.06	619373	MBOAT4	0.17
10894	LYVE1	0.06	130399	ACVR1C	0.17
5950	RBP4	0.07	1646	AKR1C2	0.17
762	CA4	0.09	80763	C12orf39	0.17
54997	TESC	0.09	2159	F10	0.18
3489	IGFBP6	0.09	84889	SLC7A3	0.18
3952	LEP	0.09	1308	COL17A1	0.18
213	ALB	0.09	83699	SH3BGRL2	0.18
3131	HLF	0.10	84417	C2orf40	0.18
4023	LPL	0.10	4081	MAB21L1	0.18
10633	RASL10A	0.11	3484	IGFBP1	0.18
364	AQP7	0.11	5239	PGM5	0.19
1908	EDN3	0.11	4969	OGN	0.19
1811	SLC26A3	0.11	2719	GPC3	0.19
91851	CHRD1	0.11	116362	RBP7	0.19
729359	PLIN4	0.13	948	CD36	0.19
1149	CIDEA	0.13	5764	PTN	0.19
5959	RDH5	0.13	3043	HBB	0.19
5348	FXYD1	0.14	56920	SEMA3G	0.20
5346	PLIN1	0.14	94274	PPP1R14A	0.20
10249	GLYAT	0.14	57447	NDRG2	0.20
158800	RHOXF1	0.14	84795	PYROXD2	0.20
221476	PII6	0.14	84649	DGAT2	0.20
3040	HBA2	0.14	2690	GHR	0.20
6939	TCF15	0.14	22802	CLCA4	0.20
79645	EFCAB1	0.14	5179	PENK	0.20
80343	SEL1L2	0.14	6663	SOX10	0.20
9413	FAM189A2	0.15	6649	SOD3	0.21
26289	AK5	0.15	54922	RASIP1	0.21
25891	PAMR1	0.15	8406	SRPX	0.21
3679	ITGA7	0.15	1446	CSN1S1	0.21
1264	CNN1	0.15	7123	CLEC3B	0.22
92304	SCGB3A1	0.15	9647	PPM1F	0.22
2167	FABP4	0.15	1842	ECM2	0.22
23285	KIAA1107	0.15	3909	LAMA3	0.22
7145	TNS1	0.16	8639	AOC3	0.23
4881	NPR1	0.16	2934	GSN	0.23
1028	CDKN1C	0.16	9370	ADIPOQ	0.23
1036	CDO1	0.16	3202	HOXA5	0.23
130271	PLEKHH2	0.16	9452	ITM2A	0.23
8736	MYOM1	0.16	6290	SAA3P	0.23
8908	GYG2	0.16	4604	MYBPC1	0.23
79785	RERGL	0.16	1128	CHRM1	0.23
221091	LRRN4CL	0.17	83878	USHBP1	0.24
3991	LIPE	0.17	63970	TP53AIP1	0.24

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
27175	TUBG2	0.24	79192	IRX1	0.28
1346	COX7A1	0.24	3400	ID4	0.28
6376	CX3CL1	0.24	57519	STARD9	0.29
50486	G0S2	0.24	57666	FBRSL1	0.29
6285	S100B	0.24	3590	IL11RA	0.29
443	ASPA	0.24	57664	PLEKHA4	0.29
947	CD34	0.25	197257	LDHD	0.29
84632	AFAP1L2	0.25	66036	MTMR9	0.29
3866	KRT15	0.25	2321	FLT1	0.29
147463	ANKRD29	0.25	126	ADH1C	0.29
2878	GPX3	0.25	1363	CPE	0.29
7079	TIMP4	0.25	56131	PCDHB4	0.29
54345	SOX18	0.25	22915	MMRN1	0.29
51277	DNAJC27	0.25	7069	THRSP	0.29
84870	RSPO3	0.25	57161	PEL12	0.30
55323	LARP6	0.25	770	CA11	0.30
6387	CXCL12	0.25	53342	IL17D	0.30
137835	TMEM71	0.25	79987	SVEP1	0.30
5212	VIT	0.25	857	CAV1	0.30
26577	PCOLCE2	0.25	222166	C7orf41	0.30
845	CASQ2	0.25	27190	IL17B	0.30
6422	SFRP1	0.25	116159	CYYR1	0.30
10351	ABCA8	0.26	4487	MSX1	0.30
10840	ALDH1L1	0.26	9068	ANGPTL1	0.30
65983	GRAMD3	0.26	10411	RAPGEF3	0.30
84327	ZBED3	0.26	3199	HOXA2	0.30
57124	CD248	0.26	2944	GSTM1	0.30
3235	HOXD9	0.26	2920	CXCL2	0.30
2192	FBLN1	0.26	201134	CEP112	0.31
91653	BOC	0.26	220001	VWCE	0.31
4147	MATN2	0.26	83888	FGFBP2	0.31
126669	SHE	0.27	6366	CCL21	0.31
2788	GNG7	0.27	6711	SPTBN1	0.31
129804	FBLN7	0.27	85378	TUBGCP6	0.31
270	AMPD1	0.27	26040	SETBP1	0.31
79656	BEND5	0.27	4692	NDN	0.31
58503	PROL1	0.27	25890	ABI3BP	0.31
3316	HSPB2	0.27	23531	MMD	0.31
729440	CCDC61	0.27	30846	EHD2	0.31
54438	GFOD1	0.27	6196	RPS6KA2	0.31
5243	ABCB1	0.27	2009	EML1	0.31
810	CALML3	0.27	6289	SAA2	0.31
6898	TAT	0.27	345275	HSD17B13	0.31
5648	MASP1	0.28	2701	GJA4	0.32
25999	CLIP3	0.28	112609	MRAP2	0.32
125875	CLDND2	0.28	727	C5	0.32
7102	TSPAN7	0.28	477	ATP1A2	0.32
1879	EBF1	0.28	9627	SNCAIP	0.32
23252	OTUD3	0.28	4435	CITED1	0.32
5493	PPL	0.28	10974	C10orf116	0.32
83987	CCDC8	0.28	11005	SPINK5	0.32
9073	CLDN8	0.28	80325	ABTB1	0.33
221981	THSD7A	0.28	221395	GPR116	0.33
64102	TNMD	0.28	10014	HDAC5	0.33

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
137872	ADHFE1	0.33	1489	CTF1	0.37
27151	CPAMD8	0.33	35	ACADS	0.37
387923	SERP2	0.33	3749	KCNC4	0.37
145581	LRFN5	0.33	140738	TMEM37	0.37
6263	RYR3	0.33	2791	GNG11	0.37
2354	FOSB	0.33	23604	DAPK2	0.37
51302	CYP39A1	0.33	10217	CTDSPL	0.37
4128	MAOA	0.34	23550	PSD4	0.37
117248	GALNTL2	0.34	4306	NR3C2	0.37
10268	RAMP3	0.34	119587	CPXM2	0.37
7730	ZNF177	0.34	7942	TFEB	0.37
10873	ME3	0.34	3815	KIT	0.37
7461	CLIP2	0.34	1805	DPT	0.37
7049	TGFBR3	0.34	23242	COBL	0.37
79901	CYBRD1	0.34	4313	MMP2	0.37
5152	PDE9A	0.34	4139	MARK1	0.37
50805	IRX4	0.34	9104	RGN	0.37
8644	AKR1C3	0.34	2329	FMO4	0.37
5915	RARB	0.34	25802	LMOD1	0.38
2770	GNAII	0.34	4239	MFAP4	0.38
54996	2-Mar	0.35	10392	NOD1	0.38
79791	FBXO31	0.35	6794	STK11	0.38
54776	PPP1R12C	0.35	85458	DIXDC1	0.38
9079	LDB2	0.35	4123	MAN2C1	0.38
57104	PNPLA2	0.35	54476	RNF216	0.38
30008	EFEMP2	0.35	9920	KBTD11	0.38
91461	PKDCC	0.35	6329	SCN4A	0.38
23368	PPP1R13B	0.35	10253	SPRY2	0.38
23461	ABCA5	0.35	1910	EDNRB	0.38
9572	NR1D1	0.35	9249	DHRS3	0.38
23338	PHF15	0.35	22869	ZNF510	0.38
114800	CCDC85A	0.35	3384	ICAM2	0.38
2550	GABBR1	0.35	8613	PPAP2B	0.38
4638	MYLK	0.35	1950	EGF	0.38
2327	FMO2	0.35	55273	TMEM100	0.38
139411	PTCHD1	0.35	6297	SALL2	0.38
10391	CORO2B	0.35	9365	KL	0.38
25854	FAM149A	0.35	8863	PER3	0.38
55701	ARHGEF40	0.36	8404	SPARCL1	0.38
1759	DNM1	0.36	2202	EFEMP1	0.38
22849	CPEB3	0.36	8369	HIST1H4G	0.38
57716	PRX	0.36	5187	PER1	0.39
1628	DBP	0.36	30815	ST6GALNAC6	0.39
80031	SEMA6D	0.36	256364	EML3	0.39
259217	HSPA12A	0.36	57381	RHOJ	0.39
6909	TBX2	0.36	761	CA3	0.39
1511	CTSG	0.36	83989	FAM172A	0.39
79971	WLS	0.36	1408	CRY2	0.39
90865	IL33	0.36	2281	FKBP1B	0.39
11343	MGLL	0.36	51222	ZNF219	0.39
55800	SCN3B	0.36	54540	FAM193B	0.39
1949	EFNB3	0.36	4053	LTBP2	0.39
284217	LAMA1	0.36	55184	DZANK1	0.39
22927	HABP4	0.37	5740	PTGIS	0.39

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
23645	PPP1R15A	0.39	84814	PPAPDC3	0.42
342574	KRT27	0.39	79365	BHLHE41	0.42
83543	AIF1L	0.39	316	AOX1	0.42
624	BDKRB2	0.39	23380	SRGAP2	0.42
347	APOD	0.39	84033	OBSCN	0.42
84935	C13orf33	0.39	90353	CTU1	0.42
858	CAV2	0.39	9013	TAF1C	0.42
5138	PDE2A	0.40	474344	GIMAP6	0.42
114928	GPRASP2	0.40	84883	AIFM2	0.42
58190	CTDSP1	0.40	58480	RHOA	0.42
513	ATP5D	0.40	65982	ZSCAN18	0.42
57684	ZBTB26	0.40	666	BOK	0.42
7041	TGFB1I1	0.40	79762	C1orf115	0.42
5787	PTPRB	0.40	525	ATP6V1B1	0.42
7294	TXK	0.40	4675	NAP1L3	0.42
56301	SLC7A10	0.40	3257	HPS1	0.43
55937	APOM	0.40	55781	RIOK2	0.43
6368	CCL23	0.40	63947	DMRTC1	0.43
55020	TTC38	0.40	1969	EPHA2	0.43
134265	AFAP1L1	0.40	25927	CNRIP1	0.43
4485	MST1	0.40	57685	CACHD1	0.43
51559	NT5DC3	0.40	29997	GLTSCR2	0.43
7169	TPM2	0.40	26051	PPP1R16B	0.43
51705	EMCN	0.40	83604	TMEM47	0.43
8938	BALAP3	0.40	2308	FOXO1	0.43
10365	KLF2	0.40	55225	RAVER2	0.43
59	ACTA2	0.40	54839	LRRC49	0.43
80309	SPHKAP	0.40	122953	JDP2	0.43
3779	KCNMB1	0.41	29775	CARD10	0.43
10826	C5orf4	0.41	166	AES	0.43
219654	ZCCHC24	0.41	25924	MYRIP	0.43
92162	TMEM88	0.41	2852	GPER	0.43
7450	VWF	0.41	51421	AMOTL2	0.43
10266	RAMP2	0.41	124936	CYB5D2	0.43
25875	LETMD1	0.41	1294	COL7A1	0.43
1938	EEF2	0.41	127435	PODN	0.43
121551	BTBD11	0.41	84952	CGNL1	0.43
2119	ETV5	0.41	83483	PLVAP	0.43
9696	CROCC	0.41	1958	EGR1	0.43
1031	CDKN2C	0.41	230	ALDOC	0.43
9037	SEMA5A	0.41	65987	KCTD14	0.43
3397	ID1	0.41	4804	NGFR	0.44
84707	BEX2	0.41	64852	TUT1	0.44
57616	TSHZ3	0.41	84253	GARNL3	0.44
1471	CST3	0.41	5866	RAB3IL1	0.44
55214	LEPREL1	0.41	10608	MXD4	0.44
3914	LAMB3	0.41	4211	MEIS1	0.44
57478	USP31	0.41	83547	RILP	0.44
3783	KCNN4	0.41	9172	MYOM2	0.44
8839	WISP2	0.41	57192	MCOLN1	0.44
1583	CYP11A1	0.42	255877	BCL6B	0.44
10124	ARL4A	0.42	56904	SH3GLB2	0.44
738	C11orf2	0.42	51285	RASL12	0.44
29800	ZDHHC1	0.42	3425	IDUA	0.44

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
23135	KDM6B	0.44	402117	VWC2L	0.46
171024	SYNPO2	0.44	81490	PTDSS2	0.46
10350	ABCA9	0.44	283748	PLA2G4D	0.46
3691	ITGB4	0.44	23523	CABIN1	0.46
2348	FOLR1	0.44	6146	RPL22	0.46
11145	PLA2G16	0.44	85360	SYDE1	0.46
554	AVPR2	0.45	60468	BACH2	0.46
64072	CDH23	0.45	57451	ODZ2	0.46
80177	MYCT1	0.45	4013	VWA5A	0.46
5957	RCVRN	0.45	339768	ESPNL	0.46
408	ARRB1	0.45	3860	KRT13	0.46
144699	FBXL14	0.45	7094	TLN1	0.46
83719	YPEL3	0.45	4232	MEST	0.46
22841	RAB11FIP2	0.45	1410	CRYAB	0.46
283927	NUDT7	0.45	57452	GALNTL1	0.47
293	SLC25A6	0.45	63935	PCIF1	0.47
90507	SCRN2	0.45	25873	RPL36	0.47
37	ACADVL	0.45	9812	KIAA0141	0.47
112744	IL17F	0.45	51665	ASB1	0.47
6709	SPTAN1	0.45	64123	ELTD1	0.47
8086	AAAS	0.45	6122	RPL3	0.47
7423	VEGFB	0.45	222962	SLC29A4	0.47
64221	ROBO3	0.45	23102	TBC1D2B	0.47
7273	TTN	0.45	3476	IGBP1	0.47
2657	GDF1	0.45	93408	MYL10	0.47
59271	C21orf63	0.45	5310	PKD1	0.47
132160	PPM1M	0.45	4628	MYH10	0.47
27244	SESN1	0.45	221935	SDK1	0.47
51310	SLC22A17	0.45	23328	SASH1	0.47
4828	NMB	0.45	8522	GAS7	0.47
54360	CYTL1	0.45	10023	FRAT1	0.47
203245	NAIF1	0.45	7301	TYRO3	0.47
23166	STAB1	0.45	2767	GNA11	0.47
2121	EVC	0.45	9457	FHL5	0.47
116496	FAM129A	0.45	4094	MAF	0.47
23239	PHLPP1	0.45	65268	WNK2	0.47
51673	TPPP3	0.45	54585	LZTFL1	0.47
64094	SMOC2	0.45	375449	MAST4	0.47
6383	SDC2	0.45	138311	FAM69B	0.47
2180	ACSL1	0.45	160622	GRASP	0.47
23770	FKBP8	0.45	22837	COBLL1	0.47
55901	THSD1	0.46	51435	SCARA3	0.47
25895	METTL21B	0.46	217	ALDH2	0.47
23731	C9orf5	0.46	6236	RRAD	0.47
126393	HSPB6	0.46	8322	FZD4	0.47
4056	LTC4S	0.46	653275	CFC1B	0.47
79825	CCDC48	0.46	10908	PNPLA6	0.47
10810	WASF3	0.46	57526	PCDH19	0.47
29911	HOOK2	0.46	8424	BBOX1	0.47
583	BBS2	0.46	9905	SGSM2	0.48
28984	C13orf15	0.46	10435	CDC42EP2	0.48
1465	CSRP1	0.46	23087	TRIM35	0.48
55258	THNSL2	0.46	60314	C12orf10	0.48
161198	CLEC14A	0.46	1073	CFL2	0.48

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
3699	ITIH3	0.48	5256	PHKA2	0.49
92922	CCDC102A	0.48	6237	RRAS	0.49
65057	ACD	0.48	5288	PIK3C2G	0.49
9095	TBX19	0.48	10252	SPRY1	0.49
6441	SFTPD	0.48	79026	AHNAK	0.49
22846	VASH1	0.48	9693	RAPGEF2	0.49
51066	C3orf32	0.48	51226	COPZ2	0.49
23179	RGL1	0.48	158326	FREM1	0.49
4664	NAB1	0.48	1956	EGFR	0.49
50511	SYCP3	0.48	5360	PLTP	0.49
6430	SRSF5	0.48	290	ANPEP	0.49
11078	TRIOBP	0.48	1756	DMD	0.49
78991	PCYOX1L	0.48	5118	PCOLCE	0.49
6623	SNCG	0.48	56654	NPDC1	0.49
23384	SPECC1L	0.48	9254	CACNA2D2	0.49
53826	FXYD6	0.48	55536	CDCA7L	0.49
9397	NMT2	0.48	124975	GGT6	0.49
6041	RNASEL	0.48	1906	EDN1	0.49
113510	HELQ	0.48	81029	WNT5B	0.49
64788	LMF1	0.48	2646	GCKR	0.49
2217	FCGRT	0.48	9811	CTIF	0.50
79720	VPS37B	0.48	145376	PPP1R36	0.50
6764	ST5	0.48	222865	TMEM130	0.50
252969	NEIL2	0.48	92999	ZBTB47	0.50
8987	STBD1	0.48	168002	DACT2	0.50
41	ACCN2	0.48	6829	SUPT5H	0.50
7905	REEP5	0.48	9992	KCNE2	0.50
5919	RARRES2	0.48	58509	C19orf29	0.50
10544	PROCR	0.48	79706	PRKRIP1	0.50
6876	TAGLN	0.48	1153	CIRBP	0.50
8436	SDPR	0.49	9639	ARHGEF10	0.50
23500	DAAM2	0.49	4054	LTBP3	0.50
130132	RFTN2	0.49	1120	CHKB	0.50
80310	PDGFD	0.49	286046	XKR6	0.50
4215	MAP3K3	0.49	9590	AKAP12	0.50
282775	OR5J2	0.49	64115	C10orf54	0.50
51161	C3orf18	0.49	2067	ERCC1	0.50
29098	RANGRF	0.49	7507	XPA	0.50
53336	CPXCR1	0.49	22897	CEP164	0.50
9081	PRY	0.49	652	BMP4	0.50
9459	ARHGEF6	0.49	55702	CCDC94	0.50
2995	GYPC	0.49	57613	KIAA1467	0.50
23057	NMNAT2	0.49	28514	DLL1	0.50
4669	NAGLU	0.49	169270	ZNF596	0.50
6452	SH3BP2	0.49	83982	IFI27L2	0.50
51458	RHCG	0.49	2247	FGF2	0.50
1112	FOXP3	0.49	26248	OR2K2	0.50
29954	POMT2	0.49	84303	CHCHD6	0.50
9612	NCOR2	0.49	3615	IMPDH2	0.50
3198	HOXA1	0.49	1813	DRD2	0.50
5311	PKD2	0.49	80148	PQLC1	0.50
2946	GSTM2	0.49	390081	OR52E4	0.50
2109	ETFB	0.49	352954	GATS	0.50
56062	KLHL4	0.49	90871	C9orf123	0.50

Table 2. Down-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
6915	TBXA2R	0.50	50945	TBX22	0.52
64288	ZNF323	0.50	5204	PFDN5	0.52
5195	PEX14	0.50	5338	PLD2	0.52
84557	MAP1LC3A	0.50	94	ACVRL1	0.52
6164	RPL34	0.50	54039	PCBP3	0.52
8835	SOCS2	0.50	7691	ZNF132	0.52
2735	GLI1	0.50	338	APOB	0.52
26022	TMEM98	0.50	84658	EMR3	0.52
3908	LAMA2	0.50	283232	TMEM80	0.52
1825	DSC3	0.50	5430	POLR2A	0.52
5730	PTGDS	0.50	54623	PAF1	0.52
162515	SLC16A11	0.51	11070	TMEM115	0.52
274	BIN1	0.51	10395	DLC1	0.52
79654	HECTD3	0.51	57140	RNPEPL1	0.52
22863	ATG14	0.51	79781	IQCA1	0.52
25949	SYF2	0.51	1838	DTNB	0.52
84872	ZC3H10	0.51	51386	EIF3L	0.52
23187	PHLDB1	0.51	56919	DHX33	0.52
5434	POLR2E	0.51	57542	KLHDC5	0.52
6181	RPLP2	0.51	3628	INPP1	0.52
6141	RPL18	0.51	4520	MTF1	0.52
84747	UNC119B	0.51	8547	FCN3	0.52
23399	CTDNEP1	0.51	60401	EDA2R	0.52
599	BCL2L2	0.51	8082	SSPN	0.52
197258	FUK	0.51	80755	AARSD1	0.52
5207	PFKFB1	0.51	710	SERPING1	0.52
8131	NPRL3	0.51	56246	MRAP	0.52
25839	COG4	0.51	10555	AGPAT2	0.52
10816	SPINT3	0.51	949	SCARB1	0.52
60485	SAV1	0.51	23743	BHMT2	0.52
5681	PSKH1	0.51	3910	LAMA4	0.52
80318	GKAP1	0.51	60370	AVPI1	0.52
57088	PLSCR4	0.51	5021	OXTR	0.52
93129	ORAI3	0.51	55997	CFC1	0.52
5829	PXN	0.51	23144	ZC3H3	0.52
56776	FMN2	0.51	150709	ANKAR	0.52
85456	TNKS1BP1	0.51	6591	SNAI2	0.52
283	ANG	0.51	10129	FRY	0.52
7035	TFPI	0.51	5166	PDK4	0.52
51232	CRIM1	0.51	146433	IL34	0.52
112616	CMTM7	0.51	118812	MORN4	0.53
22981	NINL	0.51	10516	FBLN5	0.53
8727	CTNNA1	0.51	9463	PICK1	0.53
9902	MRC2	0.51	127495	LRRC39	0.53
10900	RUNDC3A	0.51	7753	ZNF202	0.53
51299	NRN1	0.51	79827	CLMP	0.53
79632	FAM184A	0.52	203260	CCDC107	0.53
80820	EEPD1	0.52	83657	DYNLRB2	0.53

Table 3. Up-regulated genes found in dataset GSE29174.

There are 437 up-regulated genes in this list. Q-value reported by SAM were 0 for all genes in this list.

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
1300	COL10A1	42.74	54443	ANLN	5.79
3007	HIST1H1D	29.72	6710	SPTB	5.71
8366	HIST1H4B	25.58	7272	TTK	5.64
6286	S100P	25.19	10635	RAD51AP1	5.49
1301	COL11A1	24.72	4069	LYZ	5.37
3627	CXCL10	17.83	55183	RIF1	5.34
4283	CXCL9	15.88	891	CCNB1	5.34
1387	CREBBP	12.83	91543	RSAD2	5.31
27299	ADAMDEC1	12.78	81610	FAM83D	5.24
54986	ULK4	12.46	64581	CLEC7A	5.10
55771	PRR11	12.02	10051	SMC4	5.02
54790	TET2	11.25	4085	MAD2L1	4.96
6241	RRM2	10.60	55872	PBK	4.83
3433	IFIT2	10.49	991	CDC20	4.82
6999	TDO2	9.73	9221	NOLC1	4.74
1656	DDX6	9.72	2124	EVI2B	4.66
55088	C10orf118	9.37	375248	ANKRD36	4.66
9648	GCC2	9.24	1164	CKS2	4.64
6696	SPP1	8.92	1230	CCR1	4.62
2803	GOLGA4	8.57	890	CCNA2	4.56
83540	NUF2	7.73	127933	UHMK1	4.49
10112	KIF20A	7.66	10274	STAG1	4.45
9833	MELK	7.59	597	BCL2A1	4.43
55165	CEP55	7.50	55355	HJURP	4.41
10142	AKAP9	7.44	54210	TREM1	4.36
9447	AIM2	7.42	253558	LCLAT1	4.26
2706	GJB2	7.33	1033	CDKN3	4.24
6498	SKIL	7.13	79801	SHCBP1	4.23
219285	SAMD9L	7.06	126731	C1orf96	4.21
10261	IGSF6	7.01	6772	STAT1	4.20
2335	FN1	6.95	55729	ATF7IP	4.14
699	BUB1	6.75	6713	SQLE	4.14
1058	CENPA	6.75	157570	ESCO2	4.10
332	BIRC5	6.73	79871	RPAP2	4.09
51203	NUSAP1	6.59	9493	KIF23	4.09
259266	ASPM	6.54	4751	NEK2	4.05
1063	CENPF	6.49	10631	POSTN	4.03
165918	RNF168	6.44	23515	MORC3	4.02
9232	PTTG1	6.34	7153	TOP2A	4.02
5996	RGS1	6.07	10403	NDC80	4.00
29089	UBE2T	5.96	10915	TCERG1	3.99
22974	TPX2	5.94	57650	KIAA1524	3.99
4321	MMP12	5.91	23049	SMG1	3.93
983	CDK1	5.89	80231	CXorf21	3.87
85444	LRRCC1	5.87	5111	PCNA	3.86

Table 3. Up-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
29121	CLEC2D	3.83	79682	MLF1IP	3.11
4090	SMAD5	3.80	29123	ANKRD11	3.09
2123	EVI2A	3.80	5429	POLH	3.09
57695	USP37	3.79	701	BUB1B	3.07
133418	EMB	3.76	200030	NBPF11	3.06
4131	MAP1B	3.76	55677	IWS1	3.06
9787	DLGAP5	3.75	160418	TMTC3	3.04
9768	KIAA0101	3.74	9147	NEMF	3.04
54625	PARP14	3.73	11320	MGAT4A	3.04
2215	FCGR3B	3.71	5238	PGM3	3.03
9134	CCNE2	3.70	2820	GPD2	3.02
3117	HLA-DQA1	3.68	388886	FAM211B	3.01
10380	BPNT1	3.67	7852	CXCR4	3.00
79056	PRRG4	3.63	57082	CASC5	2.99
10673	TNFSF13B	3.63	22926	ATF6	2.98
8467	SMARCA5	3.61	7594	ZNF43	2.98
115908	CTHRC1	3.61	968	CD68	2.97
3428	IFI16	3.61	7171	TPM4	2.96
1520	CTSS	3.61	11004	KIF2C	2.96
10797	MTHFD2	3.57	10808	HSPH1	2.95
55681	SCYL2	3.57	84909	C9orf3	2.94
9749	PHACTR2	3.57	1894	ECT2	2.93
94240	EPSTI1	3.56	1629	DBT	2.92
64151	NCAPG	3.51	116969	ART5	2.90
25879	DCAF13	3.51	3227	HOXC11	2.88
116064	LRRC58	3.47	3149	HMGB3	2.87
29899	GPSM2	3.47	10437	IFI30	2.87
135114	HINT3	3.45	57489	ODF2L	2.87
27333	GOLIM4	3.43	2151	F2RL2	2.86
55839	CENPN	3.43	23215	PRRC2C	2.85
23213	SULF1	3.41	128710	C20orf94	2.85
81671	VMP1	3.39	23594	ORC6	2.84
9889	ZBED4	3.36	5205	ATP8B1	2.83
3092	HIP1	3.34	51430	C1orf9	2.80
51512	GTSE1	3.34	57405	SPC25	2.80
92797	HELB	3.34	112401	BIRC8	2.80
51426	POLK	3.30	3606	IL18	2.80
5611	DNAJC3	3.30	115362	GBP5	2.80
6596	HLTF	3.28	50515	CHST11	2.79
9910	RABGAP1L	3.25	83461	CDCA3	2.79
528	ATP6VIC1	3.23	10744	PTTG2	2.78
3833	KIFC1	3.23	51765	MST4	2.77
197131	UBR1	3.20	10926	DBF4	2.76
29923	HILPDA	3.20	27125	AFF4	2.75
28998	MRPL13	3.19	10615	SPAG5	2.75
58527	C6orf115	3.19	55143	CDCA8	2.74
79000	C1orf135	3.19	51602	NOP58	2.74
9857	CEP350	3.18	51478	HSD17B7	2.73
84296	GINS4	3.18	2209	FCGR1A	2.73
81034	SLC25A32	3.15	9958	USP15	2.72
55723	ASF1B	3.14	5469	MED1	2.72
7110	TMF1	3.14	8813	DPM1	2.70
84081	NSRP1	3.14	6731	SRP72	2.70
23075	SWAP70	3.12	9991	PTBP3	2.70

Table 3. Up-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
6726	SRP9	2.69	79866	BORA	2.41
55215	FANCI	2.68	7072	TIA1	2.40
57590	WDFY1	2.67	55632	G2E3	2.40
55142	HAUS2	2.66	2213	FCGR2B	2.40
23047	PDS5B	2.66	3987	LIMS1	2.39
5373	PMM2	2.66	829	CAPZA1	2.39
11065	UBE2C	2.66	26973	CHORDC1	2.38
23085	ERC1	2.66	435	ASL	2.38
389197	C4orf50	2.65	29979	UBQLN1	2.38
11260	XPOT	2.65	8548	BLZF1	2.37
29980	DONSON	2.65	9694	TTC35	2.37
64399	HHIP	2.64	55055	ZWILCH	2.36
6453	ITSN1	2.63	4481	MSR1	2.36
29108	PYCARD	2.63	10213	PSMD14	2.35
9877	ZC3H11A	2.62	9966	TNFSF15	2.35
81624	DIAPH3	2.62	51582	AZIN1	2.35
79723	SUV39H2	2.61	54843	SYTL2	2.34
55789	DEPDC1B	2.61	9039	UBA3	2.33
10097	ACTR2	2.59	933	CD22	2.33
23036	ZNF292	2.58	5685	PSMA4	2.33
22936	ELL2	2.57	9885	OSBPL2	2.33
8477	GPR65	2.57	9262	STK17B	2.33
23397	NCAPH	2.57	56942	C16orf61	2.32
3015	H2AFZ	2.54	10767	HBS1L	2.32
55749	CCAR1	2.53	87178	PNPT1	2.32
25937	WWTR1	2.52	6303	SAT1	2.32
360023	ZBTB41	2.51	7316	UBC	2.32
5080	PAX6	2.51	4205	MEF2A	2.32
4193	MDM2	2.51	85465	EPT1	2.31
24137	KIF4A	2.51	84640	USP38	2.31
9212	AURKB	2.51	5810	RAD1	2.30
168850	ZNF800	2.50	64397	ZFP106	2.29
55109	AGGF1	2.49	5706	PSMC6	2.29
23185	LARP4B	2.49	22948	CCT5	2.29
51571	FAM49B	2.49	10672	GNA13	2.29
51077	FCF1	2.49	339344	MYPOP	2.28
23167	EFR3A	2.49	7292	TNFSF4	2.28
23468	CBX5	2.48	57103	C12orf5	2.28
5396	PRRX1	2.48	388403	YPEL2	2.28
10096	ACTR3	2.47	54876	DCAF16	2.27
10308	ZNF267	2.47	113235	SLC46A1	2.27
6782	HSPA13	2.47	11177	BAZ1A	2.27
3832	KIF11	2.47	339175	METTTL2A	2.26
917	CD3G	2.47	26586	CKAP2	2.26
80821	DDHD1	2.46	55785	FGD6	2.26
52	ACPI	2.46	24145	PANX1	2.25
4179	CD46	2.46	253461	ZBTB38	2.25
10499	NCOA2	2.44	23232	TBC1D12	2.25
60558	GUF1	2.44	995	CDC25C	2.25
55676	SLC30A6	2.43	55974	SLC50A1	2.25
6646	SOAT1	2.43	472	ATM	2.25
5440	POLR2K	2.43	23008	KLHDC10	2.24
84955	NUDCD1	2.42	10024	TROAP	2.24
54739	XAF1	2.42	9521	EEF1E1	2.24

Table 3. Up-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
84295	PHF6	2.23	7402	UTRN	2.09
7295	TXN	2.23	55589	BMP2K	2.08
2710	GK	2.23	158747	MOSPD2	2.08
10905	MAN1A2	2.22	56886	UGGT1	2.07
6780	STAU1	2.22	203100	HTRA4	2.07
10282	BET1	2.22	55279	ZNF654	2.07
134430	WDR36	2.21	54499	TMCO1	2.07
4299	AFF1	2.21	81930	KIF18A	2.07
6747	SSR3	2.21	142686	ASB14	2.06
7334	UBE2N	2.21	55209	SETD5	2.06
5965	RECQL	2.21	9736	USP34	2.04
4605	MYBL2	2.2	116285	ACSM1	2.04
6093	ROCK1	2.19	2201	FBN2	2.04
161725	OTUD7A	2.19	963	CD53	2.04
23518	R3HDM1	2.18	55159	RFWD3	2.03
2239	GPC4	2.18	9871	SEC24D	2.03
28977	MRPL42	2.18	9887	SMG7	2.02
64859	OBFC2A	2.18	23376	UFL1	2.02
3845	KRAS	2.18	79646	PANK3	2.01
51388	NIP7	2.18	50613	UBQLN3	2.00
7586	ZKSCAN1	2.18	201595	STT3B	2.00
10762	NUP50	2.17	59345	GNB4	1.99
7328	UBE2H	2.17	5876	RABGGTB	1.99
10730	YME1L1	2.17	79820	CATSPERB	1.99
23093	TLL5	2.17	6637	SNRPG	1.99
6790	AURKA	2.17	51330	TNFRSF12A	1.99
22889	KIAA0907	2.17	9928	KIF14	1.99
10875	FGL2	2.17	286097	EFHA2	1.98
23161	SNX13	2.17	9131	AIFM1	1.98
9169	SCAF11	2.16	488	ATP2A2	1.98
1788	DNMT3A	2.15	23042	PDXDC1	1.98
9088	PKMYT1	2.15	7114	TMSB4X	1.98
23033	DOPEY1	2.13	9123	SLC16A3	1.98
89882	TPD52L3	2.13	54454	ATAD2B	1.97
6556	SLC11A1	2.13	23143	LRCH1	1.97
64216	TFB2M	2.13	4212	MEIS2	1.97
3071	NCKAP1L	2.13	1457	CSNK2A1	1.97
51068	NMD3	2.13	80012	PHC3	1.97
509	ATP5C1	2.13	128497	SPATA25	1.96
953	ENTPD1	2.13	186	AGTR2	1.96
51105	PHF20L1	2.13	53981	CPSF2	1.96
5062	PAK2	2.13	56996	SLC12A9	1.96
9205	ZMYM5	2.12	1584	CYP11B1	1.96
55157	DARS2	2.12	133619	PRRC1	1.96
8520	HAT1	2.11	4288	MKI67	1.96
79739	TLL7	2.11	9014	TAF1B	1.96
9495	AKAP5	2.10	55858	TMEM165	1.96
3181	HNRNPA2B1	2.10	2212	FCGR2A	1.96
389898	UBE2NL	2.10	10075	HUWE1	1.96
29850	TRPM5	2.10	220988	HNRNPA3	1.96
3070	HELLS	2.10	80146	UXS1	1.95
331	XIAP	2.09	122011	CSNK1A1L	1.95
55751	TMEM184C	2.09	150468	CKAP2L	1.95
2146	EZH2	2.09	84624	FNDC1	1.95

Table 3. Up-regulated genes found in dataset GSE29174. (continue)

NCBI gene ID	Gene Symbol	Fold Change	NCBI gene ID	Gene Symbol	Fold Change
26057	ANKRD17	1.95	7332	UBE2L3	1.92
128061	C1orf131	1.95	3336	HSPE1	1.92
64090	GAL3ST2	1.94	54800	KLHL24	1.92
130507	UBR3	1.93	2290	FOXG1	1.91
2298	FOXD4	1.93	50848	F11R	1.91
123169	LEO1	1.93	10627	MYL12A	1.91
57187	THOC2	1.93	5074	PAWR	1.91
148789	B3GALNT2	1.93	6476	SI	1.91
58508	MLL3	1.92	1009	CDH11	1.90
5701	PSMC2	1.92	29066	ZC3H7A	1.90
148066	ZNRF4	1.92	51319	RSRC1	1.90
6670	SP3	1.92			



Table 4. Raw data count of each prediction database.

Prediction database used in this work and their corresponding unique prediction count were listed. Prediction count: a unique miRNA-target pair was counted as one prediction.

Target prediction database	Prediction count
TargetScan 6.0	4305160
miRanda (Aug. 2010)	7333699
picTar (5sp)	56058



Table 5. Summary of constructed miRNA-regulated networks.

L0 gene: genes connected directly to the miRNA (i.e. direct target of miRNA); **L1 gene**: genes not connected directly to the miRNA.

miRBase Accession	miR name	Total gene count	L0 count	L1 count
MIMAT0002819	hsa-miR-193b-3p	16	1	15
MIMAT0000250	hsa-miR-139-5p	28	10	18
MIMAT0000437	hsa-miR-145-5p	86	22	64
MIMAT0000423	hsa-miR-125b-5p	211	16	195
MIMAT0000443	hsa-miR-125a-5p	206	14	192
MIMAT0000097	hsa-miR-99a-5p	14	1	13
MIMAT0000265	hsa-miR-204-5p	64	18	46
MIMAT0000076	hsa-miR-21-5p	91	16	75
MIMAT0000064	hsa-let-7c	96	20	76
MIMAT0000421	hsa-miR-122-5p	5	3	2
MIMAT0000098	hsa-miR-100-5p	14	1	13
MIMAT0000272	hsa-miR-215	3	3	0
MIMAT0000271	hsa-miR-214-3p	14	8	6
MIMAT0000738	hsa-miR-383	34	3	31
MIMAT0002856	hsa-miR-520d-3p	146	23	123
MIMAT0000077	hsa-miR-22-3p	46	11	35
MIMAT0002820	hsa-miR-497-5p	267	32	235
MIMAT0000089	hsa-miR-31-5p	34	3	31

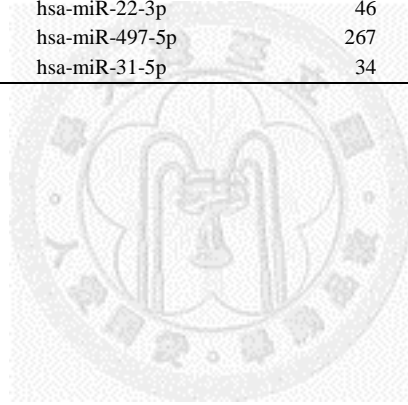


Table 6. Specific enriched GO terms of each miRNA-regulated PINs.

Genes annotated with the specific GO term in the PIN were also listed in this table. **Adj. p-value:** multiple-test adjusted p-value calculated by the method described in the work of Benjamini and Yekutieli [42].

MIMAT0002856(hsa-miR-520d-3p)		
GO term	Genes	Adj. p-value
GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	SH3KBP1, HDAC2, RET, ABI1, LYN, GRB2, SORBS1, CLTC, CLTA, CDC42, CASP9, RAF1, SRC, AP2A1, AP2B1, MAPK3, ARHGEF7, PRKCA, RPS6, PRKAR2B, MAPK1, ARHGEF6, CDK1, SH3GL2, EIF4G1, HDAC1, ECT2, MKNK1, CASP3, PRKACA, ADRB2, PRKAR2A, EIF4B, SHC1, RAC1	2.77E-29
GO:0048011, Nerve growth factor receptor signaling pathway	HDAC2, GRB2, CLTC, CLTA, CASP9, RAF1, SRC, AP2A1, AP2B1, MAPK3, ARHGEF7, PRKCA, PRKAR2B, MAPK1, ARHGEF6, CDK1, SH3GL2, HDAC1, ECT2, CASP3, PRKACA, PRKAR2A, SHC1, RAC1	5.09E-24
GO:0007173, Epidermal growth factor receptor signaling pathway	SH3KBP1, GRB2, CLTC, CLTA, CDC42, CASP9, RAF1, SRC, AP2A1, AP2B1, MAPK3, ARHGEF7, PRKCA, PRKAR2B, MAPK1, CDK1, SH3GL2, PRKACA, PRKAR2A, SHC1	2.96E-22
GO:0043067, Regulation of programmed cell death	HDAC2, STK17B, ESR1, ABL1, LYN, TP53, GABRB3, PAK2, LCK, CASP9, RAF1, PLK1, ARHGEF7, PRKCA, RPS6, SH3RF1, MAPK1, IFT57, ARHGAP10, ARHGEF6, CDK1, APAF1, HDAC1, ECT2, CASP3, SOX10, EP300, ARAF, TFAP2A, ADRB2, HCK, KLHL20, CASP8, HIP1, RAC1	4.76E-19
GO:0042058, Regulation of epidermal growth factor receptor signaling pathway	SH3KBP1, ESR1, GRB2, CLTC, CLTA, CDC42, AP2A1, AP2B1, ARHGEF7, SH3GL2, SHC1	2.36E-12
GO:0008543, Fibroblast growth factor receptor signaling pathway	GRB2, CASP9, RAF1, SRC, MAPK3, PRKCA, PRKAR2B, MAPK1, CDK1, MKNK1, PRKACA, PRKAR2A, SHC1	2.39E-12
GO:0043068, Positive regulation of programmed cell death	STK17B, ABL1, LYN, TP53, LCK, CASP9, ARHGEF7, PRKCA, RPS6, SH3RF1, MAPK1, ARHGEF6, APAF1, ECT2, CASP3, EP300, TFAP2A, ADRB2, CASP8, HIP1, RAC1	3.09E-12
GO:0010942, Positive regulation of cell death	STK17B, ABL1, LYN, TP53, LCK, CASP9, ARHGEF7, PRKCA, RPS6, SH3RF1, MAPK1, ARHGEF6, APAF1, ECT2, CASP3, EP300, TFAP2A, ADRB2, CASP8, HIP1, RAC1	4.49E-12
GO:0006917, Induction of apoptosis	STK17B, ABL1, TP53, LCK, CASP9, ARHGEF7, PRKCA, SH3RF1, MAPK1, ARHGEF6, APAF1, ECT2, CASP3, EP300, CASP8, HIP1, RAC1	6.91E-11
GO:0012502, Induction of programmed cell death	STK17B, ABL1, TP53, LCK, CASP9, ARHGEF7, PRKCA, SH3RF1, MAPK1, ARHGEF6, APAF1, ECT2, CASP3, EP300, CASP8, HIP1, RAC1	7.42E-11

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0002856(hsa-miR-520d-3p)		
GO term	Genes	Adj. p-value
GO:0042059, Negative regulation of epidermal growth factor receptor signaling pathway	SH3KBP1, GRB2, CLTC, CLTA, CDC42, AP2A1, AP2B1, ARHGEF7, SH3GL2	8.63E-11
GO:0015630, Microtubule cytoskeleton	STMN1, SORBS1, SMAD4, CLTC, CDC42, LCK, RACGAP1, PLK1, PRKAR2B, YES1, MAPK1, IFT57, CDK1, ECT2, PRKACA, RB1, EP300, CCNB1, CHAF1B, TFAP2A, CASP8, PRKAR2A	4.75E-10
GO:0060548, Negative regulation of cell death	HDAC2, ESR1, TP53, SMAD4, RAF1, PLK1, PRKCA, RPS6, SH3RF1, CDK1, HDAC1, CASP3, SOX10, ARAF, TFAP2A, HCK, KLHL20	6.27E-08
GO:0008286, Insulin receptor signaling pathway	GRB2, SORBS1, RAF1, MAPK3, RPS6, MAPK1, CDK1, EIF4G1, EIF4B, SHC1	2.15E-07
GO:0043069, Negative regulation of programmed cell death	HDAC2, ESR1, TP53, RAF1, PLK1, PRKCA, RPS6, SH3RF1, CDK1, HDAC1, CASP3, SOX10, ARAF, TFAP2A, HCK, KLHL20	3.13E-07
GO:0008284, Positive regulation of cell proliferation	HDAC2, ESR1, LYN, CDC42, E2F1, PRKCA, MAPK1, CDK1, RHOG, HDAC1, NCK1, SOX10, CCNB1, ADRB2, HCK, SHC1	8.34E-07
GO:0051988, Regulation of attachment of spindle microtubules to kinetochore	CDC42, RACGAP1, ECT2, CCNB1	3.27E-05
GO:0008629, Induction of apoptosis by intracellular signals	ABL1, TP53, CASP9, APAF1, CASP3, EP300, CASP8	5.83E-05
MIMAT0002820(hsa-miR-497-5p)		
GO term	Genes	Adj. p-value
GO:0043067, Regulation of programmed cell death	ESR1, MEN1, ABL1, HIPK3, PPARGC1A, SIAH1, SH3RF1, PAK2, LCK, MED1, PPARG, CBX4, ARHGEF7, YWHAB, RXRA, ACVR1, MAPK1, CASP3, CASP6, AR, PTPRF, MDM2, BRCA1, MLH1, RAB27A, PIAS4, FAF1, RAC1, VHL, SKI, NR4A1, LYN, TP53, PSMC2, GATA1, GATA6, GATA3, RAF1, CDKN1B, PLK1, PSMD11, HOXA13, RPS6, ESR2, ARHGAP10, ARHGEF6, SMAD3, SKIL, RYR2, PSEN1, HCK, TRAF2	2.67E-25
GO:0043068, Positive regulation of programmed cell death	MEN1, ABL1, SIAH1, SH3RF1, LCK, PPARG, ARHGEF7, YWHAB, RXRA, MAPK1, CASP3, CASP6, PTPRF, BRCA1, MLH1, RAB27A, PIAS4, FAF1, RAC1, NR4A1, LYN, TP53, GATA6, CDKN1B, HOXA13, RPS6, ESR2, ARHGEF6, SMAD3, RYR2, PSEN1, TRAF2	1.74E-17
GO:0010942, Positive regulation of cell death	MEN1, ABL1, SIAH1, SH3RF1, LCK, PPARG, ARHGEF7, YWHAB, RXRA, MAPK1, CASP3, CASP6, PTPRF, BRCA1, MLH1, RAB27A, PIAS4, FAF1, RAC1, NR4A1, LYN, TP53, GATA6, CDKN1B, HOXA13, RPS6, ESR2, ARHGEF6, SMAD3, RYR2, PSEN1, TRAF2	3.08E-17

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0002820(hsa-miR-497-5p)		
GO term	Genes	Adj. p-value
GO:0008285, Negative regulation of cell proliferation	MEN1, MED1, PPARG, RXRA, CASP3, AR, PTPRF, VDR, VHL, SKI, LYN, TP53, TOB1, GATA1, GATA3, RAF1, HNF4A, CDKN1B, BRD7, MED25, ESR2, ABI1, SMAD1, SMAD2, SMAD3, SMAD4, SOX7	3.85E-14
GO:0015629, Actin cytoskeleton	ABL1, SORBS1, FLNA, SEPT7, ANLN, MACF1, HAP1, SH3PXD2A, IQGAP2, BRCA1, ACTC1, ACTA1, MYL2, MYLK, SORBS2, ARPC4, ARPC5, ACTR2, ACTR3, ARPC1B, WASF1, WASF2, HCK	2.52E-13
GO:0006917, Induction of apoptosis	ABL1, SH3RF1, LCK, PPARG, ARHGEF7, YWHAB, MAPK1, CASP3, CASP6, BRCA1, MLH1, RAB27A, RAC1, NR4A1, TP53, CDKN1B, ARHGEF6, SMAD3, RYR2, PSEN1, TRAF2	9.69E-11
GO:0012502, Induction of programmed cell death	ABL1, SH3RF1, LCK, PPARG, ARHGEF7, YWHAB, MAPK1, CASP3, CASP6, BRCA1, MLH1, RAB27A, RAC1, NR4A1, TP53, CDKN1B, ARHGEF6, SMAD3, RYR2, PSEN1, TRAF2	1.06E-10
GO:0007178, Transmembrane receptor protein serine/threonine kinase signaling pathway	ACVR1, SMURF2, SKI, GDF6, BMP6, ZNF8, GATA4, HNF4A, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, RYR2	1.22E-10
GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	SORBS1, CDC42, SRC, MAPK3, ARHGEF7, YWHAB, MAPK1, SH3GL2, CASP3, MDM2, EIF4G1, RAC1, SH3KBP1, NR4A1, LYN, GRB2, RAF1, CDKN1B, RPS6, ABI1, ARHGEF6, MKNK1, PSEN1, EIF4B	1.67E-10
GO:0090092, Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	MEN1, ACVR1, SMURF2, SKI, GDF6, TP53, BMP6, GATA4, GATA6, HOXA13, SMAD2, SMAD3, SMAD4, SKIL	7.51E-10
GO:0030509, BMP signaling pathway	ACVR1, SMURF2, SKI, GDF6, BMP6, ZNF8, SMAD1, SMAD4, SMAD5, RYR2	3.25E-09
GO:0060548, Negative regulation of cell death	ESR1, HIPK3, PPARGC1A, SH3RF1, MED1, CBX4, ACVR1, CASP3, AR, MDM2, VHL, TP53, GATA1, GATA6, GATA3, RAF1, CDKN1B, PLK1, RPS6, SMAD3, SMAD4, PSEN1, HCK	7.88E-09
GO:0007173, Epidermal growth factor receptor signaling pathway	CDC42, SRC, MAPK3, ARHGEF7, YWHAB, MAPK1, SH3GL2, MDM2, SH3KBP1, NR4A1, GRB2, RAF1, CDKN1B	9.22E-09
GO:0030521, Androgen receptor signaling pathway	PPARGC1A, MED14, MED1, AR, BRCA1, MED12, PIAS1, RAN, NR1I3	1.42E-08
GO:0043069, Negative regulation of programmed cell death	ESR1, HIPK3, PPARGC1A, SH3RF1, MED1, CBX4, ACVR1, CASP3, AR, MDM2, VHL, TP53, GATA1, GATA6, GATA3, RAF1, CDKN1B, PLK1, RPS6, SMAD3, PSEN1, HCK	2.44E-08
GO:0048011, Nerve growth factor receptor signaling pathway	SRC, MAPK3, ARHGEF7, YWHAB, MAPK1, SH3GL2, CASP3, MDM2, RAC1, NR4A1, GRB2, RAF1, CDKN1B, ARHGEF6, PSEN1	2.64E-08
GO:0032956, Regulation of actin cytoskeleton organization	ABL1, LRP1, ARPC4, ARPC5, ACTR3, ARPC1B, SMAD3, NCK1, SORBS3, HCK, LIMK1	6.42E-06

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0002820(hsa-miR-497-5p)		
GO term	Genes	Adj. p-value
GO:0008543, Fibroblast growth factor receptor signaling pathway	SRC, MAPK3, YWHAB, MAPK1, MDM2, NR4A1, GRB2, RAF1, CDKN1B, MKNK1	9.06E-06
GO:0042059, Negative regulation of epidermal growth factor receptor signaling pathway	CDC42, ARHGEF7, SH3GL2, PTPRF, SH3KBP1, GRB2, PSEN1	1.11E-05
GO:0042058, Regulation of epidermal growth factor receptor signaling pathway	ESR1, CDC42, ARHGEF7, SH3GL2, PTPRF, SH3KBP1, GRB2, PSEN1	1.17E-05
GO:0007179, Transforming growth factor beta receptor signaling pathway	ACVR1, SMURF2, SKI, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5	1.67E-05
GO:0015630, Microtubule cytoskeleton	STMN1, RIF1, SORBS1, CDC42, LCK, RACGAP1, YES1, YWHAB, MAPK1, SEPT7, KIF23, CDC16, MACF1, BRCA1, FEZ1, NCOR1, PLK1, CHD3, SMAD4, CEP350, CDC27, PSEN1	2.14E-05
GO:0017015, Regulation of transforming growth factor beta receptor signaling pathway	MEN1, SMURF2, SKI, TP53, SMAD2, SMAD3, SMAD4, SKIL	2.30E-05
GO:0070302, Regulation of stress-activated protein kinase signaling cascade	MEN1, ZEB2, HIPK3, SH3RF1, CDC42, MAPK3, MAPK1, LYN, NCOR1, TRAF2	2.74E-05
GO:0001959, Regulation of cytokine-mediated signaling pathway	HSP90AB1, MED1, PPARG, PTPRF, NR1H2, PIAS1, IL36RN, HIPK1	6.35E-05
GO:0008284, Positive regulation of cell proliferation	ESR1, CDC42, MED1, RARA, MAPK1, AR, MDM2, NR4A1, LYN, FZR1, BMP6, GATA1, GATA4, GATA6, CDKN1B, NCK1, HCLS1, HCK	7.29E-05
MIMAT0000423(hsa-miR-125b-5p)		
GO term	Genes	Adj. p-value
GO:0043067, Regulation of programmed cell death	HMGA2, PML, PRNP, FGF2, XRCC4, BRCA1, IGFBP3, HDAC3, CTNNB1, CD5, CDK1, NKX2-5, MEF2C, PRKCI, CASP2, PSMA4, PSMA3, CFDP1, CAV1, FAF1, YWHAB, HIF1A, RELA, TCF7L2, TNFSF12, PSEN2, TP53, TOP2A, TNFRSF4, BID, MYC, JUN, OGT, CDKN1A, HOXA13, RNF7, PPP2R4, HDAC2, HDAC1, SNCA, PTEN, NFKBIA, IFI16, NOL3, TRAF2, HSP90B1	4.56E-24
GO:0060548, Negative regulation of cell death	HMGA2, PRNP, FGF2, XRCC4, HDAC3, CTNNB1, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, HIF1A, RELA, TCF7L2, PSEN2, TP53, TNFRSF4, MYC, JUN, CDKN1A, RNF7, HDAC2, HDAC1, SNCA, PTEN, MGMT, NFKBIA, NOL3, HSP90B1	8.04E-17
GO:0008284, Positive regulation of cell proliferation	HMGA2, FGF2, XRCC4, CDC25B, CTNNB1, EGR1, AGGF1, CDK1, NKX2-5, MEF2C, PRKCI, IRS1, HIF1A, RELA, HCLS1, TNFSF12, ARNT, PTPRC, TNFSF4, TNFRSF4, MYC, JUN, FGF1, CDKN1A, HDAC2, HDAC1, NOLC1, PTEN	2.20E-15

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000423(hsa-miR-125b-5p)		
GO term	Genes	Adj. p-value
GO:0043069, Negative regulation of programmed cell death	HMGA2, PRNP, XRCC4, HDAC3, CTNNB1, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, HIF1A, RELA, TCF7L2, PSEN2, TP53, TNFRSF4, MYC, JUN, CDKN1A, RNF7, HDAC2, HDAC1, SNCA, PTEN, NFKBIA, NOL3, HSP90B1	3.62E-15
GO:0043068, Positive regulation of programmed cell death	HMGA2, PML, BRCA1, IGFBP3, CTNNB1, CD5, MEF2C, PRKCI, CASP2, CAV1, FAF1, YWHAB, TNFSF12, PSEN2, TP53, TOP2A, BID, JUN, OGT, CDKN1A, HOXA13, RNF7, PPP2R4, PTEN, IFI16, TRAF2	4.51E-14
GO:0010942, Positive regulation of cell death	HMGA2, PML, BRCA1, IGFBP3, CTNNB1, CD5, MEF2C, PRKCI, CASP2, CAV1, FAF1, YWHAB, TNFSF12, PSEN2, TP53, TOP2A, BID, JUN, OGT, CDKN1A, HOXA13, RNF7, PPP2R4, PTEN, IFI16, TRAF2	7.15E-14
GO:0006916, Anti-apoptosis	PRNP, HDAC3, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, RELA, TCF7L2, PSEN2, RNF7, HDAC1, SNCA, NFKBIA, NOL3, HSP90B1	1.25E-10
GO:0008285, Negative regulation of cell proliferation	SERPINF1, SRF, PML, PRNP, FGF2, CSNK2B, IGFBP3, CTNNB1, CAV1, HMGA1, VDR, CDH5, HSF1, COL18A1, TP53, MYC, JUN, CDKN1A, PAK1, PTEN	2.08E-09
GO:0015630, Microtubule cytoskeleton	STMN1, KIF1C, RANGAP1, CDC25B, BRCA1, HDAC3, CTNNB1, PIN4, HSPH1, RANBP9, CDK1, SPTAN1, YWHAQ, DVL1, FKBP4, YWHAB, CCDC85B, MAPT, PSEN2, TOP2A, SPIB, MYC, OGT, APEX1, PAFAH1B1	2.24E-09
GO:0048011, Nerve growth factor receptor signaling pathway	HDAC3, CDK1, MEF2C, PRKCI, CASP2, IRS1, YWHAB, RELA, PSEN2, HDAC2, HDAC1, PTEN, ATF1, NFKBIA	1.99E-08
GO:0050678, Regulation of epithelial cell proliferation	SERPINF1, PGR, FGF2, CTNNB1, AGGF1, CAV1, HIF1A, TNFSF12, ARNT, MYC, JUN, FGF1, PTEN	3.41E-08
GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	FGF2, HDAC3, CDK1, MEF2C, PRKCI, CASP2, IRS1, FIBP, PTPN1, YWHAB, RELA, PSEN2, FGF1, HDAC2, HDAC1, PTEN, ATF1, NFKBIA, EIF4EBP1	6.39E-08
GO:0006917, Induction of apoptosis	PML, BRCA1, CD5, CASP2, CAV1, YWHAB, TNFSF12, PSEN2, TP53, BID, OGT, CDKN1A, RNF7, PTEN, IFI16, TRAF2	1.68E-07
GO:0012502, Induction of programmed cell death	PML, BRCA1, CD5, CASP2, CAV1, YWHAB, TNFSF12, PSEN2, TP53, BID, OGT, CDKN1A, RNF7, PTEN, IFI16, TRAF2	1.81E-07
GO:0035666, TRIF-dependent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.40E-06
GO:0034138, Toll-like receptor 3 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.71E-06
GO:0051693, Actin filament capping	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	3.43E-06

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000423(hsa-miR-125b-5p)		
GO term	Genes	Adj. p-value
GO:0002756, MyD88-independent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	3.82E-06
GO:0034134, Toll-like receptor 2 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	5.33E-06
GO:0034130, Toll-like receptor 1 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	5.33E-06
GO:0030835, Negative regulation of actin filament depolymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	5.58E-06
GO:0015629, Actin cytoskeleton	WAS, CDH1, BRCA1, SPTB, SPTBN1, SPTAN1, CTDPI, STX1A, SPTA1, PAK1, SNCA, ADD1, EPB41, EPB49	5.58E-06
GO:0002755, MyD88-dependent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	7.66E-06
GO:0034142, Toll-like receptor 4 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	1.14E-05
GO:0050679, Positive regulation of epithelial cell proliferation	FGF2, CTNNB1, AGGF1, HIF1A, TNFSF12, ARNT, MYC, JUN, FGF1	1.14E-05
GO:0030834, Regulation of actin filament depolymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	1.27E-05
GO:0030837, Negative regulation of actin filament polymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	2.71E-05
GO:0002224, Toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.96E-05
GO:0008629, Induction of apoptosis by intracellular signals	PML, BRCA1, YWHAB, TP53, BID, CDKN1A, RNF7, IFI16	3.52E-05
MIMAT0000076(hsa-miR-21-5p)		
GO term	Genes	Adj. p-value
GO:0043067, Regulation of programmed cell death	SPRY2, TP53, ADAMTSL4, ETS1, TDGF1, RAF1, HOXA5, HOXA13, MSX1, MSX2, NKX2-5, CBL, INHBB, COL4A3, ACVR1C, TRAF2	1.92E-05

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000076(hsa-miR-21-5p)		
GO term	Genes	Adj. p-value
GO:0007173, Epidermal growth factor receptor signaling pathway	SPRY2, SPRY1, GRB2, PTPN11, TDGF1, RAF1, CBL	9.44E-05
MIMAT0000250(hsa-miR-139-5p)		
GO term	Genes	Adj. p-value
GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	LTBP1, FBN1, FBN2	2.61E-05
MIMAT0000089(hsa-miR-31-5p)		
GO term	Genes	Adj. p-value
GO:0007187, G-protein signaling, coupled to cyclic nucleotide second messenger	GNA12, GNA13, DRD5, MTNR1A, S1PR3, TSHR, S1PR4	8.74E-07
GO:0019935, Cyclic-nucleotide-mediated signaling	GNA12, GNA13, DRD5, MTNR1A, S1PR3, TSHR, S1PR4	1.60E-06
GO:0007188, G-protein signaling, coupled to cAMP nucleotide second messenger	GNA12, GNA13, DRD5, S1PR3, TSHR, S1PR4	6.82E-06
GO:0048011, Nerve growth factor receptor signaling pathway	ARHGEF1, PRKCD, ARHGEF12, PRKACA, PRKCE, MCF2, ARHGEF11	6.93E-06
GO:0019933, CAMP-mediated signaling	GNA12, GNA13, DRD5, S1PR3, TSHR, S1PR4	1.05E-05
GO:0043067, Regulation of programmed cell death	PTK2B, PRKCD, TGFBR1, ARHGEF12, CTNNB1, PRKCE, TIA1, MCF2, FASTK, ARHGEF11, F2R	1.25E-05
GO:0003376, Sphingosine-1-phosphate signaling pathway	S1PR3, S1PR2, S1PR4	3.22E-05
GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	PTK2B, ARHGEF1, PRKCD, ARHGEF12, PRKACA, PRKCE, MCF2, ARHGEF11	5.20E-05
GO:0043068, Positive regulation of programmed cell death	TGFBR1, ARHGEF12, CTNNB1, PRKCE, TIA1, MCF2, FASTK, ARHGEF11	7.67E-05
GO:0010942, Positive regulation of cell death	TGFBR1, ARHGEF12, CTNNB1, PRKCE, TIA1, MCF2, FASTK, ARHGEF11	8.78E-05
GO:0006917, Induction of apoptosis	TGFBR1, ARHGEF12, PRKCE, TIA1, MCF2, FASTK, ARHGEF11	9.32E-05

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000089(hsa-miR-31-5p)		
GO term	Genes	Adj. p-value
GO:0012502, Induction of programmed cell death	TGFBR1, ARHGEF12, PRKCE, TIA1, MCF2, FASTK, ARHGEF11	9.51E-05
MIMAT0000437(hsa-miR-145-5p)		
GO term	Genes	Adj. p-value
GO:0030509, BMP signaling pathway	BMP6, ZNF8, ACVR1, SMAD1, SMAD4, RYR2, SMAD5, SMURF2, GDF6	1.37E-11
GO:0007178, Transmembrane receptor protein serine/threonine kinase signaling pathway	BMP6, ZNF8, ACVR1, SMAD1, SMAD4, RYR2, SMAD5, SMURF2, GDF6	2.56E-08
GO:0090092, Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	MEN1, TP53, BMP6, HOXA13, ACVR1, SMAD4, SULF1, SMURF2, GDF6	8.22E-08
MIMAT0000064(hsa-let-7c)		
GO term	Genes	Adj. p-value
GO:0043067, Regulation of programmed cell death	RRM2B, ACVR1, TP53, RASA1, TGFBR1, PSMA3, BIRC5, ACTN2, HOXA13, IRS2, FASTK, VAV1, PSMB6, BCL2, CDK1, HDAC1, SOX10, TIA1, AKT1, AURKB	3.43E-08
GO:0043069, Negative regulation of programmed cell death	RRM2B, ACVR1, TP53, RASA1, TGFBR1, BIRC5, IRS2, BCL2, CDK1, HDAC1, SOX10, AKT1, AURKB	6.58E-06
GO:0060548, Negative regulation of cell death	RRM2B, ACVR1, TP53, RASA1, TGFBR1, BIRC5, IRS2, BCL2, CDK1, HDAC1, SOX10, AKT1, AURKB	8.92E-06
GO:0015630, Microtubule cytoskeleton	INCENP, SNTB2, SEPT1, TACC1, BIRC5, RACGAP1, PIN4, CDCA8, CDK1, PHF1, AKT1, AURKB, NINL, CCDC85B	5.15E-05
MIMAT0000443(hsa-miR-125a-5p)		
GO term	Genes	Adj. p-value
GO:0043067, Regulation of programmed cell death	HMGA2, PML, PRNP, FGF2, XRCC4, BRCA1, IGFBP3, HDAC3, CTNNB1, CD5, CDK1, NKX2-5, MEF2C, PRKCI, CASP2, PSMA4, PSMA3, CFDP1, CAV1, FAF1, YWHAB, HIF1A, RELA, TCF7L2, TNFSF12, PSEN2, TP53, TOP2A, TNFRSF4, BID, MYC, JUN, OGT, CDKN1A, RNF7, PPP2R4, HDAC2, HDAC1, SNCA, PTEN, NFKBIA, IFI16, NOL3, TRAF2, HSP90B1	1.62E-23
GO:0060548, Negative regulation of cell death	HMGA2, PRNP, FGF2, XRCC4, HDAC3, CTNNB1, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, HIF1A, RELA, TCF7L2, PSEN2, TP53, TNFRSF4, MYC, JUN, CDKN1A, RNF7, HDAC2, HDAC1, SNCA, PTEN, MGMT, NFKBIA, NOL3, HSP90B1	4.53E-17
GO:0008284, Positive regulation of cell proliferation	HMGA2, FGF2, XRCC4, CDC25B, CTNNB1, EGR1, AGGF1, CDK1, NKX2-5, MEF2C, PRKCI, IRS1, HIF1A, RELA, HCLS1, TNFSF12, ARNT, PTPRC, TNFSF4, TNFRSF4, MYC, JUN, FGF1, CDKN1A, HDAC2, HDAC1, NOLC1, PTEN	1.23E-15

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000443(hsa-miR-125a-5p)		
GO term	Genes	Adj. p-value
GO:0043069, Negative regulation of programmed cell death	HMGA2, PRNP, XRCC4, HDAC3, CTNNB1, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, HIF1A, RELA, TCF7L2, PSEN2, TP53, TNFRSF4, MYC, JUN, CDKN1A, RNF7, HDAC2, HDAC1, SNCA, PTEN, NFKBIA, NOL3, HSP90B1	2.02E-15
GO:0043068, Positive regulation of programmed cell death	HMGA2, PML, BRCA1, IGFBP3, CTNNB1, CD5, MEF2C, PRKCI, CASP2, CAV1, FAF1, YWHAB, TNFSF12, PSEN2, TP53, TOP2A, BID, JUN, OGT, CDKN1A, RNF7, PPP2R4, PTEN, IFI16, TRAF2	2.80E-13
GO:0010942, Positive regulation of cell death	HMGA2, PML, BRCA1, IGFBP3, CTNNB1, CD5, MEF2C, PRKCI, CASP2, CAV1, FAF1, YWHAB, TNFSF12, PSEN2, TP53, TOP2A, BID, JUN, OGT, CDKN1A, RNF7, PPP2R4, PTEN, IFI16, TRAF2	4.35E-13
GO:0006916, Anti-apoptosis	PRNP, HDAC3, CDK1, NKX2-5, MEF2C, PRKCI, CFDP1, RELA, TCF7L2, PSEN2, RNF7, HDAC1, SNCA, NFKBIA, NOL3, HSP90B1	9.04E-11
GO:0008285, Negative regulation of cell proliferation	SERPINF1, SRF, PML, PRNP, FGF2, CSNK2B, IGFBP3, CTNNB1, CAV1, HMGA1, VDR, CDH5, HSF1, COL18A1, TP53, MYC, JUN, CDKN1A, PAK1, PTEN	1.32E-09
GO:0015630, Microtubule cytoskeleton	STMN1, KIF1C, RANGAP1, CDC25B, BRCA1, HDAC3, CTNNB1, PIN4, HSPH1, RANBP9, CDK1, SPTAN1, YWHAQ, DVL1, FKBP4, YWHAB, CCDC85B, MAPT, PSEN2, TOP2A, SPIB, MYC, OGT, APEX1, PAFAH1B1	1.32E-09
GO:0048011, Nerve growth factor receptor signaling pathway	HDAC3, CDK1, MEF2C, PRKCI, CASP2, IRS1, YWHAB, RELA, PSEN2, HDAC2, HDAC1, PTEN, ATF1, NFKBIA	1.50E-08
GO:0050678, Regulation of epithelial cell proliferation	SERPINF1, PGR, FGF2, CTNNB1, AGGF1, CAV1, HIF1A, TNFSF12, ARNT, MYC, JUN, FGF1, PTEN	2.66E-08
GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	FGF2, HDAC3, CDK1, MEF2C, PRKCI, CASP2, IRS1, FIBP, PTPN1, YWHAB, RELA, PSEN2, FGF1, HDAC2, HDAC1, PTEN, ATF1, NFKBIA, EIF4EBP1	4.37E-08
GO:0006917, Induction of apoptosis	PML, BRCA1, CD5, CASP2, CAV1, YWHAB, TNFSF12, PSEN2, TP53, BID, OGT, CDKN1A, RNF7, PTEN, IFI16, TRAF2	1.22E-07
GO:0012502, Induction of programmed cell death	PML, BRCA1, CD5, CASP2, CAV1, YWHAB, TNFSF12, PSEN2, TP53, BID, OGT, CDKN1A, RNF7, PTEN, IFI16, TRAF2	1.31E-07
GO:0035666, TRIF-dependent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.01E-06
GO:0034138, Toll-like receptor 3 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.27E-06
GO:0051693, Actin filament capping	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	2.97E-06

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000443(hsa-miR-125a-5p)		
GO term	Genes	Adj. p-value
GO:0002756, MyD88-independent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	3.20E-06
GO:0015629, Actin cytoskeleton	WAS, CDH1, BRCA1, SPTB, SPTBN1, SPTAN1, CTDP1, STX1A, SPTA1, PAK1, SNCA, ADD1, EPB41, EPB49	4.25E-06
GO:0034134, Toll-like receptor 2 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	4.43E-06
GO:0034130, Toll-like receptor 1 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	4.43E-06
GO:0030835, Negative regulation of actin filament depolymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	4.72E-06
GO:0002755, MyD88-dependent toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	6.41E-06
GO:0050679, Positive regulation of epithelial cell proliferation	FGF2, CTNNB1, AGGF1, HIF1A, TNFSF12, ARNT, MYC, JUN, FGF1	9.31E-06
GO:0034142, Toll-like receptor 4 signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	9.31E-06
GO:0030834, Regulation of actin filament depolymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	1.09E-05
GO:0030837, Negative regulation of actin filament polymerization	SPTB, SPTBN1, SPTAN1, SPTA1, ADD1, EPB49	2.37E-05
GO:0002224, Toll-like receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	2.48E-05
GO:0008629, Induction of apoptosis by intracellular signals	PML, BRCA1, YWHAB, TP53, BID, CDKN1A, RNF7, IFI16	2.92E-05
GO:0050851, Antigen receptor-mediated signaling pathway	WAS, MEF2C, RELA, PSEN2, PTPRC, PAK1, PTEN, NFKBIA	8.55E-05
GO:0002221, Pattern recognition receptor signaling pathway	ATF2, CDK1, MEF2C, FOS, RELA, JUN, ATF1, NFKBIA	9.15E-05

Table 6. Specific enriched GO terms of each miRNA-regulated PINs. (continue)

MIMAT0000443(hsa-miR-125a-5p)		
GO term	Genes	Adj. p-value
GO:0001936, Regulation of endothelial cell proliferation	FGF2, AGGF1, CAV1, HIF1A, TNFSF12, ARNT, JUN	9.47E-05
MIMAT0000077(hsa-miR-22-3p)		
GO term	Genes	Adj. p-value
GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	FBN2, FBN1, LTBP1	1.45E-05
MIMAT0000265(hsa-miR-204-5p)		
GO term	Genes	Adj. p-value
GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	FBN1, FBN2, LTBP1	6.31E-05

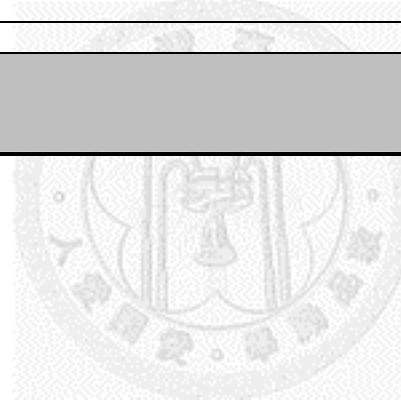


Table 7. P-value of activation status of each miRNA-regulated PINs.

*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

miRNA	Activation p-value of normal samples	Activation p-value of tumor samples
MIMAT0002856(hsa-miR-520d-3p)	0.389964	0.002018**
MIMAT0000265(hsa-miR-204-5p)	0.150553	0.033994*
MIMAT0000272(hsa-miR-215)	1	1
MIMAT0000271(hsa-miR-214-3p)	0.352249	1
MIMAT0002820(hsa-miR-497-5p)	0.032329*	0.582514
MIMAT0000076(hsa-miR-21-5p)	0.507891	0.998652
MIMAT0000738(hsa-miR-383)	0.285448	0.241169
MIMAT0000250(hsa-miR-139-5p)	0.672676	0.175408
MIMAT0000064(hsa-let-7c)	0.877611	0.014578*
MIMAT0000089(hsa-miR-31-5p)	0.009438**	0.729054
MIMAT0000077(hsa-miR-22-3p)	0.352565	0.06039
MIMAT0000098(hsa-miR-100-5p)	0.006016***	0.082638
MIMAT0000097(hsa-miR-99a-5p)	0.006016***	0.082638
MIMAT0000443(hsa-miR-125a-5p)	2.28E-07***	0.806962
MIMAT0002819(hsa-miR-193b-3p)	0.605696	0.719147
MIMAT0000423(hsa-miR-125b-5p)	3.25E-07***	0.74226
MIMAT0000437(hsa-miR-145-5p)	0.000132***	0.448992
MIMAT0000421(hsa-miR-122-5p)	1	0.155739



Table 8. Result of ROC curve analysis on NTUH miRNA array data.

ROC analysis was done to validate the diagnostic value of the miRNA in the miRNA-regulated PINs. **AUC**: area under (ROC) curve; *****: p-value < 0.05; *******: p-value < 0.001.

miRBase Accession	miRNA name	AUC	p-value
MIMAT0002856	hsa-miR-520d-3p	0.49	0.549112
MIMAT0000265	hsa-miR-204-5p	0.98	6.47E-10***
MIMAT0000272	hsa-miR-215	0.21	0.999782
MIMAT0000271	hsa-miR-214-3p	0.68	0.010387*
MIMAT0002820	hsa-miR-497-5p	0.99	2.75E-10***
MIMAT0000076	hsa-miR-21-5p	0.78	0.000184***
MIMAT0000738	hsa-miR-383	0.60	0.106284
MIMAT0000423	hsa-miR-125b-5p	0.99	2.48E-10***
MIMAT0000064	hsa-let-7c	0.93	3.79E-08***
MIMAT0000089	hsa-miR-31-5p	0.80	8.63E-05***
MIMAT0000077	hsa-miR-22-3p	0.27	0.99749
MIMAT0000098	hsa-miR-100-5p	0.98	5.55E-10***
MIMAT0000097	hsa-miR-99a-5p	0.99	2.55E-10***
MIMAT0000443	hsa-miR-125a-5p	0.31	0.990694
MIMAT0002819	hsa-miR-193b-3p	0.41	0.86128
MIMAT0000250	hsa-miR-139-5p	0.99	2.42E-10***
MIMAT0000437	hsa-miR-145-5p	0.96	3.14E-09***
MIMAT0000421	hsa-miR-122-5p	0.48	0.597483

Table 9. GOBO survival analysis results.

Genes which was annotated with the specified GO term of proteins in the PIN would be used as input gene set for GOBO analysis. *: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

miRNA	GO term	P-value
MIMAT0000064(hsa-let-7c)	GO:0015630, Microtubule cytoskeleton	9.97E-06***
	GO:0043067, Regulation of programmed cell death	0.268067
	GO:0043069, Negative regulation of programmed cell death	0.0390439
	GO:0060548, Negative regulation of cell death	0.0390439
MIMAT0000076(hsa-miR-21-5p)	GO:0007173, Epidermal growth factor receptor signaling pathway	0.721139
	GO:0043067, Regulation of programmed cell death	0.266312
MIMAT0000077(hsa-miR-22-3p)	GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	0.940727
	GO:0003376, Sphingosine-1-phosphate signaling pathway	0.062202
MIMAT0000089(hsa-miR-31-5p)	GO:0006917, Induction of apoptosis	0.048584*
	GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	0.050408
	GO:0007187, G-protein signaling, coupled to cyclic nucleotide second messenger	0.289466
	GO:0007188, G-protein signaling, coupled to cAMP nucleotide second messenger	0.687572
	GO:0010942, Positive regulation of cell death	0.356228
	GO:0012502, Induction of programmed cell death	0.048584*
	GO:0019933, CAMP-mediated signaling	0.687572
	GO:0019935, Cyclic-nucleotide-mediated signaling	0.289466
	GO:0043067, Regulation of programmed cell death	0.694486
	GO:0043068, Positive regulation of programmed cell death	0.356228
MIMAT0000250(hsa-miR-139-5p)	GO:0048011, Nerve growth factor receptor signaling pathway	0.154543
	GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	0.940727
MIMAT0000265(hsa-miR-204-5p)	GO:0035583, Negative regulation of transforming growth factor beta receptor signaling pathway by extracellular sequestering of TGFbeta	0.940727
MIMAT0000423(hsa-miR-125b-5p)	GO:0002224, Toll-like receptor signaling pathway	0.380928
	GO:0002755, MyD88-dependent toll-like receptor signaling pathway	0.380928
	GO:0002756, MyD88-independent toll-like receptor signaling pathway	0.380928
	GO:0006916, Anti-apoptosis	0.0593
	GO:0006917, Induction of apoptosis	0.618064
	GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	0.776269
	GO:0008284, Positive regulation of cell proliferation	0.882324
	GO:0008285, Negative regulation of cell proliferation	0.883393
	GO:0008629, Induction of apoptosis by intracellular signals	0.073118
	GO:0010942, Positive regulation of cell death	0.972892
	GO:0012502, Induction of programmed cell death	0.618064
GO:0015629, Actin cytoskeleton	0.596528	

Table 9. GOBO survival analysis results. (cont.)

*: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

miRNA	GO term	P-value
MIMAT0000423(hsa-miR-125b-5p)	GO:0015630, Microtubule cytoskeleton	0.028245*
	GO:0030834, Regulation of actin filament depolymerization	0.654383
	GO:0030835, Negative regulation of actin filament depolymerization	0.654383
	GO:0030837, Negative regulation of actin filament polymerization	0.654383
	GO:0034130, Toll-like receptor 1 signaling pathway	0.380928
	GO:0034134, Toll-like receptor 2 signaling pathway	0.380928
	GO:0034138, Toll-like receptor 3 signaling pathway	0.380928
	GO:0034142, Toll-like receptor 4 signaling pathway	0.380928
	GO:0035666, TRIF-dependent toll-like receptor signaling pathway	0.380928
	GO:0043067, Regulation of programmed cell death	0.643418
	GO:0043068, Positive regulation of programmed cell death	0.972892
	GO:0043069, Negative regulation of programmed cell death	0.492576
	GO:0048011, Nerve growth factor receptor signaling pathway	0.171634
	GO:0050678, Regulation of epithelial cell proliferation	0.002205**
	GO:0050679, Positive regulation of epithelial cell proliferation	0.205483
	GO:0051693, Actin filament capping	0.654383
GO:0060548, Negative regulation of cell death	0.413665	
MIMAT0000437(hsa-miR-145-5p)	GO:0007178, Transmembrane receptor protein serine/threonine kinase signaling pathway	0.196953
	GO:0030509, BMP signaling pathway	0.196953
	GO:0090092, Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.843529
MIMAT0000443(hsa-miR-125a-5p)	GO:0001936, Regulation of endothelial cell proliferation	0.115146
	GO:0002221, Pattern recognition receptor signaling pathway	0.380928
	GO:0002224, Toll-like receptor signaling pathway	0.380928
	GO:0002755, MyD88-dependent toll-like receptor signaling pathway	0.380928
	GO:0002756, MyD88-independent toll-like receptor signaling pathway	0.380928
	GO:0006916, Anti-apoptosis	0.0593
	GO:0006917, Induction of apoptosis	0.618064
	GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	0.776269
	GO:0008284, Positive regulation of cell proliferation	0.882324
	GO:0008285, Negative regulation of cell proliferation	0.883393
	GO:0008629, Induction of apoptosis by intracellular signals	0.073118
	GO:0010942, Positive regulation of cell death	0.972892
	GO:0012502, Induction of programmed cell death	0.618064

Table 9. GOBO survival analysis results. (cont.)

*: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

miRNA	GO term	P-value
MIMAT0000443(hsa-miR-125a-5p)	GO:0015629, Actin cytoskeleton	0.596528
	GO:0015630, Microtubule cytoskeleton	0.028245*
	GO:0030834, Regulation of actin filament depolymerization	0.654383
	GO:0030835, Negative regulation of actin filament depolymerization	0.654383
	GO:0030837, Negative regulation of actin filament polymerization	0.654383
	GO:0034130, Toll-like receptor 1 signaling pathway	0.380928
	GO:0034134, Toll-like receptor 2 signaling pathway	0.380928
	GO:0034138, Toll-like receptor 3 signaling pathway	0.380928
	GO:0034142, Toll-like receptor 4 signaling pathway	0.380928
	GO:0035666, TRIF-dependent toll-like receptor signaling pathway	0.380928
	GO:0043067, Regulation of programmed cell death	0.643418
	GO:0043068, Positive regulation of programmed cell death	0.972892
	GO:0043069, Negative regulation of programmed cell death	0.492576
	GO:0048011, Nerve growth factor receptor signaling pathway	0.171634
	GO:0050678, Regulation of epithelial cell proliferation	0.002205**
	GO:0050679, Positive regulation of epithelial cell proliferation	0.205483
	GO:0050851, Antigen receptor-mediated signaling pathway	0.103325
	GO:0051693, Actin filament capping	0.654383
GO:0060548, Negative regulation of cell death	0.413665	
MIMAT0002820(hsa-miR-497-5p)	GO:0001959, Regulation of cytokine-mediated signaling pathway	0.06699
	GO:0006917, Induction of apoptosis	0.142401
	GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	0.837635
	GO:0007173, Epidermal growth factor receptor signaling pathway	0.387447
	GO:0007178, Transmembrane receptor protein serine/threonine kinase signaling pathway	0.240167
	GO:0007179, Transforming growth factor beta receptor signaling pathway	0.017876*
	GO:0008284, Positive regulation of cell proliferation	0.430255
	GO:0008285, Negative regulation of cell proliferation	0.149994
	GO:0008543, Fibroblast growth factor receptor signaling pathway	0.521978
	GO:0010942, Positive regulation of cell death	0.237692
	GO:0012502, Induction of programmed cell death	0.142401
	GO:0015629, Actin cytoskeleton	0.228804
	GO:0015630, Microtubule cytoskeleton	0.18331
	GO:0017015, Regulation of transforming growth factor beta receptor signaling pathway	0.128773

Table 9. GOBO survival analysis results. (cont.)

*: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

miRNA	GO term	P-value
MIMAT0002820(hsa-miR-497-5p)	GO:0030509, BMP signaling pathway	0.39837
	GO:0030521, Androgen receptor signaling pathway	0.383811
	GO:0032956, Regulation of actin cytoskeleton organization	0.91762
	GO:0042058, Regulation of epidermal growth factor receptor signaling pathway	0.934045
	GO:0042059, Negative regulation of epidermal growth factor receptor signaling pathway	0.789492
	GO:0043067, Regulation of programmed cell death	0.111544
	GO:0043068, Positive regulation of programmed cell death	0.237692
	GO:0043069, Negative regulation of programmed cell death	0.856892
	GO:0048011, Nerve growth factor receptor signaling pathway	0.471986
	GO:0060548, Negative regulation of cell death	0.667437
	GO:0070302, Regulation of stress-activated protein kinase signaling cascade	0.561032
	GO:0090092, Regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.182314
	GO:0006917, Induction of apoptosis	0.489781
	GO:0007169, Transmembrane receptor protein tyrosine kinase signaling pathway	0.171689
GO:0007173, Epidermal growth factor receptor signaling pathway	0.449696	
GO:0008284, Positive regulation of cell proliferation	0.05916	
GO:0008286, Insulin receptor signaling pathway	0.237933	
GO:0008543, Fibroblast growth factor receptor signaling pathway	0.159318	
GO:0008629, Induction of apoptosis by intracellular signals	0.502822	
GO:0010942, Positive regulation of cell death	0.076906	
GO:0012502, Induction of programmed cell death	0.489781	
MIMAT0002856(hsa-miR-520d-3p)	GO:0015630, Microtubule cytoskeleton	0.000292***
	GO:0042058, Regulation of epidermal growth factor receptor signaling pathway	0.762633
	GO:0042059, Negative regulation of epidermal growth factor receptor signaling pathway	0.826854
	GO:0043067, Regulation of programmed cell death	0.740092
	GO:0043068, Positive regulation of programmed cell death	0.076906
	GO:0043069, Negative regulation of programmed cell death	0.067499
	GO:0048011, Nerve growth factor receptor signaling pathway	0.014926*
	GO:0051988, Regulation of attachment of spindle microtubules to kinetochores	7.48E-06***
	GO:0060548, Negative regulation of cell death	0.213035

REFERENCE

1. Jemal A, Center MM, DeSantis C, Ward EM: **Global patterns of cancer incidence and mortality rates and trends.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:1893-1907.
2. Veronesi U, Boyle P, Goldhirsch A, Orecchia R, Viale G: **Breast cancer.** *Lancet* 2005, **365**:1727-1741.
3. Cho E, Spiegelman D, Hunter DJ, Chen WY, Stampfer MJ, Colditz GA, Willett WC: **Premenopausal fat intake and risk of breast cancer.** *J Natl Cancer Inst* 2003, **95**:1079-1085.
4. Leong AS, Zhuang Z: **The changing role of pathology in breast cancer diagnosis and treatment.** *Pathobiology* 2011, **78**:99-114.
5. (EBCTCG) EBCTCG: **Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.** *Lancet* 2005, **365**:1687-1717.
6. Colleoni M, Viale G, Zahrieh D, Pruneri G, Gentilini O, Veronesi P, Gelber RD, Curigliano G, Torrisi R, Luini A, et al: **Chemotherapy is more effective in patients with breast cancer not expressing steroid hormone receptors: a study of preoperative treatment.** *Clin Cancer Res* 2004, **10**:6622-6628.
7. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, et al: **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.** *J Clin Oncol* 2007, **25**:118-145.
8. Hudis CA: **Trastuzumab--mechanism of action and use in clinical practice.**

- N Engl J Med* 2007, **357**:39-51.
9. Huttenhofer A, Schattner P, Polacek N: **Non-coding RNAs: hope or hype?** *Trends Genet* 2005, **21**:289-297.
 10. Kim VN, Han J, Siomi MC: **Biogenesis of small RNAs in animals.** *Nat Rev Mol Cell Biol* 2009, **10**:126-139.
 11. Hutvagner G, Zamore PD: **A microRNA in a multiple-turnover RNAi enzyme complex.** *Science* 2002, **297**:2056-2060.
 12. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G: **miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs.** *Genes Dev* 2002, **16**:720-728.
 13. Nelson PT, Hatzigeorgiou AG, Mourelatos Z: **miRNP:mRNA association in polyribosomes in a human neuronal cell line.** *RNA* 2004, **10**:387-394.
 14. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T: **Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs.** *Mol Cell* 2004, **15**:185-197.
 15. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, et al: **Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.** *Proc Natl Acad Sci U S A* 2002, **99**:15524-15529.
 16. Zhang B, Pan X, Cobb GP, Anderson TA: **microRNAs as oncogenes and tumor suppressors.** *Dev Biol* 2007, **302**:1-12.
 17. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T: **MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer.** *Gastroenterology* 2007, **133**:647-658.
 18. Rask L, Balslev E, Jorgensen S, Eriksen J, Flyger H, Moller S, Hogdall E,

- Litman T, Nielsen BS: **High expression of miR-21 in tumor stroma correlates with increased cancer cell proliferation in human breast cancer.** *APMIS* 2011, **119**:663-673.
19. Zaman MS, Shahryari V, Deng G, Thamminana S, Saini S, Majid S, Chang I, Hirata H, Ueno K, Yamamura S, et al: **Up-regulation of microRNA-21 correlates with lower kidney cancer survival.** *PLoS One* 2012, **7**:e31060.
20. Nadiminty N, Tummala R, Lou W, Zhu Y, Shi XB, Zou JX, Chen H, Zhang J, Chen X, Luo J, et al: **MicroRNA let-7c Is Downregulated in Prostate Cancer and Suppresses Prostate Cancer Growth.** *PLoS One* 2012, **7**:e32832.
21. Wang XF, Shi ZM, Wang XR, Cao L, Wang YY, Zhang JX, Yin Y, Luo H, Kang CS, Liu N, et al: **MiR-181d acts as a tumor suppressor in glioma by targeting K-ras and Bcl-2.** *J Cancer Res Clin Oncol* 2012, **138**:573-584.
22. Zhang Y, Yan LX, Wu QN, Du ZM, Chen J, Liao DZ, Huang MY, Hou JH, Wu QL, Zeng MS, et al: **miR-125b is methylated and functions as a tumor suppressor by regulating the ETS1 proto-oncogene in human invasive breast cancer.** *Cancer Res* 2011, **71**:3552-3562.
23. Sachdeva M, Mo YY: **MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1.** *Cancer Res* 2010, **70**:378-387.
24. Guo X, Wu Y, Hartley RS: **MicroRNA-125a represses cell growth by targeting HuR in breast cancer.** *RNA Biol* 2009, **6**:575-583.
25. Jin H, Tuo W, Lian H, Liu Q, Zhu XQ, Gao H: **Strategies to identify microRNA targets: new advances.** *N Biotechnol* 2010, **27**:734-738.
26. Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92-105.
27. Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR,

- Blencowe BJ, Frey BJ, Morris QD: **Using expression profiling data to identify human microRNA targets.** *Nat Methods* 2007, **4**:1045-1049.
28. Liu B, Li J, Tsykin A, Liu L, Gaur AB, Goodall GJ: **Exploring complex miRNA-mRNA interactions with Bayesian networks by splitting-averaging strategy.** *BMC Bioinformatics* 2009, **10**:408.
29. Tseng CW, Lin CC, Chen CN, Huang HC, Juan HF: **Integrative network analysis reveals active microRNAs and their functions in gastric cancer.** *BMC Syst Biol* 2011, **5**:99.
30. Farazi TA, Horlings HM, Ten Hoeve JJ, Mihailovic A, Halfwerk H, Morozov P, Brown M, Hafner M, Reyal F, van Kouwenhove M, et al: **MicroRNA sequence and expression analysis in breast tumors by deep sequencing.** *Cancer Res* 2011, **71**:4443-4453.
31. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, Horvath S: **Strategies for aggregating gene expression data: the collapseRows R function.** *BMC Bioinformatics* 2011, **12**:322.
32. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39**:D152-157.
33. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP: **Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs.** *Nat Struct Mol Biol* 2011, **18**:1139-1146.
34. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.** *Cell* 2005, **120**:15-20.
35. Betel D, Wilson M, Gabow A, Marks DS, Sander C: **The microRNA.org**

- resource: targets and expression.** *Nucleic Acids Res* 2008, **36**:D149-153.
36. Lall S, Grun D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P, et al: **A genome-wide map of conserved microRNA targets in C. elegans.** *Curr Biol* 2006, **16**:460-471.
37. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**:495-500.
38. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98**:5116-5121.
39. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al: **Human Protein Reference Database--2009 update.** *Nucleic Acids Res* 2009, **37**:D767-772.
40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
41. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28**:495-501.
42. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency.** *Annals of Statistics* 2001, **29**:1165-1188.
43. Maere S, Heymans K, Kuiper M: **BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics* 2005, **21**:3448-3449.

44. Sing T, Sander O, Beerenwinkel N, Lengauer T: **ROCR: visualizing classifier performance in R.** *Bioinformatics* 2005, **21**:3940-3941.
45. Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982, **143**:29-36.
46. Ringner M, Fredlund E, Hakkinen J, Borg A, Staaf J: **GOBO: gene expression-based outcome for breast cancer online.** *PLoS One* 2011, **6**:e17911.
47. Derfoul A, Juan AH, Difilippantonio MJ, Palanisamy N, Ried T, Sartorelli V: **Decreased microRNA-214 levels in breast cancer cells coincides with increased cell proliferation, invasion and accumulation of the Polycomb Ezh2 methyltransferase.** *Carcinogenesis* 2011, **32**:1607-1614.
48. Heyn H, Engelmann M, Schreek S, Ahrens P, Lehmann U, Kreipe H, Schlegelberger B, Beger C: **MicroRNA miR-335 is crucial for the BRCA1 regulatory cascade in breast cancer development.** *Int J Cancer* 2011, **129**:2797-2806.
49. Sempere LF, Christensen M, Silahtaroglu A, Bak M, Heath CV, Schwartz G, Wells W, Kauppinen S, Cole CN: **Altered MicroRNA expression confined to specific epithelial cell subpopulations in breast cancer.** *Cancer Res* 2007, **67**:11612-11620.
50. Iyevleva AG, Kuligina E, Mitiushkina NV, Togo AV, Miki Y, Imyanitov EN: **High level of miR-21, miR-10b, and miR-31 expression in bilateral vs. unilateral breast carcinomas.** *Breast Cancer Res Treat* 2012, **131**:1049-1059.
51. Mathivanan S, Periaswamy B, Gandhi TK, Kandasamy K, Suresh S, Mohmood R, Ramachandra YL, Pandey A: **An evaluation of human protein-protein interaction data in the public domain.** *BMC Bioinformatics* 2006, **7 Suppl**

- 5:S19.
52. Halbert CH, Kumanyika S, Bowman M, Bellamy SL, Briggs V, Brown S, Bryant B, Delmoor E, Johnson JC, Purnell J, et al: **Participation rates and representativeness of African Americans recruited to a health promotion program.** *Health Educ Res* 2010, **25**:6-13.
 53. Nicolas Diaz-Chico B, German Rodriguez F, Gonzalez A, Ramirez R, Bilbao C, Cabrera de Leon A, Aguirre Jaime A, Chirino R, Navarro D, Diaz-Chico JC: **Androgens and androgen receptors in breast cancer.** *J Steroid Biochem Mol Biol* 2007, **105**:1-15.
 54. Sethi N, Kang Y: **Dysregulation of developmental pathways in bone metastasis.** *Bone* 2011, **48**:16-22.
 55. Liao SJ, Zhou YH, Yuan Y, Li D, Wu FH, Wang Q, Zhu JH, Yan B, Wei JJ, Zhang GM, Feng ZH: **Triggering of Toll-like receptor 4 on metastatic breast cancer cells promotes alphavbeta3-mediated adhesion and invasive migration.** *Breast Cancer Res Treat* 2011.
 56. Jiang P, Enomoto A, Takahashi M: **Cell biology of the movement of breast cancer cells: intracellular signalling and the actin cytoskeleton.** *Cancer Lett* 2009, **284**:122-130.