國　立　交　通　大　學

生物資訊研究所

博 士 論 文

分析微小核醣核酸在腫瘤細胞中的調控機制

Systematic Analysis of microRNA Regulations in Tumor Cells

研 究 生：徐唯哲

指導教授：黃憲達 博士

中華民國九十八年三月

分析微小核醣核酸在腫瘤細胞中的調控機制

Systematic Analysis of microRNA Regulations in Tumor Cells

研 究 生：徐唯哲　　Student ： Paul Wei-Che Hsu
指導教授：黃憲達 博士　　Advisor：Hsien-Da Huang

國 立 交 通 大 學
生 物 資 訊 研 究 所
博 士 論 文

A Thesis
Submitted to Institute of Bioinformatics
College of the Biological Science & Technology
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Ph.D.
in
Bioinformatics

July 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年三月

分析腫瘤細胞中微小核醣核酸的調控機制

學生:徐唯哲　　指導教授:黃憲達 博士

國立交通大學　生物資訊研究所

摘要

微小核醣核酸(microRNA/miRNA)是一段長度約為 22 核苷酸的非編碼的寡核醣核酸分子，它們主要功能是藉由抑制轉譯或是降解 mRNA 來降低基因的表現。近年來發現微小核醣核酸調控許多的基因與細胞功能有關，造成細胞凋亡，分化和發育。許多研究顯示出人類腫瘤形成(oncogenesis)與異常的微小核醣核酸表現量有關，但是大多數的調控機制仍有待發覺。在這篇研究中，我們分析了人類兩百多種在癌細胞中表現異常的微小核醣核酸，並測量其組織特異性, 定義出 20 種在腫瘤形成中扮演重要角色的微小核醣核酸。這些微小核醣核酸分別在八種不同的癌細胞中扮演致癌基因或抑癌基因，藉由大量表現或不表現來調控下游的基因達到致癌目的。從計算不同癌細胞的基因表現與序列分析中，我們預測哪些轉錄因子會調控微小核醣核酸的表達，以及這些微小核醣核酸的標靶基因是否參與腫瘤形成。我們將分析結果建構出八種癌症疾病中微小核醣核酸和基因之間的調控網路，希望可以作為癌症研究或臨床實驗的參考。

Systematic Analysis of miRNA Regulations in Tumor Cells

Student: Paul Wei-Che Hsu Advisor : Dr. Hsien-Da Huang

Institute of Bioinformatics, National Chiao Tung University

Abstract

MicroRNAs (miRNAs) are small non-coding RNA molecules of ~22 nt sequences that have an important role in the translational inhibition and degradation of mRNA to downregulate gene expression. Recent work supports miRNAs downregulate gene expression during various crucial cell processes such as apoptosis, differentiation and development. Recent studies have suggested that oncogenesis may be link with aberrant expression of miRNAs, but in most case it is still not clear what mechanism of miRNA leads to cancer formation. In this study, we analyzed more than 200 miRNAs which are aberrantly expressed in tumor cells, and the tissue specificity is tested by the miRNA expression among normal tissues. We identified 20 oncomirs which are up or down-expressed to regulate downstream genes and involved in oncogenesis of eight cancer types. We also predicted the transcription factor binding sites (TFBS) in miRNA promoter and identified miRNA targets which are tumour suppressors or oncogenes. Those analyses are integrated for miRNA regulatory network construction in eight cancers, and we hope our achievements can support cancer research and clinical trial.

致謝

我最幸運的是能在博士班生涯開始時，便投入 miRNA 的研究。雖然五年前在台灣研究 miRNA 的實驗室屈指可數，能得到相關研究的資訊是少之又少，但是在指導教授黃憲達博士帶領下，我們創造了奇蹟；在人類和病毒的 miRNA 研究方面不但在台灣佔有領先的地位，建構的資料庫也受到國際學術單位的重視，這要歸功於黃博士當初高瞻遠矚的規劃，在此要衷心感謝他這些年對我的教誨與提拔。

另外要感謝 miRNA 小組最早的成員勝達和立人，在研究當初不但培養出堅強的革命情感，彼此的默契與合作也創造出驚人的實力。感謝熙淵總是在我需要幫助時適時出現，永遠記得你為了幫我申請留法獎學金時揮汗陪我一起跑公文的感人畫面。感謝威霽和博凱在研究期間提供不少協助，即使是小事情也會熱心地幫助我。感謝宗夷和文綺在論文上的幫忙，讓我的論文充實不少。感謝 Alan 在基因調控方面的協助，並讓 wet lab 能夠繼續運轉下去。還要感謝實驗室所有同仁，願意承擔實驗室一切瑣事並盡力完成。

最後感謝鄒安平博士對我的研究和論文提供精闢的見解以及指教。由於 miRNA 研究日新月異，限於時間倉促，這篇論文仍有許多不完美之處，但是對於 miRNA 在癌症方面的研究著實提供了新的方向。

將這篇論文獻給和我一起掙扎奮鬥的太座碧蘭和兒女浩榮和浩華，感謝他們陪著我這段日子一起喜怒哀樂，並衷心希望能永遠如此幸福美滿。

This thesis is based on the following publications:

1. Tsai, W.C., **Hsu, P.W.**, Lai, T.C., Chau, G.Y., Lin, C.W., Chen, C.M., Lin, C.D., Liao, Y.L., Wang, J.L., Chau, Y.P. et al. (2008) MicroRNA-122, a tumor suppressor microRNA that regulates intrahepatic metastasis of hepatocellular carcinoma. Hepatology, 49, 1571-1582.

2. Hsu, S.D., C.H. Chu, A.P. Tsou, S.J. Chen, H.C. Chen, **P.W. Hsu,** Y.H. Wong, Y.H. Chen, G.H. Chen, and H.D. Huang (2008) "miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes" Nucleic Acids Research, Vol 36, D165-D169.

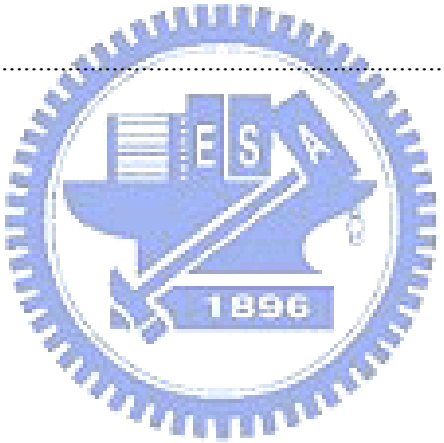
3. **Hsu, P.W.**, L.Z. Lin, S.D. Hsu, H.D. Huang (2007) "ViTa: a database of host microRNA targets on viruses" Nucleic Acids Research, Vol 35, D381-D385.

4. **Hsu, P.W.** , H.D. Huang, S.D. Hsu, L.Z. Lin, A.P. Tsou, C.P. Tseng, P.F. Stadler, S. Washietl, and I.L. Hofacker. (2006) "miRNAMap: genomic maps for miRNA genes and their target genes in vertebrate genomes" Nucleic Acids Research, Vol 34, D135-D139.

Table of Contents

List of Figures

List of Tables



Chapter 1 Introduction

The term “microRNA (miRNA)” was first introduced in 2001 (1), but in 1993, the miRNAs have been discovered in *C.elegan* and known the function of regulation in developmental stage (2). Since their discovery, miRNAs have been found in many organisms. In the latest version of miRBase (3), a database of miRNAs, 8619 miRNAs have been discovered in 87 different species. The numbers of human miRNAs are 695, more than tripled in the records of 2004 (Figure 1.1). In spite the fact that a large number of miRNAs have been identified, most of them are unknown the functions. Recently, miRNA expression has been linked to cancer, and to perform a comprehensive study of miRNA functions in oncogenesis. To systematically analyze miRNA regulations in tumor cells, we collected cancer-related miRNAs from literatures and integrated related information of miRNAs and gene expression profiles (**Chapter 2**). We also conducted the prediction of cis-regulatory elements of miRNA genes (**Chapter 3**) and miRNA targets identification (**Chapter 4**). We worked with the miRNAs which are directly associated with cancers and constructed the regulatory networks of oncomirs in 8 different cancer types (**Chapter 5)**. We hope to provide an effective analytical platform to determine miRNA regulatory systems that are contributors to oncogenesis and it might aid cancer diagnosis, including possibilities in tumor classification, disease prognosis, early cancer detection and therapeutic decision making.

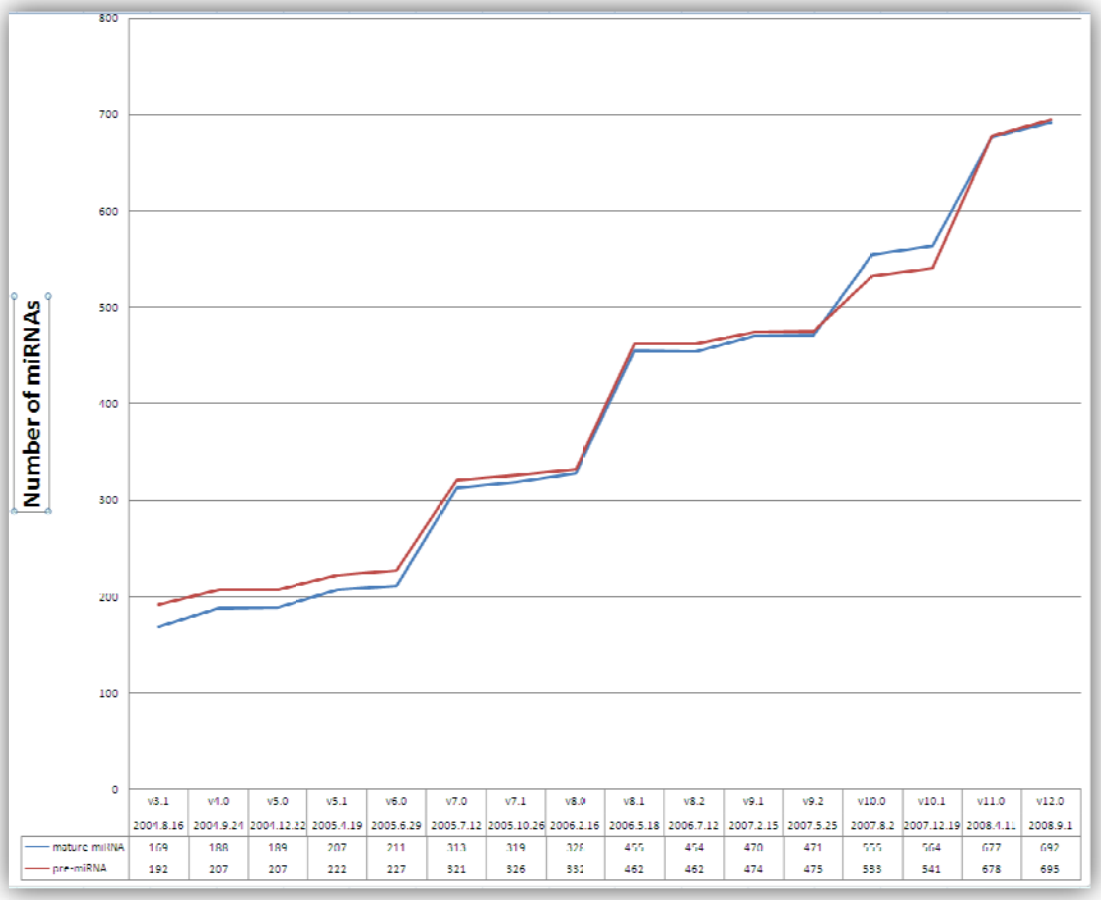


Figure 1.1 Numbers of human miRNAs have tripled since 2004.

1.1 Biological background

1.1.1 miRNA Biogenesis and Function

miRNAs are small non-coding RNAs of ~22 nt sequences that function to regulate gene expression by interfering the post-transcriptional level, resulting in degradation of mRNAs and repression of translation by the base pair to 3' untranslated regions (3'-UTR) of the mRNAs. Recent work supports miRNAs downregulate gene expression during various crucial cell processes such as apoptosis, differentiation and development.

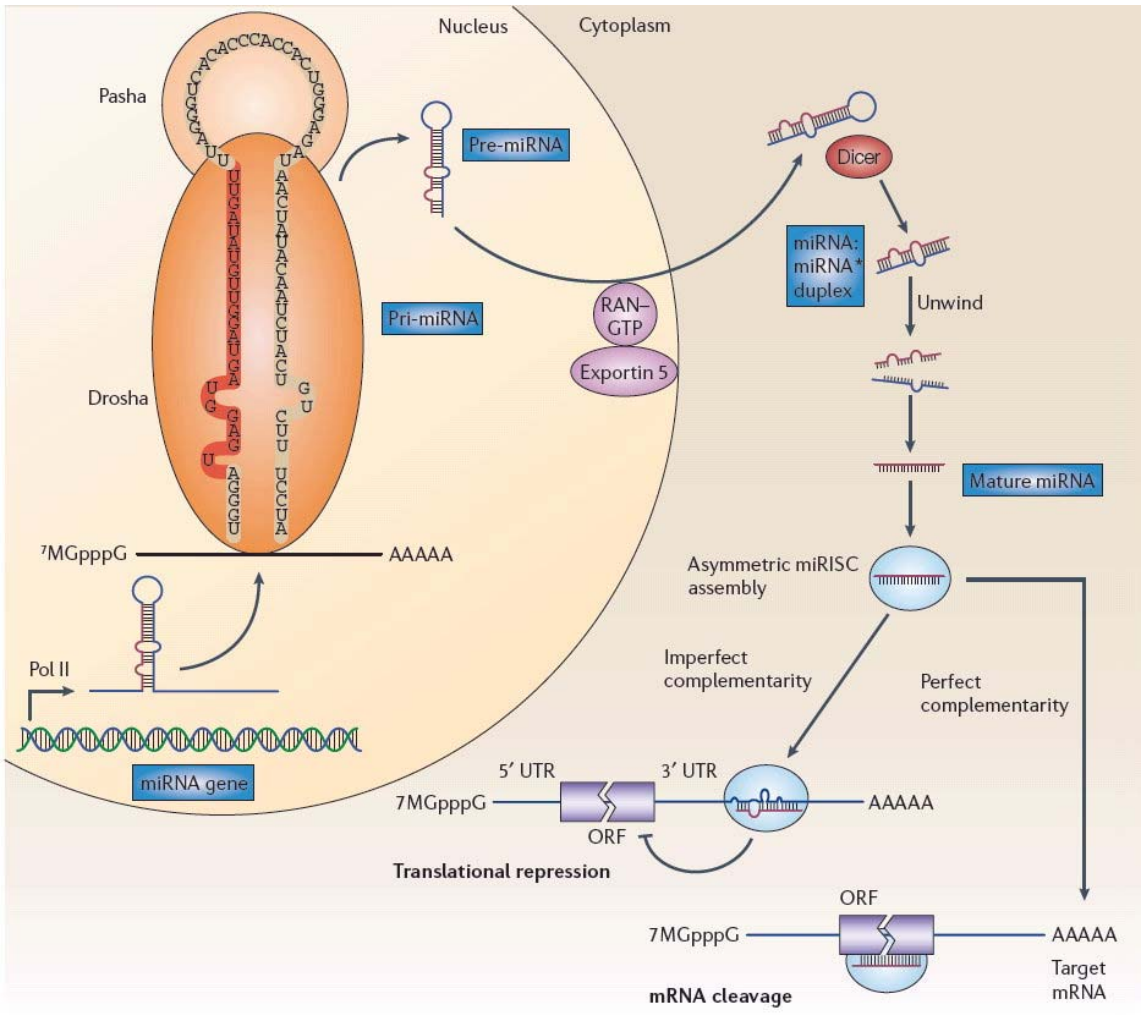


Figure 1.2 The biogenesis and function of miRNAs (Esquela-Kerscher and Slack, 2006).

MicroRNA (miRNA) genes are generally transcribed by RNA Polymerase II (Pol II) in the nucleus to form large pri-miRNA transcripts, which are with 5' 7-methyl guanosine cap and 3' poly-A tail (Figure 1.2). These pri-miRNA transcripts are processed by the RNase III enzyme Drosha and its co-factor DGCR8 (also known as Pasha) (4) to release about 70 nucleotide pre-miRNA precursor product. In this process, Drosha functions as the catalytic subunit, while DGCR8 recognizes the RNA substrate. This processing is crucial for the vast majority of miRNAs, although a small subgroup of miRNAs found in short introns can bypass this step (5). The pre-miRNA is then exported to the cytoplasm by exportin-5 and turned into 22-nt

miRNA duplex by RNase III enzyme Dicer (6) (Figure 1.2). One strand of the miRNA duplex is incorporated into an effector complex known as RNA-induced silencing complex (RISC) with Argonaute proteins (7). Recent study shows that nuclear RISC, consisting only of Ago2 and mature miRNA, is loaded in the cytoplasm and imported into the nucleus (8). Diederichs and Haber indicate that Ago2 can serve as Dicer enzyme to cleave pre-miRNA (9). The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression depend on the degree of complementarity between the miRNA and its target. miRNAs that bind to mRNA targets with imperfect complementarity block target gene expression at the level of protein translation. However, miRNAs can function as siRNA to affect mRNA stability (10). Complementary sites for miRNAs using this mechanism are generally found in the 3' untranslated regions (3'-UTRs) of the target mRNA genes. miRNAs that bind to their mRNA targets induce target-mRNA cleavage. miRNA-binding site such as seed regions (Watson–Crick consecutive base pairing between mRNAs and the miRNA at position 2–7 counted from its 5' end) (11) located in 3'-UTRs of mRNAs are important to translational repression and mRNA degradation (12). miRNAs using this mechanism bind to miRNA complementary sites that are generally found in the coding sequence or open reading frame (ORF) of the mRNA target.

Figure 1.3 shows the function of miRNAs in different ways. Plant miRNAs differ from animal miRNAs in that many plant miRNAs have perfect homology to their target mRNAs, and they act through the RNAi pathway to cause mRNA degradation (13). However, some plant with imperfect complementarity to their target sites and act a function similar to animal miRNAs. Plant miRNAs are also known to target chromatin modifications, such as histone methylation and DNA methylation (14).

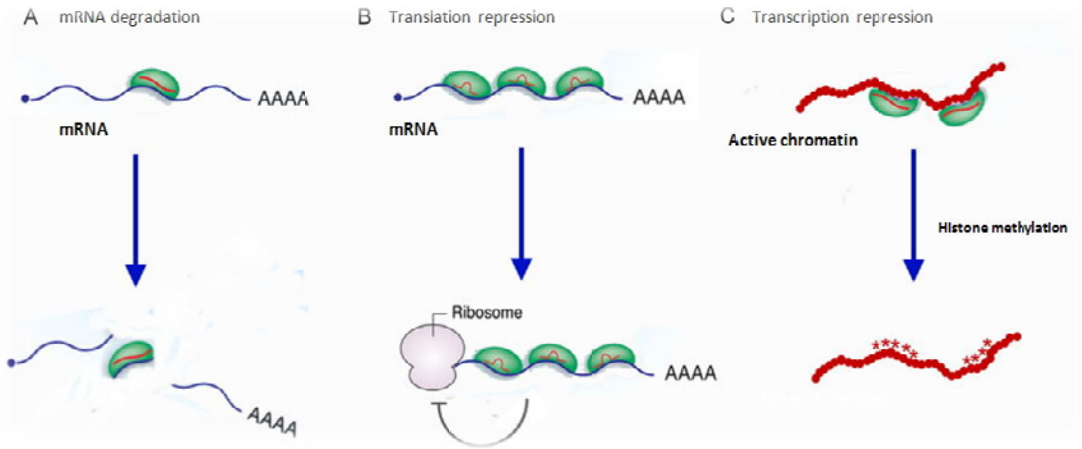


Figure 1.3 miRNA function: (A) mRNA degradation, common in plants, (B) Translation repression, common in animals, (C) Transcription repression by histone or DNA methylation, common in yeasts, plants.

1.1.2 Intronic and intergenic miRNAs

According to the miRNA locations of mRNA transcripts, five full-length pri-miRNAs have been characterized in Figure 1.4 (15), miR-155, miR-23a-27a-24-2 and miR-21 are intergenic miRNA which locate on non-protein-coding transcripts; whereas those miRNAs locate on the protein coding transcripts, like miR-26 and miR-198, are intronic miRNA and exonic miRNA.

Most human miRNAs are genomically isolated, several are found in miRNA clusters that are transcribed and expressed coordinately. Cai et al. analyzed the pri-miRNA precursors of 15 human miRNAs, five isolated and the others in clusters, have shown that all derive from pri-miRNA precursors that bear a 5' 7-methyl guanosine cap and a 3' poly-A tail (15). The pri-miRNA precursor for the clustered human miRNAs miR-23amiR-27amiR-24-2 is an unspliced 2.2 kilobase (kb) RNA, and the 3.4 kb pri-miRNA for the isolated human miRNA miR-21 is also unspliced. In contrast, the pri-miRNA for human miR-155 contains two introns and is polyadenylated at two alternate poly-A sites to give spliced pri-miRNA precursors of 0.6 and 1.4 kb.

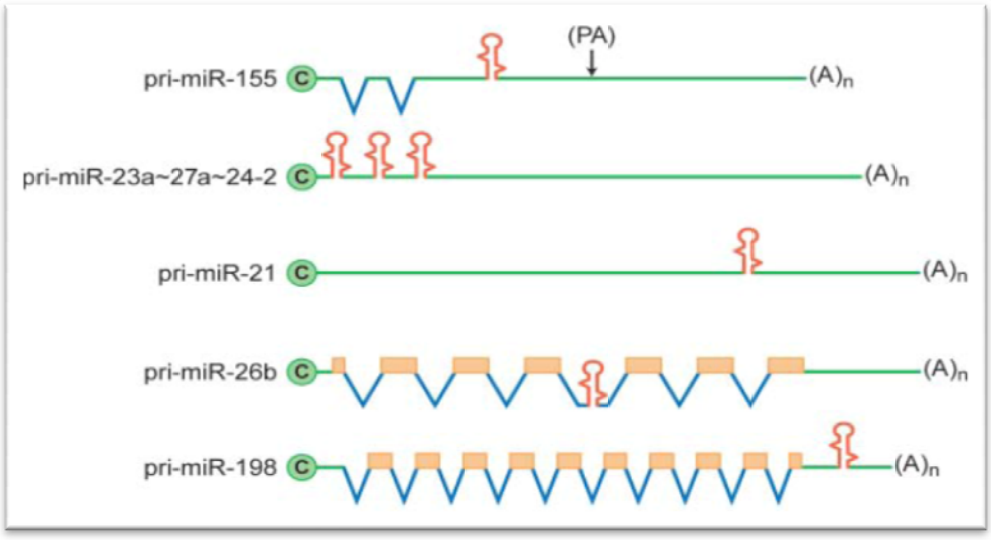


Figure 1.4 The structures of primary transcripts of microRNA (Cullen et al., 2004), PA, alternate poly-A-site.

The computational analysis of 695 human miRNAs, Identification of human miRNA host genes and transcription units (Figure 1.5), concluded that 51 were located in exon regions, 291 in intergenic regions of non-protein-coding transcripts, and 365 were located in intronic regions of protein coding transcripts. Twenty-four miRNAs were located in 3'-UTR or 5'-UTR regions. Twenty-five were found in both exonic and intronic locations, depending on the alternative splicing pattern of the flanking gene. Approximately 40% miRNAs are intergenic miRNAs and 52% miRNAs are located within the introns of protein coding mRNAs.

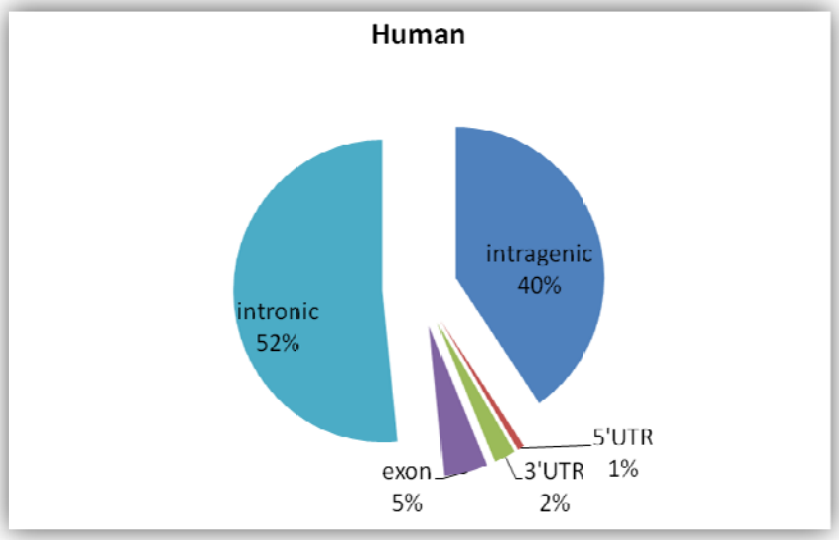


Figure 1.5 Approximately 40% miRNAs are expressed from non-protein-coding transcripts.

Some studies suggest that miRNAs are transcribed in parallel with their host transcripts (16), Baskerville and Bartel used oligonucleotide array to perform the expression patterns of 175 human miRNAs across 24 different human organs, a strong correlation expression profiles was observed in the expression profiles between miRNAs and their host genes, whereas less correlation was observed between miRNAs and their upstream or downstream genes. Table 1.3 shows that intronic miRNAs are usually coordinately expressed with their host gene mRNA, implying that they also generally derive from a common transcript (17).

Table 1.1 Correlation between expression of intronic miRNAs and host gene (Baskerville and Bartel, 2005).

Gene name	Ensembl ID[a]	MicroRNA	Host gene corr.	Upstream corr.	Downstream corr.	In situ
CROC4	125462	hsa-miR-9-1	0.999	0.083	−0.054	Ab (Jeffrey et al. 2000)
PDE2A	186642	hsa-miR-139	0.990	−0.293	0.493	Ab (Sadhu et al. 1999)
C20orf166	174407	hsa-miR-133a-2	0.988	−0.105	0.185	
		hsa-miR-1-1	0.968	−0.045	0.226	
PGSF1	176840	hsa-miR-7-3	0.961	−0.394	−0.123	
ABLIM2	163995	hsa-miR-95	0.960	−0.302	−0.089	
LOC254559	LOC254559	hsa-miR-9-3	0.950	−0.122	−0.150	
AATYK	181409	hsa-miR-338	0.921	0.906	−0.123	Ab (Baker et al. 2001; Tomomura et al. 2001)
EGFL7	172889	hsa-miR-126	0.888	0.727	−0.246	Ab, Nuc (Parker et al. 2004)
R3HDM	48991	hsa-miR-128a	0.856	−0.037	−0.363	
MCM7[b]	166508	hsa-miR-25	0.838	−0.548	−0.377	
TRPM3	83067	hsa-miR-204	0.796	−0.047	−0.240	
TLN2	171914	hsa-miR-190	0.663	−0.269	−0.172	
PANK3	120137	hsa-miR-103-1	0.638	−0.203	0.079	
PTPRN	54356	hsa-miR-153-1	0.626	0.579	−0.178	
CTDSP2	175215	hsa-miR-26a-2	0.609	0.492	0.265	
SMC4L1	113810	hsa-miR-15b	0.509	−0.134	−0.098	
		hsa-miR-16-2	0.504	−0.128	−0.021	
UREB1	86758	hsa-miR-98	0.503	−0.209	−0.512	
		hsa-let-7f-2	0.379	−0.242	−0.194	
PTPRN2	155093	hsa-miR-153-2	0.499	0.302	0.536	
CTDSP1	144579	hsa-miR-26b	0.453	0.292	−0.079	
ARPP-21	172995	hsa-miR-128b	0.444	0.622	−0.171	
MEST	106484	hsa-miR-335	0.442	0.129	−0.088	Nuc (Mayer et al. 2000)
NFYC	66136	hsa-miR-30c-1	0.406	−0.071	0.419	
Q8TDA7	174496	hsa-miR-99a	0.335	0.217	0.221	
		hsa-let-7c	0.067	0.712	−0.053	
		hsa-miR-125b-2	0.297	−0.037	0.221	
PANK2	125779	hsa-miR-103-2	0.270	−0.138	0.450	
PRO2730	164091	has-let-7g	0.268	0.516	−0.288	
RCL1	120158	hsa-miR-101-2	0.226	−0.494	0.196	
DNM1	79805	hsa-miR-199a-1	0.020	−0.124	−0.444	
C9orf5	106771	hsa-miR-32	−0.22	0.111	−0.290	
CTDSPL	144677	hsa-miR-26a-1	−0.285[c]	−0.202	0.099	

[a]Ensembl ID numbers begin with the prefix "ENSG00000," LOC254559 is a Genecards ID.
[b]A score of >16 in at least one tissue was required to be included in the analysis. *MCM7* contains two other miRNAs, miR-106b and miR-93, which were excluded based on this criterion although their correlation coefficients are consistent with coordinate expression.
[c]The anti-correlation between miR-26a-1 and *CTDSPL* can be explained by the *CTDSP2* transcript being the primary source of miR-26a. *CTDSP2*, the host gene of miR-26a-2, is generally expressed at much higher levels than is *CTDSPL*, and its expression is anti-correlated ($r = -0.106$) with that of *CTDSPL*. Similar scenarios could explain the low correlations between the expression of other miRNAs and their host genes.
Nuc, nucleic acid hybridization; Ab, antibody staining or immunofluorescence.

Chapter 1 Introduction

Liang et al. used TaqMan® MicroRNA Assays (a new type of real time reverse transcription (RT)-PCR-based miRNA assays) to analyze intronic miRNAs from a published literature based on the following rules: (1) the host gene and the miRNAs are transcribed from the same strand of DNA; (2) the host gene is a protein-coding gene with defined gene name and protein domains that link to its possible biological functions; (3) the miRNA does not have extra copies in other part of the genome since the transcription of each copy of the miRNA gene could be regulated by different mechanisms that would confound the result of our analyses. Among the 31 miRNAs qualified (Table 1.2), 22 of them had significant correlation ($p < 0.05$) with their host genes in expression among 19 tissue types.

Table 1.2 Correlation of expression patterns in human normal tissues between intronic miRNAs and their host genes (Liang et al., 2007).

miRNA	Host Gene	p value*	Gene Description
miR-106b	MCM7	0.047**	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)
miR-107	PANK1	0.022	pantothenate kinase 1
miR-126	EGFL7	0.027	EGF-like-domain, multiple 7
miR-128a	R3HDM1	0.002	R3H domain (binds single-stranded nucleic acids) containing 1
miR-128b	ARPP-21	0.001	cyclic AMP-regulated phosphoprotein, 21 kD
miR-139	PDE2A	0.007	phosphodiesterase 2A, cGMP-stimulated
miR-140	AIP2	**0.249**	WW domain containing E3 ubiquitin protein ligase 2
miR-148b	COPZ1	**0.472**	coatomer protein complex, subunit zeta 1
miR-149	GPC1	0.053	glypican 1
miR-151	PTK2	0.031	protein tyrosine kinase 2; focal adhesion kinase 1
miR-15b	SMC4L1	0.003	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)
miR-186	ZNF265	**0.335**	zinc finger protein 265
miR-188	CLCN5	0.07	chloride channel 5 (nephrolithiasis 2, X-linked, Dent disease)
miR-190	TLN2	0.001	talin 2
miR-196b	HOXA9	0.005	homeo box A9
miR-204	TRPM3	0.009	transient receptor potential cation channel, subfamily M, member 3
miR-208	MYH6	6×10^{-30}	myosin, heavy polypeptide 6, cardiac muscle, alpha
miR-211	TRPM1	0.004	transient receptor potential cation channel, subfamily M, member 1
miR-224	GABRE	0.006	gamma-aminobutyric acid (GABA) A receptor, epsilon
miR-25	MCM7	0.022	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)
miR-28	LPP	0.023	LIM domain containing preferred translocation partner in lipoma
miR-30e	NFYC	**0.618**	nuclear transcription factor Y, gamma
miR-326	ARRB1	0.013	arrestin, beta 1
miR-33	SREBF2	**0.287**	sterol regulatory element binding transcription factor 2
miR-335	MEST	0.006	mesoderm specific transcript homolog (mouse)
miR-338	AATK	0.0003	apoptosis-associated tyrosine kinase
miR-340	RNF130	**0.346**	ring finger protein 130
miR-342	EVL	0.017	Enah/Vasp-like
miR-346	GRID1	**0.895**	glutamate receptor, ionotropic, delta 1
miR-452	GABRE	0.0001	gamma-aminobutyric acid (GABA) A receptor, epsilon
miR-93	MCM7	0.022	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)

*When multiple clones are available in the database, the clone with the best p value was chosen. **Pearson correlation; bold-type numbers indicate those with p values > 0.07

1.1.3 **Roles of human miRNAs in cancers**

miRNAs play different and diverse roles in cancer. They can be involved in metastasis, invasion, proliferation, cell cycle, and apoptosis. Figure 1.6 illustrates that a miRNA is downregulated in cancer and targets an oncogene might act as a tumor suppressor, whereas an upregulated miRNA that targets a tumor suppressor (TS) or a gene important for differentiation might act as an oncogene (OG). Several groups have studied the miRNA expression in cancer patients and found that miRNAs are differentially expressed in normal and tumor tissues, these differences are

tumor-specific and, in some cases, are associated with prognosis (18).

The first evidence that miRNAs acting as tumour suppressor function is provided by let-7a which down-regulates expression of Ras oncogenes (19). Interestingly, let-7a family members map to fragile sites associated with lung, breast, urothelial and cervical cancer (20). Let-7a expression is reduced in tumours, potentially contributing to increase activity of the Ras pathway and effects on growth control (21). Furthermore, overexpression of let-7a in a human adenocarcinoma cell line inhibits cellular proliferation, indicating that this might be a potential therapeutic approach to treat lung cancer (22).

A variety of platforms have recently been developed for miRNA expression analyses. Those platforms made possible large profiling studies in cancer patients, confirming that miRNAs are differentially expressed in normal and tumor samples, Higher throughput expression approaches can in general be classified as hybridization-based methods using microarrays, or cloning and sequencing approaches (including miRAGE). While the latter have the advantage of being open systems that permit identification of novel miRNAs and accurate miRNA quantitation, the former are cost effective and more amenable to a large number of routine analyses. Examination of individual miRNAs can be performed by Northern hybridization and specialized real-time PCR, and can be assessed in cellular contexts through *in situ* hybridization.

Although expression profiling studies in cancer suggest that miRNAs might function as oncogenes and tumor suppressors, definitive evidence linking miRNAs with the development of cancer is scarce. Table 1.3.1 lists several studies which have investigated the roles of specific miRNAs in some cellular events.

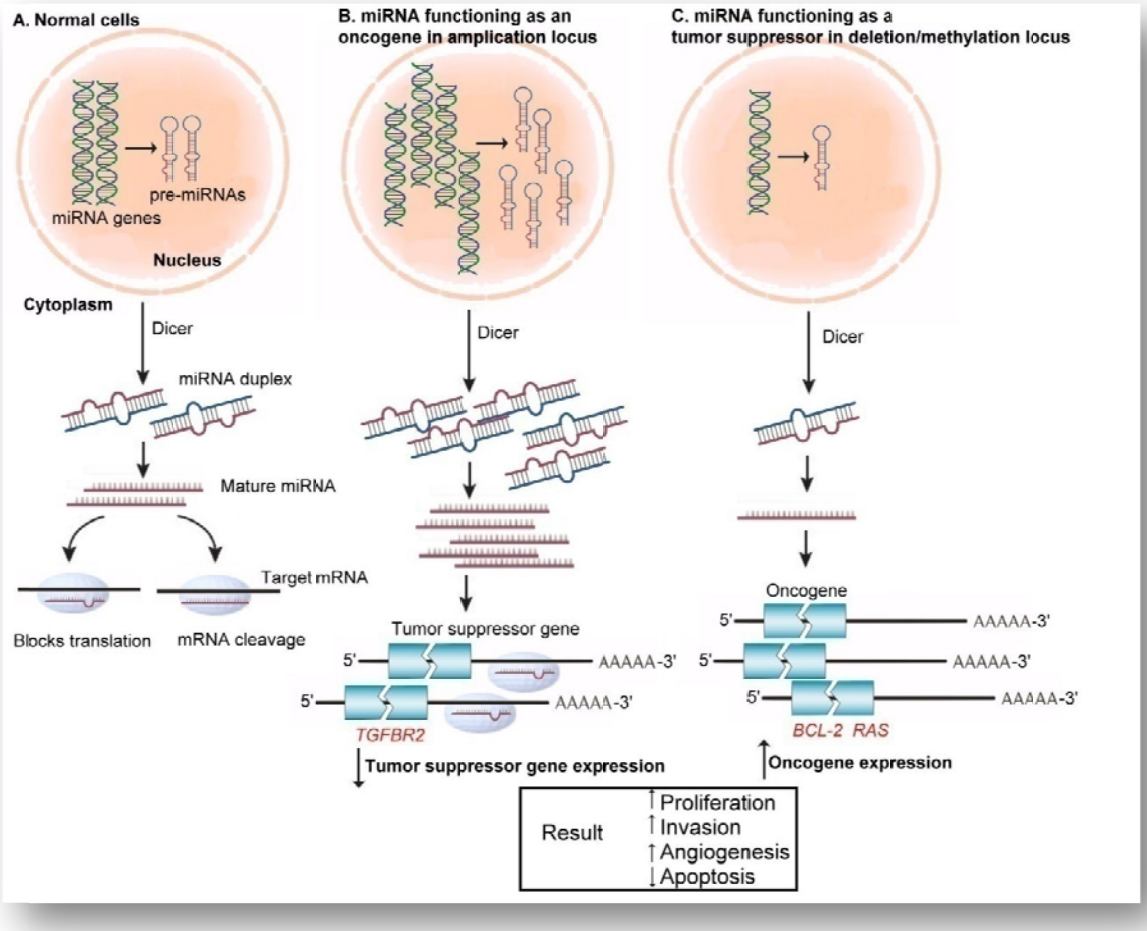


Figure 1.6 miRNAs can function as tumour suppressors and oncogenes. **A**. In normal cells, proper miRNAs result in the normal rate of growth, proliferation, differentiation and cell death. **B**. The amplification or overexpression of a miRNA that has an oncogenetic role would eliminate the expression of tumour suppressor genes and lead to cancer progression. **C**. The reduction or deletion of a miRNA that functions as a tumour suppressor leads to tumour formation.

Table 1.3 Experimental data supporting a role for miRNAs in cancer development.

miRNA	Expression in patients	Function	TS/OG	Refs
miR-122	Down regulated in liver cancer	Regulate metastasis gene ADAM17	TS	
miR-335 miR-126	Down regulated in breast cancer	Regulate a set of metastasis genes includes SOX4	TS	(23)
miR-373	Up regulated in breast cancer	Promote tumour invasion and metastasis by suppression of CD44	OG	(24)
miR-15a miR-16–1	Down regulated in CLL and in pituitary adenoma	Downregulate BCL-2	TS	(25-27)
let-7a-2	Down regulated in lung cancer	Downregulates RAS and induces apoptosis in lung cancer cell lines	TS	(19,21, 22)
miR-155	Up regulated in bad prognosis CLL and lung cancer, breast cancer, lymphoma, hodgkin and pediatric BL	Induces pre-B lymphoma and/or leukemia in mice	OG	(21,28)
miR-17–92 cluster	Up regulated in lymphomas and lung cancer	Cooperates with c-MYC; modulates E2F1	OG	(21)
miR-21	Up regulated in pancreas, glioblastoma and breast cancer	Anti-apoptotic in glioblastoma	OG	(28)
miR-106a	Up regulated in lung, gastric and prostate cancer	Downregulates RB-1	OG	(29)
miR-372, miR-373	Up regulated in testicular germ-cell tumor cell lines	Neutralize p53 function	OG	(30)
miR-142	c-MYC is translocated downstream of the miR-142 hairpin, resulting in B-lymphoma	Enhances MYC expression	OG	(31)

1.1.4 Experimental approaches for studying the function of miRNAs in cancer

Almost all of the miRNA-related studies on cancers are based on the different expression profile of miRNAs comparison in cancer cells and normal cells. Detecting mRNA expression can also be used in studies on the potential roles of miRNAs in cancers. Several experimental approaches are frequently used for cancer study:

Knockout or overexpression of miRNAs

To regulate the expression of the candidate miRNAs is a good approach to study the function of miRNAs in cancer pathogenesis. Knockdown or overexpression of a specific miRNA can study the specific roles of the miRNA in cancer initiation and development.

There are several methods to conduct this study, such as antisense inhibitors, transgenics, and point mutants. Using antisense inhibitors to block the targeted miRNA function is a good example. In this strategy, an artificial antisense RNA competes with cellular mRNAs to bind miRNAs. The antisense RNA pairs with the miRNA and inhibits the miRNA function. This has been adopted by two independent research groups to sequence-specifically inhibit miRNA- and siRNA-induced RNA silencing (32,33), and inhibit four miRNAs *in vivo* by modified antisense RNAs(34).

Point mutants of miRNAs or their targets can also be employed to study the function of miRNAs in cancers. One obvious advantage of point mutants is to study the direct interaction of miRNAs and their targeted genes. Several studies have shown that the "seed region" is important for miRNAs to recognize their targets, and increasing the mismatch in the seed sequences will significantly decrease the gene regulation function of miRNAs (35). The rules of miRNA target can be used to design point mutants of miRNAs or their targets. One or two nucleotide changes in the "seed region" of a specific candidate miRNA will dramatically decrease the possibility of the miRNA binding to its targets, resulting in the overexpression of the targets of the studied miRNAs. If these miRNAs or their targets are involved in cancer formation, this point mutation will affect the formation of cancer.

Northern blot analysis

Northern blot analysis is a reliable technique to detect gene expression at the mRNA level; it is widely used in gene expression analysis. Early on, it was adopted to study the expression of miRNA genes (2,36), and now is used as a method for detecting miRNA expression in cancer cells. For example, Hayashita et al. found that the miR-17–92 cluster is significantly overexpressed in lung cancer (37), especially with small-cell lung cancer, when compared with miRNA expression in normal cells.

Real-time PCR

Real-time PCR can also be employed to quantify miRNA expression profiles and study the potential function of miRNAs in cancer pathogenesis. Real-time PCR was first employed to measure miRNA precursors and to study the expression of 23 miRNA precursors in six cell lines (38). More recently, the real-time PCR assay was expanded to 222 miRNA precursor analysis in human cancer cell lines; different expression profiles of miRNA precursors in human cancers do exist (39). These PCR-based analyses quantify miRNA precursors and not the active mature miRNAs. The relationship between pri-miRNA and mature miRNA expression has not been thoroughly addressed. Applied Biosystems Inc. developed TaqMan® miRNA assays, a new real-time quantification method, using looped-primer RT-PCR to accurately detect mature miRNAs. Liang et al. applied the new approach to provide expression data of 345 miRNAs in 40 normal human tissues, which identified universally expressed miRNAs, and several groups of miRNAs expressed exclusively or preferentially in certain tissue types (40).

microRNA microarray

Although northern blotting is a widely used method for miRNA analysis, it has some limitations, such as unequal hybridization efficiency of individual probes (41), and difficulty in detecting multiple miRNAs simultaneously. For cancer studies, it is important to be able to compare the expression pattern of all known miRNAs between cancer cells and normal cells. Thus, it is better to have methods which detect all miRNA expression at a single time. Two-color fluorescence-based microarray technology (DNA microarray) has been widely used to detect gene expression simultaneously. Several laboratories have modified DNA microarray technology to form miRNA microarray technology (41-46) developed a custom dual-channel miRNA microarray platform, and

employed it to monitor expression levels of 124 mammalian miRNAs. They observed that the expression patterns of miRNAs are different between adult mouse tissue and embryonic stem cells (41). More recently,(47) developed a novel strategy to detect abundant miRNA expression profiles in different cell types, including several human cancers. To overcome the concerns about probe specificity in miRNA microarray analysis, they first performed hybridization in solution, and then quantified the polymer heads that are hybridized to miRNA species using multicolor flow-sorting (47). This method can be used to detect miRNA expression profiles in cancers, and even in poorly differentiated tumors (47)

Table 1.4 lists the strengths and weaknesses of these principle approaches for studying the functions of miRNAs in cancer pathogenesis (48). Although Northern blot analysis, real-time PCR, and miRNA microarray can detect the expression of certain miRNAs and determine which miRNAs may be associated with cancer formation, it is difficult to determine whether or not miRNAs play a unique role in cancers. Also, these techniques cannot directly determine the correlation between mRNA expression levels and whether the up-regulation or down-regulation of certain miRNAs is the cause of cancer or a downstream effect of the disease. Those problems can be solved by using antisense inhibitors, transgenics and point mutants. However, the techniques for regulation of specific miRNAs are still under development, and it has a long way to go before being used for clinical purposes (48). To better understand the functions of miRNAs in cancers, combining the bioinformatics strategy with the strengths of current techniques needs to be developed.

Table 1.4 Comparison of principle techniques for studying the functions of miRNAs in cancer pathogenesis (48).

Technique	Strengths	Weaknesses
Antisense inhibitor	Inhibiting specific miRNA expression to study miRNA functions in cancers.	Need to design and transform specific antisense inhibitor into targeted cells.
Transgenics	Regulating specific miRNA expression to study miRNA functions in cancers.	Need to obtain transgenic cell line to study specific miRNA functions.
Point mutant	Directly affecting the miRNA binding to targeted mRNAs, and studying the interaction of miRNAs and their targeted cancer-related genes.	Complicated design process. Need to obtain transgenic cell lines before study miRNA functions.
Northern blot	The most reliable techniques to study the expression of miRNAs in cancers.	No direct correlation between mRNA expression levels and whether the up-regulation or down-regulation of certain miRNAs is the cause of cancer or a downstream effect of the disease.
Real-time PCR	Rapidly detect miRNA expression, especially pri-miRNA expression.	No direct correlation between mRNA expression levels and whether the up-regulation or down-regulation of certain miRNAs is the cause of cancer or a downstream effect of the disease.
miRNA microarray	Simultaneously detect the expression of multiple miRNAs in cancers, may become a technique in cancer epidemiology and early cancer detection.	No direct correlation between mRNA expression levels and whether the up-regulation or down-regulation of certain miRNAs is the cause of cancer or a downstream effect of the disease.

1.2 Motivation

Although hundreds of human miRNA genes have been discovered, the functions of only a handful of these miRNAs have been experimentally determined. As the miRNA field advances, more and more researchers have identified aberrant expression of miRNAs in cancers including lymphomas, colorectal carcinoma, breast cancer, lung cancer, thyroid cancer, and hepatocellular carcinomas. It is crucial to find out what kind of regulatory mechanism results the aberrant expression of miRNAs in cancers and what kind of genes are targeted by miRNAs and results tumor formation. Therefore, a systematic analysis of miRNA function is essential to comprehensive study their roles in

cancer. Using bioinformatic tools to combine experiment data can predict the diverse roles of miRNAs which might be involved in metastasis, invasion, proliferation, cell cycle, and apoptosis. In this study, we try to provide an efficient method to systematic analysis the regulatory relationship between cancer-related miRNAs and coding genes which include transcription factors, oncogenes and tumor suppressor genes. We hope to support cancer study to forward understanding miRNA role in tumor pathogenesis.

1.3 Research goals

The goal of this study was to develop a computational approach to comprehensive analysis of the transcription factors of cancer-related miRNA and miRNA targets that are most likely to be the major biological processes in cancer pathogenesis. Those cancer-related miRNAs are analyzed by tissue specific test and the following items: transcriptional start site (TSS) prediction, cis-regulatory elements prediction, miRNA target prediction and target gene annotation (Figure 1.7). The oncomir regulatory networks involved in different oncogenesis pathway were identified after those analysis accomplished.

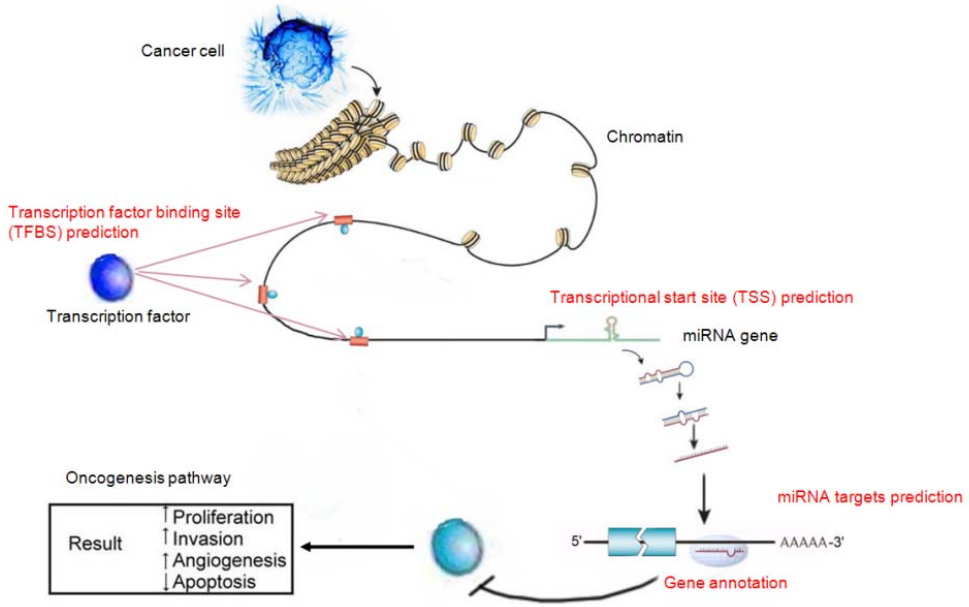


Figure 1.7 The research in this dissertation contains TSS and cis-regulatory elements prediction, miRNA targets prediction and gene annotation.

Chapter 2 Construction of miRNA information repository

2.1 Introduction

Although the first published research of miRNAs appeared fifty years ago (2), study in small RNA regulatory has been appreciated only in the last five years and becomes one of the most important events in genome-wide studies. It attributes the success to the discovery of RNA interference (RNAi), which has revolutionized approaches to down-regulate gene function. To understand the endogenous RNA interference in gene regulatory system, more and more miRNAs have been identified, and several groups have studied the miRNA expression in cancer patients by microarray, bead-based hybridization or real-time PCR and found that miRNAs are differentially expressed in normal and tumor tissues.

To comprehensive study in cancer-related miRNA function, we developed useful tool, named miRNAMap (49,50) and collected related information from miRbase database (3) and literatures in PubMed Central. After accomplishing data collection, cancer-related miRNA will be further analyzed in next two chapters.

2.2 Related works

There are several existing resources that provide updated data regarding each of these areas of research. miRBase is the main database of experimentally validated mature miRNA sequences. The miRBase database also provides integrated interfaces to comprehensive miRNA annotation and predicted gene targets. ARGONAUTE (51) provides a larger miRNA tissue expression dataset — collected from various miRNA

expression studies. DIANA TarBase (52,53) collects experimentally validated miRNA targets in eight species.

miRBase

The miRBase (http://microrna.sanger.ac.uk/) is a main database for miRNA study. It contains sequences of all experimentally verified mature miRNAs, annotated with primary literature references and the experimental method used for discovery, together with their predicted hairpin precursors, structure and function. It also integrated interfaces to comprehensive microRNA data and predicted gene targets.

In the current version 12.0, released in September 2008, there are 8619 miRNAs have been discovered in 87 different species, comparing with last version 11.0 (released in April 2008), 2227 new hairpin sequences and 2413 novel mature miRNAs have been added in the current version. This shows that the database grows very rapidly.

ARGONAUTE

Argonaute (http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface) is a database on mammalian microRNAs and their function in gene and pathway regulation. In contrast to miRBase, it has information on (i) origin of a miRNA, i.e. in which host gene it is encoded, (ii) its expression in different tissues and its known or proposed function, (iii) its potential target genes including Gene Ontology annotation as well as (iv) on miRNA families and proteins known to be involved in miRNA processing. The first release of Argonaute contains 839 miRNAs from human, mouse and rat.

TarBase

TarBase (http://microrna.gr/tarbase) is a database which collects experimentally supported miRNA targets in eight species of animals, plants and viruses. Even though

several computational programs exist to predict miRNA targets, there is a need for a comprehensive collection and description of miRNA targets with experimental support. The current version Tarbase 5.0, released in October 2008, includes more than 1300 experimentally supported targets. Each target site is described by the miRNA that binds it, the gene in which it occurs, the nature of the experiments that were conducted to test it, the sufficiency of the site to induce translational repression and/or cleavage, and the paper from which all these data were extracted. Additionally, the database is functionally linked to several other relevant and useful databases such as Ensembl, Hugo, UCSC and SwissProt.

2.3 The specific aim of data integration of miRNAs

To facilitate the annotation of the miRNA function, it is obliged to integrate relative information. We developed an integrated database, namely miRNAMap (http://mirnamap.mbc.nctu.edu.tw/) (49), to compile the miRNA genes, miRNA targets and the regulatory relationships between the miRNAs and the miRNA target genes. miRNAMap was published in 2006, is also one of the main databases to comprehensive study miRNA. It provides a variety of search functions and graphical interface to facilitate the researchers who interested in the miRNA riles in cell regulations (Figure 2.1).

The initial version of miRNAMap includes two portions: miRNAs and miRNA targets, and focuses on four species, human, mouse, rat and dog. miRNAMap collects published miRNAs from the microRBase and predict putative miRNA precursors (Figure 2.2) by RNAz (49) which is based on genome-wide mapping of conserved RNA secondary structures from UCSC clustering (54), and also predicts the mature miRNAs by an algorithm which is based on MDD (maximal dependence decomposition), all of

the known mature miRNAs have been detected their targets by miRanda (55) and TargetScans (35) results.

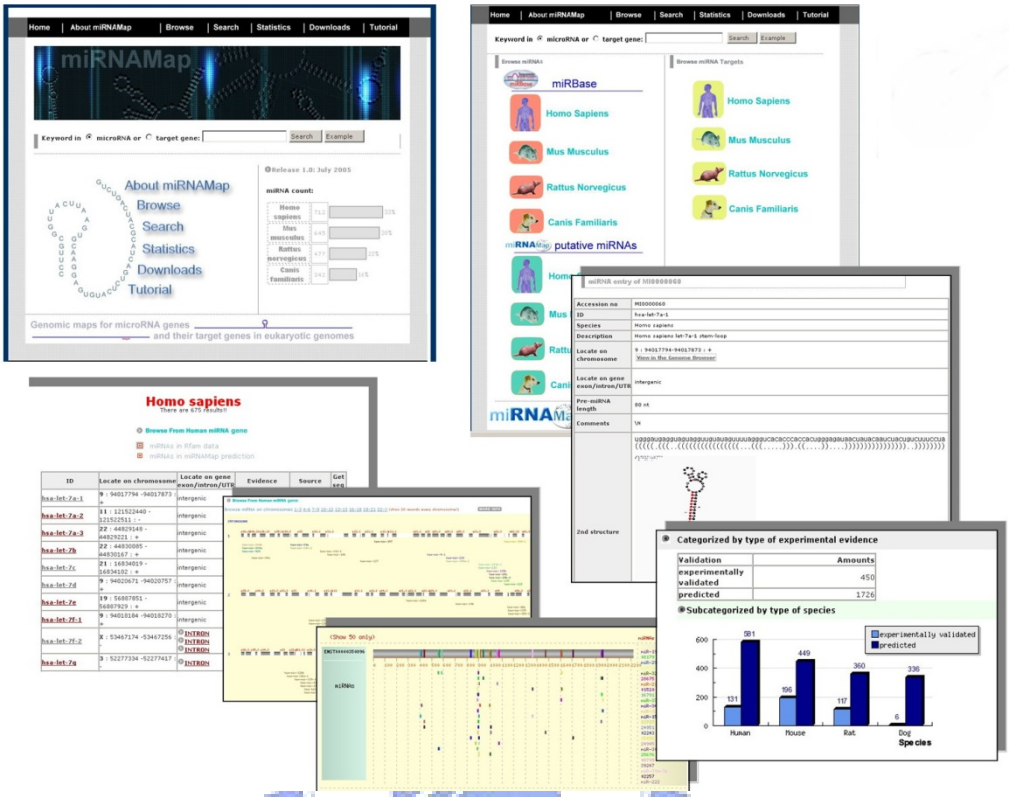


Figure 2.1 miRNAMap provides a variety of search functions and graphical interface.

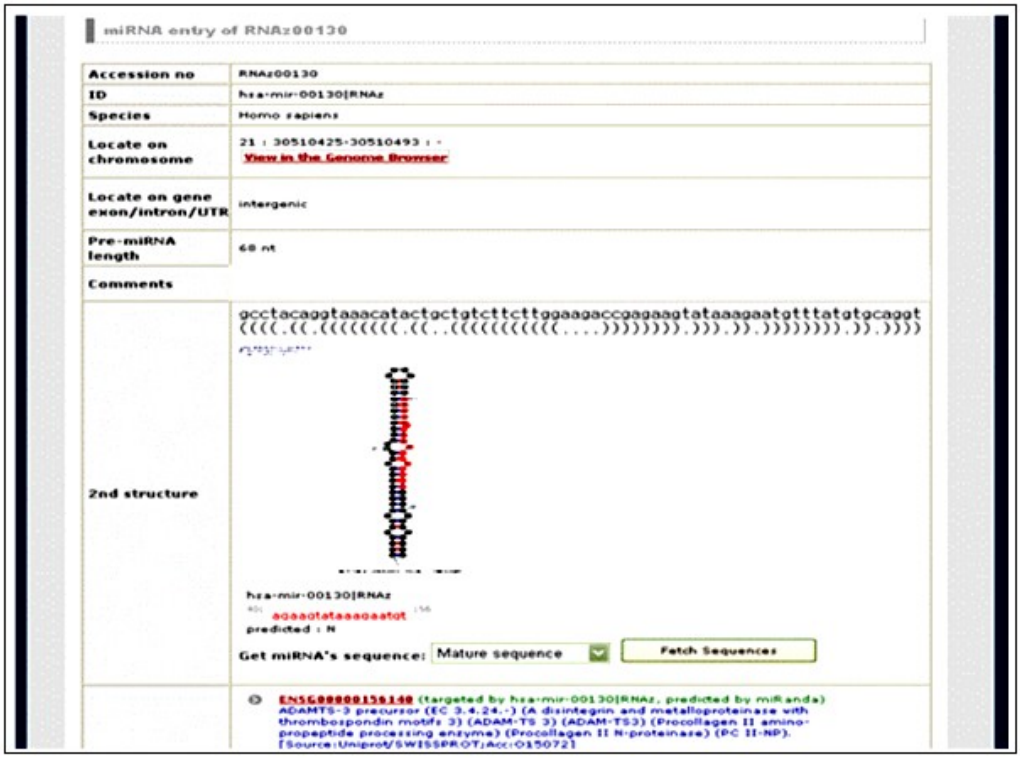


Figure 2.2 Putative miRNAs predicted by miRNAMap.

miRNAMap 2.0 contains both the collection of experimentally verified miRNA targets and computationally predicted miRNA binding sites in metazoan genomes. These data should aid researchers in the exploration of miRNAs biological function and the interpretation of the results of experiments. Table 2.1 shows the difference between previous version and miRNAMap 2.0. The main improvements incorporated to the database in the past year which include incorporating information on the accessibility of target genes predicted by Sfold (56) and the additional expression profile of miRNA and their target genes in expanded of 4 —increasing to 12 — species (Table 2.2). The assessment of the RNA accessible regions of target genes is supported and it provides the different viewpoint for the study of miRNA/target relationship. We also update the known miRNAs and annotate these miRNA targets according to combinations of widely used target prediction programs. Finally, both textual and graphical web interface are designed to facilitate the retrieval of data from the miRNAMap (Figure 2.3).

Figure 2.3 Graphical web interface of miRNAMap 2.0.

Table 2.1 The comparison of data and function between miRNAMap 1.0 and miRNAMap 2.0.

Features	miRNAMap 1.0	miRNAMap 2.0
Known miRNA entry	miRBase (version 6.0)	miRBase (version 9.2)
Species	4 mammalian	2 insects, 8 vertebrates, 2 worms
Experimental miRNA targets resource	Literature	TarBase and literature
miRNA gene expression profile	MIT microRNA profiling	MIT microRNA profiling, Q-PCR microRNA profiling
Target gene expression profile		NCBI GEO
Integrated miRNA target prediction tool	miRanda	miRanda, RNAhybrid, TargetScan
miRNA target accessible region prediction		Sfold
Relationship between miRNA and target genes		Analysis of the expression profiles between miRNA gene and its target genes.
miRNA tissue specificity	Text description	Analysis of the miRNA expression level in each tissue.
Graphical visualization	1. Pre-miRNA secondary structure 2. miRNA target sites 3. gene group search 4. literature surveyed miRNA targets 5. miRNA located information.	1. Pre-miRNA secondary structure 2. miRNA target sites 3. gene group search 4. literature and Tarbase experimental miRNA targets 5. the relationship between miRNA and its target 6. miRNA target accessible region.

Table 2.2 Numbers of mature miRNAs categorized by type of species in miRNAMap.

Species	Abbr.	Numbers
human (homo sapiens)	Hsa	542
mouse (mus musculus)	Mmu	424
zebrafish (danio rerio)	Dre	371
rat (rattus norvegicus)	Rno	261
frog (xenopus tropicalis)	Xtr	196
chicken (gallus gallus)	Gga	162
worm (caenorhabditis elegans)	Cel	135
pufferfish (fugu rubripes)	Fru	133
opossum (monodelphis domestica)	Mdo	111
fly (drosophila melanogaster)	Dme	85
mosquito (anopheles gambiae)	Aga	38
dog (canis familiaris)	Cfa	6

2.4 Materials and methods

We collected those cancer-related miRNAs which differentially expressed between normal and tumor tissues from literature. miRNA-related information was obtained from miRNAMap which is a powerful tool to provide detail information to annotate those miRNAs and target genes for further analysis with cancer-related miRNAs.

In this work we used many gene expression data from Gene Expression Omnibus (GEO) (57) which is a public database that archives and freely distributes microarray and other forms of high-throughput data submitted by the scientific community. A human genes expression profile in 79 normal tissues (NCBI GEO DataSet GDS596) is integrated in our research to analysis the regulatory relationship between miRNAs and coding genes (See Chapter 3). Several tumor microarray data were used to identify certain cis-regulatory elements of miRNAs and miRNA target genes which are involved in oncogenesis (See Chapter 3 and Chapter 4). miRNA expression profiles are also

important in our study, the tissue specific miRNAs were identified by using miRNA expression profiles among different normal tissues (40) (See Chapter 5). Figure 2.4 shows that the integration of miRNA data and related gene expression data in this study.

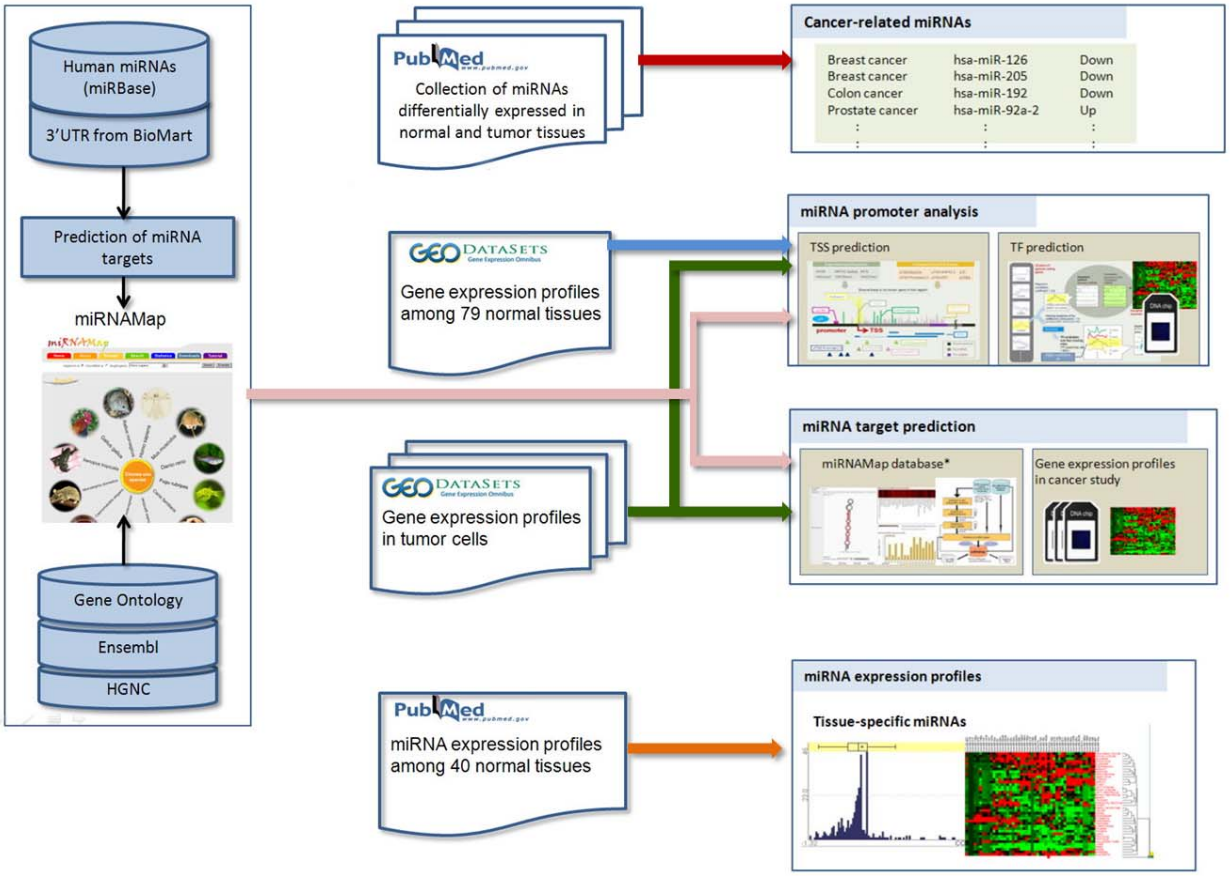


Figure 2.4 Integration of miRNA data and related gene expression data in this study.

2.4.1 Collection of cancer-related miRNAs from literatures

The first report of the association between miRNAs and cancer was published in 2002, by Calin et al. (25). Since their discovery, researchers have identified more and more aberrant expression of miRNAs in other malignancies. For efficiently searching related references, PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) was screened for papers describing cancer-related miRNAs by using the general key words like “cancer and miRNA”, we also used more than twenty cancer types as keyword like "breast cancer and miRNA", " B-cell lymphoma and miRNA". All of the cancer-related miRNAs are marked “Up” or “Down” which infer upregulated mRNA expression or downegulated mRNA expression in cancer.

2.4.2 Integration of miRNAMap database

In this study, we focused on human cancer-related miRNAs and acquired integrated information from miRNAMap database which includes locations of miRNAs on chromosomes, annotation of miRNAs and target genes.

The annotations of the coding genes were obtained from Ensembl database (58), Gene Ontology (59) and HGNC gene grouping/family data (60). The conserved regions among the genomes in the database are obtained from the UCSC Genome Browser (54). Several useful tools were integrated in miRNAMap to identify miRNA functions and structures. Table 2.3 and Table 2.4 show the integrated databases and tools in miRNAMap.

Table 2.3 The list of the integrated external data sources in miRNAMap.

Integrated Databases	Description	Literature cited
miRBase	Known microRNAs	(3,61)
UCSC Genome Browser	Conserved regions of human, mouse, rat, and dog	(54)
Lu et al. ‘s work	Gene expression profiles of known miRNAs	(47)
Gene Ontology	Gene annotations	(59)
Ensembl	Genomic sequences and gene annotations	(58)
HGNC	Gene annotations	(60)

Table 2.4 The list of the integrated annotated tools in miRNAMap.

Integrated Tools	Description	Literature cited
miRnada	Predicting of miRNA targets	(55)
TargetScan	Predicting of miRNA targets	(35),
RNAhybrid	Predicting of miRNA targets	(62)
RNAz	Identifying non-coding structural RNAs	(49)
RNAfold	RNA structural analyses	(49)
UCSC Genome Browser	Genome browser for particular locations	(54)
mmiRNA	Identifying the mature miRNA in miRNA precursor	Not published yet
Mfold	Drawing the miRNA secondary structures	(63)
Sfold	Accessibility of target genes predicted	(56)

2.4.3 **Collection of gene expression data**

2.4.3.1 Characterization of microRNA expression profiles in 40 normal human tissues

A new type of real time reverse transcription (RT)-PCR-based miRNA assays were recently developed by Applied Biosystems, Inc, that have better sensitivity and specificity compared to bead- and microarray-based technologies. We used miRNA expression profiles from Applied Biosystems, Inc.(40), the expression of 345 human miRNAs was quantitated in 40 normal human tissues that included brain, muscle, circulatory, respiratory, lymphoid, gastrointestinal, urinary, reproductive, and endocrine systems. In recent study, miRNAs expressed in certain tissue types (47). Many miRNAs with co-regulated expression patterns are located within the same genomic clusters, and candidate transcriptional factors that control the pattern of their expression may be identified by a comparative genomic strategy. Figure 2.5 shows the pattern of expression in tissues without regarding the abundance of miRNA, but identification of tissue-specific miRNAs (40). Analysis the expression profiles can obtain the information of universally expressed miRNAs, and several groups of miRNAs expressed exclusively or preferentially in certain tissue types.

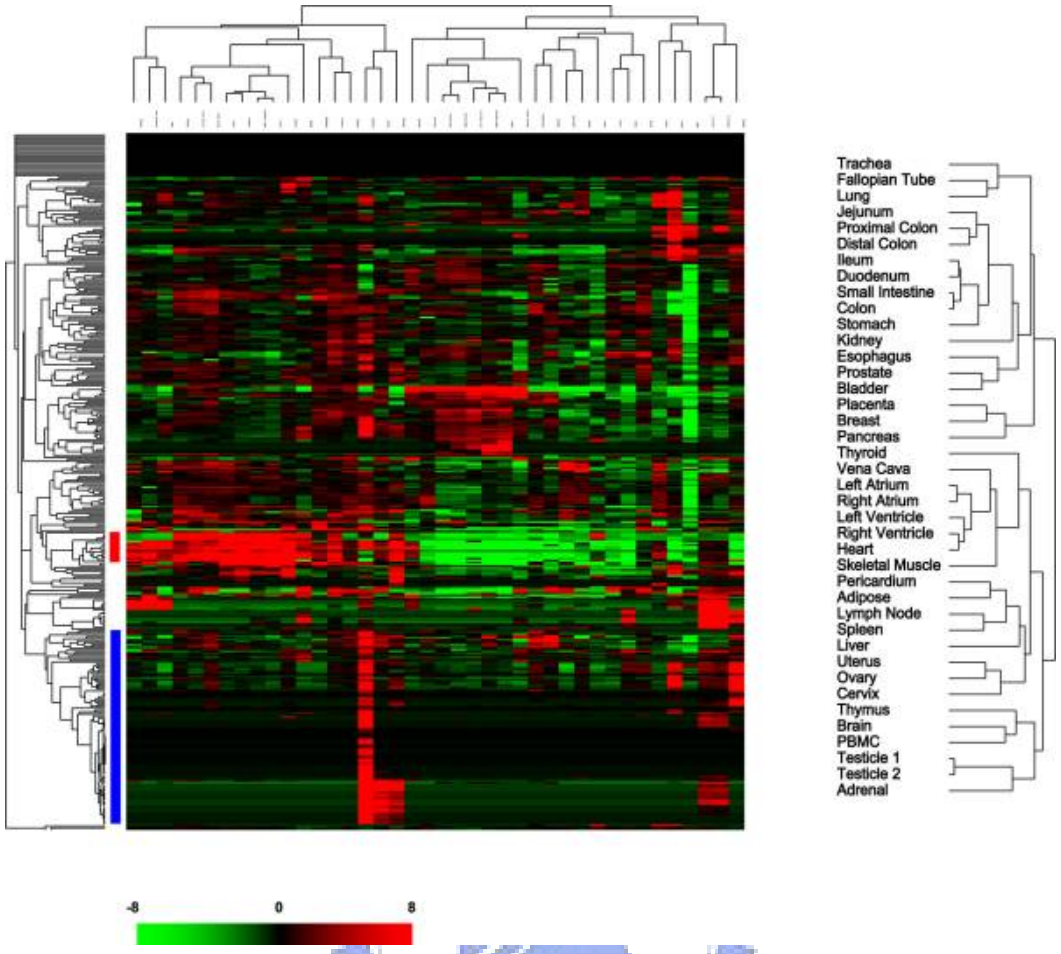


Figure 2.5 Unsupervised hierarchical clustering of the normal human tissues based on the variation of miRNA expression correlates with the anatomical locations and physiological functions of the tissues (Liang et al., 2007).

2.4.3.2 Large-scale analysis of the human transcriptome (HG-U133A) from 79 normal tissues obtained from various sources

We analyzed the expression profiles of transcription factors, miRNA target genes and the other coding gene in NCBI GEO dataset GDS596 (64). This dataset provides a large-scale analysis of human 14048 genes in 79 normal tissues from various sources. The tissue-specific pattern of mRNA expression can provide important clues about gene function. The same expression pattern involves that tissue-specific miRNA and coding gene have a close regulatory relationship, on the other hand, the opposite expression pattern indicates that the coding gene may be downregulated by miRNA or

transcription factor.

2.4.3.3 Differentially-expressed gene profiles between normal and cancer tissues

We collected eight tumor microarrays and related references from GEO and PunMed system (Table 2.5). To compare the gene expression profiles in normal and tumor cells can provide a clue to check our prediction of miRNA regulatory pathway. Using gene expression profiles can also find the most potential genes which are involved in the mechanisms of tumorgenesis and metastasis.

Table 2.5 The sources of gene expression profiles in cancer study.

Cancer	GEO Acc.,	Ref
Hepatocellular carcinoma	GSE6222	Liao et al., 2008
Brain cancer	GSE2223	Bredel et al., 2006; Markert et al., 2001
Colon cancer	GSE13067; GSE13294	Jorissen et al., 2008; Birkenkamp-Demtroder et al., 2002; Birkenkamp-Demtroder et al 2005
Breast cancer	GSE11951	Bertucci et al., 2002
Lung cancer	GSE1037	Jones et al., 2004
Ovarian cancer	GSE6008	Hendrix et al., 2006; Bignotti et al., 2007
Prostate cancer	GSE6919	Chandran al., 2007; Yu et al., 2004
Pancreatic cancer	GSE7824	Nakamura et al., 2007; Crnogorac-Jurcevic et al., 2001; Missiaglia et al 2004

The overall design and introduction of each microarray data from GEO are transcribed as follows:

Hepatocellular carcinoma

The dataset GSE6222 was used in HCC study. Total RNA was extracted from

human liver cancer at various stages (T1-1~4, T3-1~6), 2 normal human livers and HuH7 cell line. Genes showed differentially up- or downregulated by two-fold with respect to the normal human livers were identified using GeneSpring software version 7.2.

Brain cancer

The dataset GSE2223 was used in brain cancer study. Gene expression profiling in 50 glial brain tumors and 4 normal brains were used in cDNA microarrays (65). 6,706 Genes with expression in 80% of samples and whose expression levels differed by at least 2-fold, in at least one sample, from their mean expression levels across all samples were included in downstream statistical analyses.

Colon cancer

Two datasets (GSE13067 and GSE13294) were used in colon cancer study. The tissues were taken from 74 and 155 colorectal cancer patients respectively. Gene expression of about 6,800 known genes and 35,000 expressed sequence tags (ESTs) on five pools (four to six samples in each pool) of total RNA from left-sided sporadic colorectal carcinomas was used in oligonucleotide microarrays (66). 908 known genes and 4,155 ESTs were identified that changed remarkably from normal to tumor tissue. Based on intensive filtering 226 known genes and 157 ESTs were found to be highly relevant for colorectal cancer (66)

Breast cancer

The dataset GSE11951 was used and for this study, the MDA-MB-231 cell line was used as a prototypic mesenchymal and invasive cell line, spontaneously expressing high levels of CD146. Using whole-genome DNA microarrays, we investigated genes for which expression was modified by CD146 down-regulation, obtained by siRNA or shRNA technology (67)

Lung cancer

The dataset GSE1037 was used in lung cancer study. The neuroendocrine tumours from 38 patients undergoing surgery were analyzed and were classified into large-cell neuroendocrine carcinoma (LCNEC) and small-cell lung carcinoma (SCLC) as distinct groups (16). These genes were then filtered across the neuroendocrine-tumour samples passing any gene for which the log expression ratio varied either above 2·0 for any six of the 38 neuroendocrine tumours or below 0·5 for any six. This process resulted in a set of 2803 genes that were regulated across the neuroendocrine-tumour samples.

Ovarian cancer

The dataset GSE6008 was used in ovarian cancer study. The purpose of this study was to identify genes that are highly differentially expressed in metastatic serous papillary ovarian tumors (MET) when compared with primary ovarian serous carcinomas (OSPC). About 14,500 human genes were used to determine whether patterns of gene expression may differentiate OSPC from MET in 31 snap-frozen serous papillary ovarian carcinomas. Hierarchic cluster analysis of gene expression in OSPC and MET identified 156 genes that exhibited 2-fold differences ($P < 0.05$) and that distinguished OSPC from MET (68).

Prostate cancer

The dataset GSE6919 was used in prostate cancer study. The clinical characteristics of the 64 primary tumor samples were used in this study. Affymetrix GeneChip HGU95av2, HGU95b and HGU95c arrays were used. The metastatic samples are highly heterogenous in expression; differential expression analysis shows that 415 genes are upregulated and 364 genes are downregulated at least 2 fold in every patient with metastasis. (69)

Pancreatic cancer

The dataset GSE7824 was used in pancreatic cancer study. Using Affymetrix HG-U133-Plus 2.0 array and Laser Capture Microdissection techniques, the growth in different zones of the same tumor affected expression of genes were determined by human pancreatic cancer cells. Human L3.6pl pancreatic cancer cells were implanted into the pancreas of nude mice. Gene expression patterns in tumor cells within the central and peripheral zones were compared and statistical differences were determined for 1222 genes (70). Bioinformatic functional prediction analysis revealed that 346 upregulated genes in the peripheral zone were related to cytoskeleton organization and biogenesis, cell cycle, cell adhesion, cell motility, DNA replication, localization, integrin-mediated signaling pathway, development, morphogenesis, and IkB kinase/NF-kB cascade (71); and 876 upregulated genes in the central zone were related with regulation of cell proliferation, regulation of transcription, transmembrane receptor protein tyrosine kinase signaling pathway, response to stress, small GTPase mediated signal transduction, hexose metabolism, cell death, response to external stimulus, carbohydrate metabolism, and response to wounding (71).

2.5 Results

2.5.1 Cancer-related miRNAs in different cancer types

Figure 2.6 shows the numbers of literature collections about human cancer-related miRNAs in latest five year. We collected 205 miRNAs which are aberrantly expressed among twenty-seven cancers or cancer cell lines (See Appendix I) and Figure 2.7.

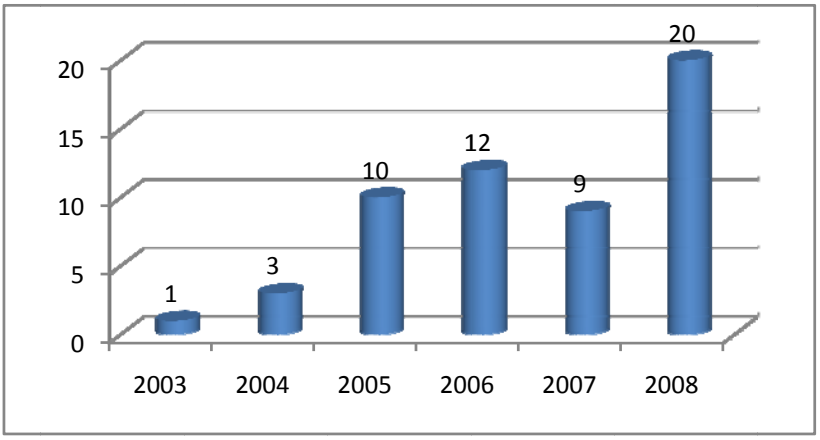


Figure 2.6 Statistics of literature collection about cancer-related miRNAs during last five years in this study.

Certain researches are reported that some miRNAs significantly are upregulated in most of the tumor types, like miR-21 which are overexpressed in bladder cancer, brain cancer,b reast cancer, cervix cancer, colon cancer, hepatocellular carcinoma, lung cancer, ovarian cancer, pancreatic cancer, prostate cancer, stomach cancer and thyroid cancer. The ubiquitous tumor suppressor gene PTEN was shown to be a direct target of miR-21, and to contribute to miR-21 effects on cell invasion in hepatocellular carcinoma. (17681183). miR-21 might play the same role as oncogene in other cancers.

Identical miRNA can be up- or downregulated in different tumor types. For example, miR-221 was downregulated in prostate cancer, but upregulated in most of other cancers. Down regulation of miR-181a was observed in brain cancer, colon cancer and hepatocellular carcinoma, but overexpression in pancreatic cancer, prostate cancer and breast cancer.

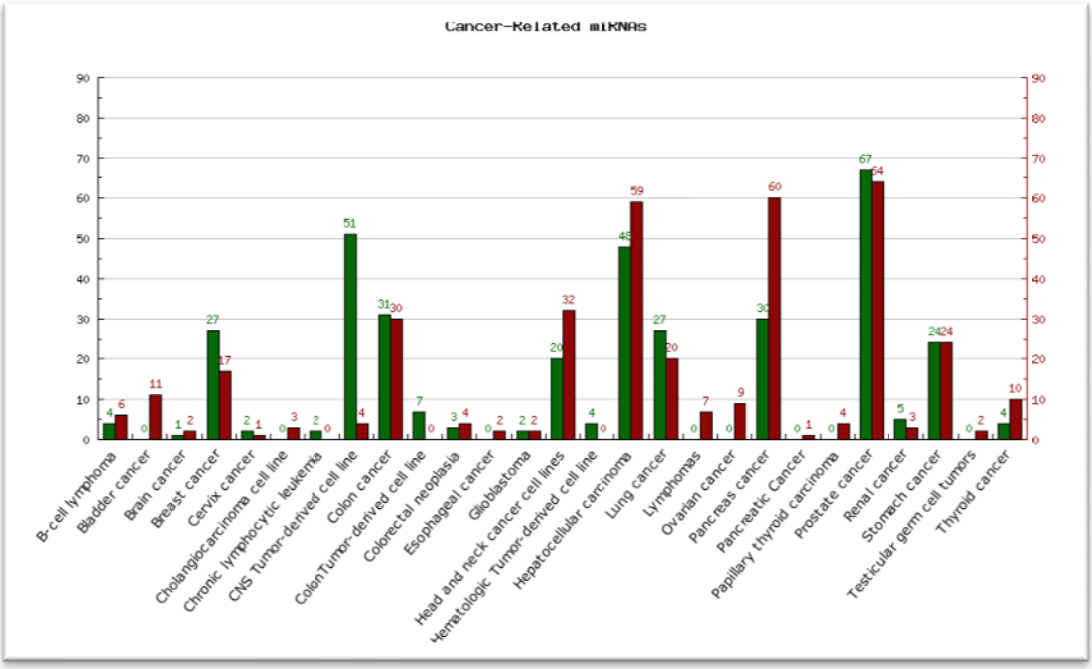


Figure 2.7 Statistics of cancer-related miRNAs in different cancer types.

Figure 2.8 shows the distribution of cancer-related miRNAs in human genome, some evidences have indicated that many miRNAs have been mapped to chromosome aberrant regions in cancers (72), the location of miRNAs in deleted regions, amplified regions and breakpoints could be involved in human cancers, For example, seven miRNAs of the miR-17-92 cluster (miR-17-5p, miR-17-3p, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92-1) located at 13q31, a region commonly amplified in lymphomas and provide an oncogenic function (18). More than half of cancer-related miRNAs are located in protein-coding genes or mRNA-like noncoding RNAs (mlncRNAs) (Figure 2.9). It suggests that approximately half of cancer-related miRNAs share the same promoters with their host genes and are expressed by a similar mechanism.

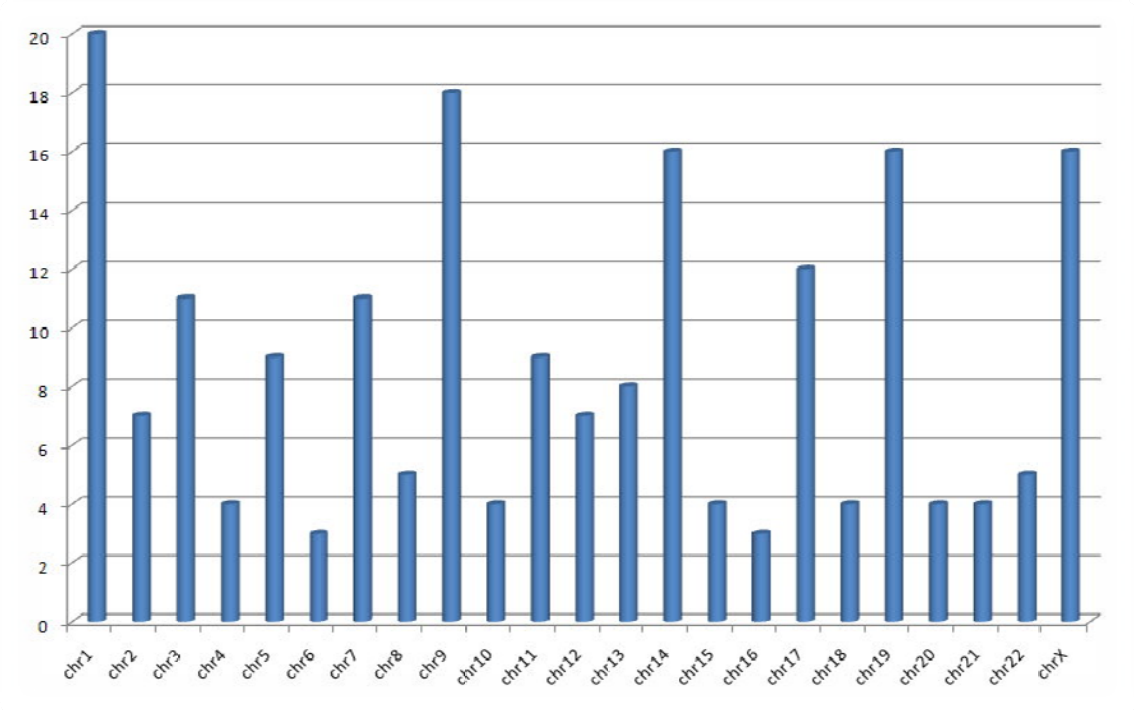


Figure 2.8 The distribution of cancer-related miRNAs in human genome.

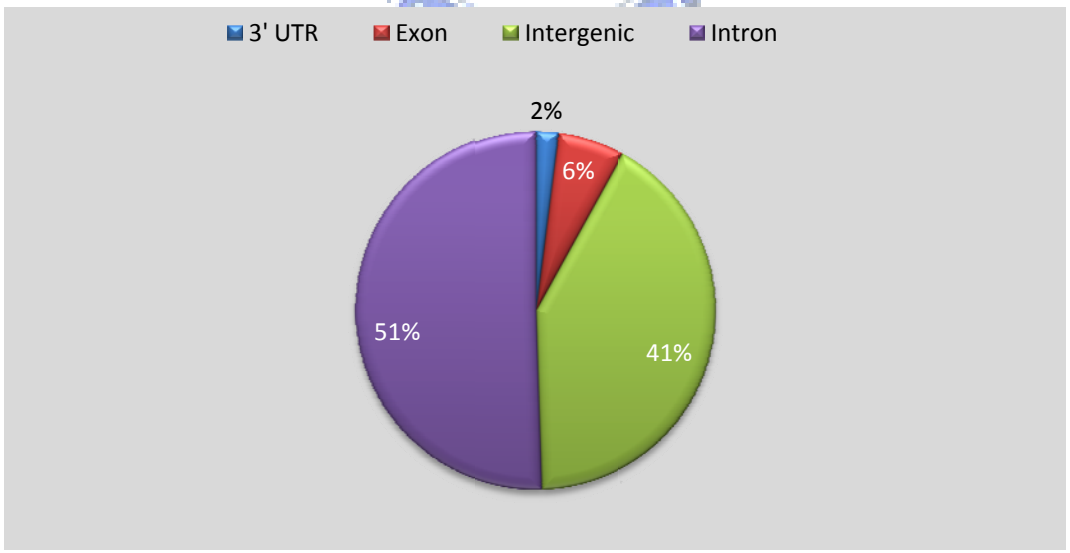


Figure 2.9 More than half of cancer-related miRNAs are itragenic miRNAs.

Chapter 3 Identification of cis-element of cancer-related miRNAs

3.1 Introduction

To discover what regulatory mechanism results downregulation or upregulation of miRNAs in tumor cells, we have to identify cis-regulatory modules and transcription factor binding sites. Alought much attention has focused on finding the targets of both miRNAs and TFs, the transcriptional elements that regulate miRNA expression remain largely unexplored. To understand the regulatory mechanism of miRNAs in cancer study, we identified complex interactions among transcription factors, miRNAs and their target genes.

The first step to understand the transcriptional regulation of intergenic miRNA is to know the exact position of transcriptional start sites (TSS) of intergenic miRNA. It is believed that the distance between the precursor of intergenic miRNA and its TSS is relatively more diverse then the distance between the coding gene and its TSS. Moreover, due to the lower expression level of miRNAs, it is more difficult to obtain full-length cDNAs of miRNA primary transcripts than coding genes. Therefore, it is crucial to determine the potential TSS of intergenic miRNAs based on previous investigation of RNA poly II transcription.

In this study, we identified the potential TSS of cancer-related miRNAs by using computational promoter prediction programs (Eponine, EP3, NNPP and Promoter2.0) and experimental data (CAGE tag, Solexa tag, EST and H3K4me3 locations). Then we used TRANSFAC database and MATCH program to search transcription binding sites in the region 3000 base-pairs (bp) upstream of the TSS. To reduce the false positive in our prediction and to determine the most potential

cis-regulatory elements, we developed a computational method of systematically identifying tissue-selective transcription factor. The expression profiles of those transcription factors were evaluated by microarray. Those transcription factors with the aberrant expression in tumor cells were chosen as most important cis-regulatory elements, and they may upregulate or downregulate their target miRNA in cancer.

3.2 Related works

There are two public databases associated with gene transcriptional regulation analysis (Figure 3.1). DBTSS (http://dbtss.hgc.jp/) is a database collecting gene transcriptional start sites, which are supported by experimental evidence, i.e., the experimentally-determined 5'-end sequence of full-length cDNAs (73). In addition, DBTSS includes human TSS supported by around 19 million of illumina Solexa tags. Another public database is Eukaryotic Promoter Database (EPD) (http://www.epd.isb-sib.ch/), which collects non-redundant experimentally confirmed promoter of RNA polymerase II transcribed in eukaryotic (74). Both DBTSS and EPD offer the information about transcriptional start sites of protein coding genes; however, none of biological database collects the transcriptional start sites for intergenic miRNAs. It is crucial to construct a resource for determining TSSs of intergenic miRNAs for further research of transcriptional regulation.

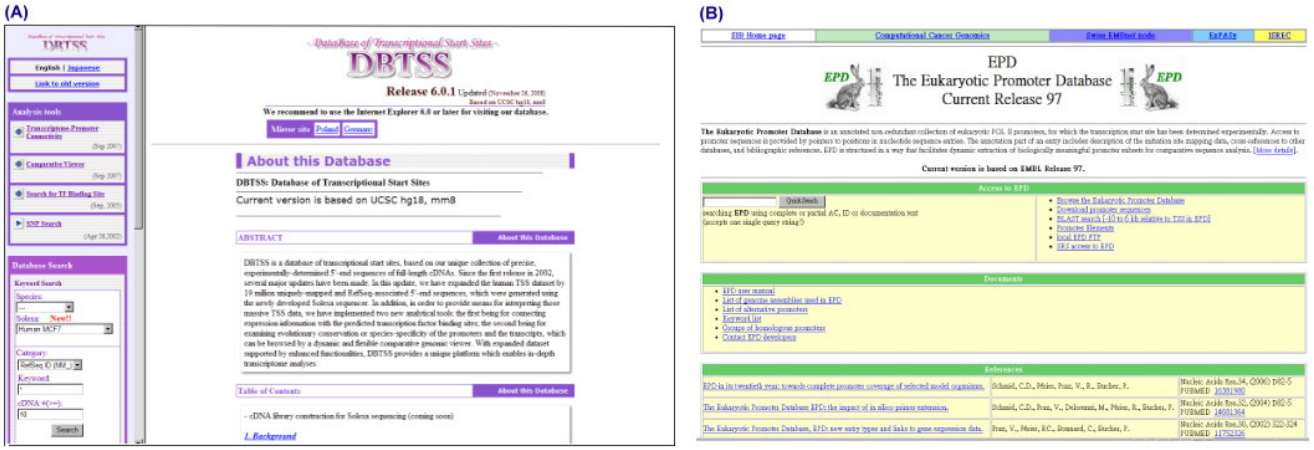


Figure 3.1 The web interfaces of (A) DBTSS database (B) EPD database.

A variety of tools have been developed for identifying putative transcriptional start sites (TSS) of protein coding genes, such as EP3 (75), Eponine (47), NNPP (76) and Promoter 2.0 (77).

EP3

EP3 (.http://bioinformatics.psb.ugent.be/) is a tool for the identification of the core region of a eukaryotic gene promoter. It uses universal properties of the promoter to detect those regions in a whole genome context. EP3 has been tested on several eukaryotes ranging from protists to human. It is efficient to identify regions that are associated with transcription initiation. EP3 provides graphical user interface (Figure 3.2) and stand-alone command line tool available for download.

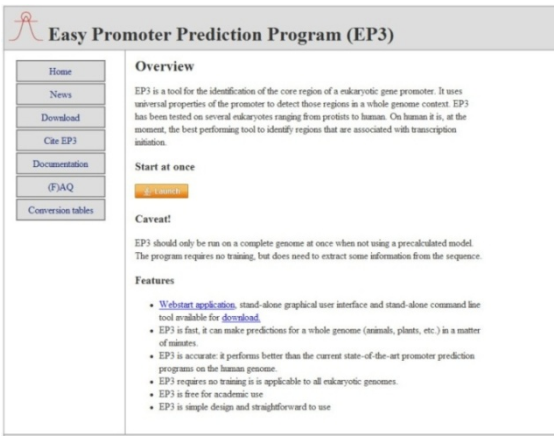


Figure 3.2 Graphical user interface of EP3.

Eponine

Eponine is a probabilistic method for detecting transcription start sites in mammalian genomic sequence, with good specificity and excellent positional accuracy. Eponine models consist of a set of DNA weight matrices. A weight matrix is a simple generative model for a short, ungapped sequence motif. Each model is associated with a position distribution relative to the transcription start site and has the properities as follows (Figure 3.3): (1) CpG enrichment downstream of the start site, (2) a TATAAA motif, with a tightly focused distribution centered at position -30 relative to the transcription start site, (3) two GC-rich matrices closely flanking the TATA box.

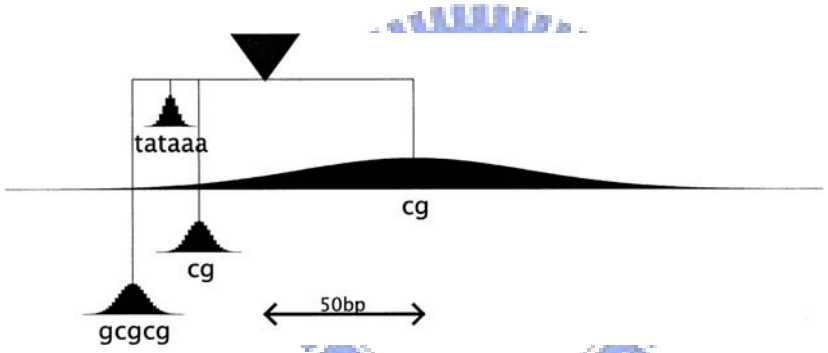


Figure 3.3 Schematic of Eponine core promotor model, showing the constraint distributions and weight-matrix consensus sequences (47).

NNPP

NNPP provides a web application to find eukaryotic and prokaryotic promoters in a DNA sequence (http://www.fruitfly.org/seq_tools/promoter.html) (Figure 3.4). The NNPP program is a time-delay neural network which consists mainly of two feature layers, one for recognizing the TATA-box and one for recognizing the "Initiator", which is the region spanning the transcription start site. A set of 419 promoter sequences from experimental data was used as training dataset (76). Both feature layers are combined into one output unit, which gives output scores between 0 and 1.

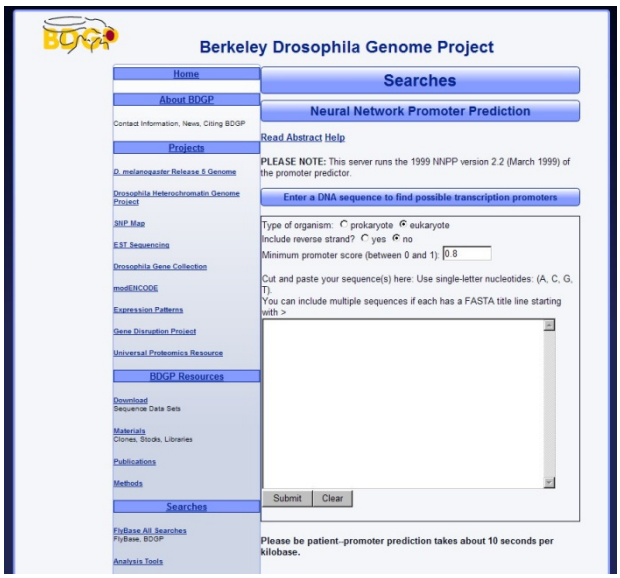


Figure 3.4 Graphical user interface of NNPP.

Promoter 2.0

Promoter 2.0 (http://www.cbs.dtu.dk/services/Promoter/) predicts TSS of vertebrate PolII promoters in DNA sequences (Figure 3.5). It has been developed as an evolution of simulated transcription factors that interact with sequences in promoter regions. It builds on principles that are common to neural networks and genetic algorithms.

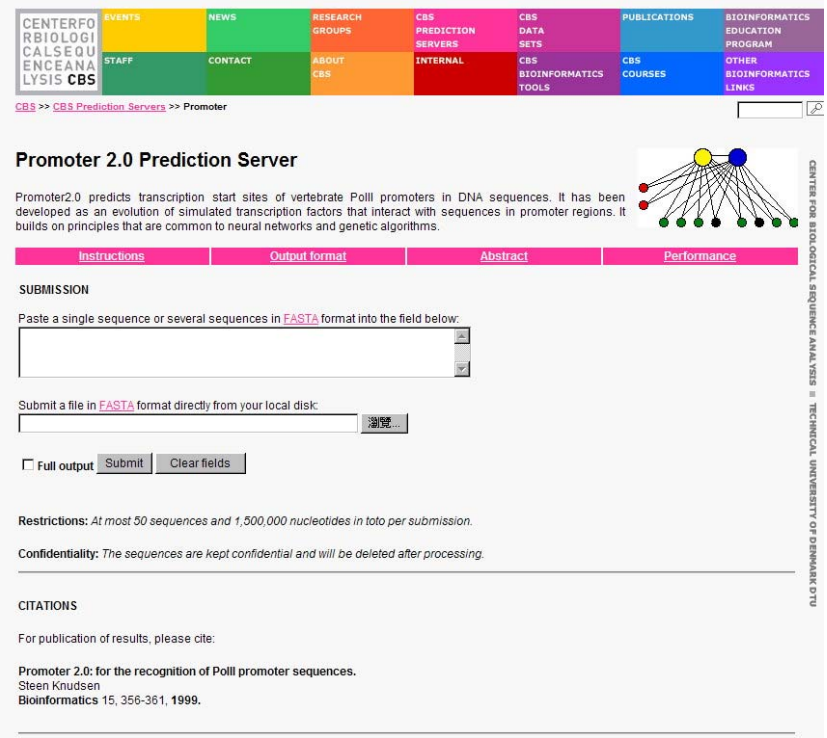


Figure 3.5 Graphical user interface of Promoter2.0.

3.3 Materials and methods

3.3.1 Predict potential TSS of intergenic miRNAs from computational and experimental data

We developed a systematic method for identifying promoter region of intergenic miRNAs for further study of finding cis-regulatory elements. We collected computationally identified TSS of intergenic miRNAs in human by using the tools as Table 3.1. The detail functions of those tools are mentioned in previous section.

Table 3.1 Transcription start sites prediction tools.

Tools	Ref	Method	Species
EP3	Abeel et al.,2008 (75)	Structural features of DNA identification	eukaryote
Eponine	Thomas et al., 2002 (47)	Relevance vector machine	mammalian
NNPP2.2	Reese et al., 2001(76)	Neural network	prokaryote/ eukaryote
Promoter 2.0	Knudsen (77)	Neural network	vertebrate

Experimentally determined 5'-end sequences were evaluated for TSS identification by mapping sequences on human genome. Cap-analysis gene expression (CAGE) tags are ~20 nts sequences derived from the mRNA sequence in the proximity of the cap site (78). Because all of the RNA polymerase II transcripts were altered at the 5' end nucleotide from a guanosine to 7-methylguanosine during transcription, the genomic position of CAGE tags can represent for the transcription start site (TSS). In addition, CAGE tags having an identical 5' terminal site were grouped into a CAGE-tag starting site (CTSS). If CAGE tags are mapped to the upstream flanking region of miRNA precursor, the location of tag can be viewed as a TSS of the intergenic miRNA. The greater number of the CAGE tags are clustered in a region, the more likely to locate the TSS. Millions of human high-quality CAGE tags were obtained from FANTOM3 database (http://fantom.gsc.riken.jp/). Only tag clusters with at least two mapped tags on the same genomic location were considered to be real TSSs, with which most false positives should be removed. This filtering mechanism resulted in 123,400 unique start sites for human genes.

Solexa tag is a type of new-generation, high-throughput sequencing tag which DNA templates are immobilized on a special surface fluorescently labeled nucleotides with specific enzyme (79). In the late update of DBTSS, they incorporated oligo-capping method with Solexa sequencing technique (73). The oligo-capping method could ensure that the cDNA is full-length because of the ability of the substitution of 5' cap structure with the 5' oligo and 5' oligo sequence could be represent as the cDNA with 5' terminal sequence. With the massively parallel sequencing technology and full-length cDNA sequence, the thousands of TSS information has been generated. The solexa tags can be obtained from DBTSS database (http://dbtss.hgc.jp/). TSSs which were supported by >= 5 sequences were counted for

the Solexa tags and 29,210 unique start sites for human.

The histone H3 is trimethylated at its lysine 4 residue (H3K4me3) at the transcriptional start sites of most genes in the genome, even when genes are not productively transcribed, and the knowledge that this covalent modification is restricted to sites of transcription initiation(80,81). We used H3K4me3 enriched loci upstream to miRNAs which are identified by Marson et al by using ChIP-Sequencing (ChIP-Seq) data. (82). H3K4me3 enriched regions are identified on the upstream of 113 intergenic miRNAs.

To accurately determine the TSS, a ranking method for each putative TSS was applied in this work. The ranking score of a site ***x*** is based on the density of predicted site around the site in the upstream of miRNA gene and define as:

$$D(x) = \sum_{i=1}^{n} \left(\frac{1}{(|\mathrm{Loc_i} - \mathrm{Loc_x}| + 1)^2} \right)$$

$$W(x) = \begin{cases} 5 & \text{if site } \boldsymbol{x} \text{ is from CAGE tags data} \\ 2 & \text{if site } \boldsymbol{x} \text{ is from DBTSS Solexa tags data} \\ 1 & \text{otherwise} \end{cases}$$

$$Score\ (x) = D(x) + W(x)$$

Where the $\mathrm{Loc_i}$ are location of predicted site and the $\mathrm{Loc_x}$ is location of the site to be ranked. There are ***n*** predicted site in the upstream of the miRNA gene. There are three factors effectively to the ranking score in our function. The first is the number of neighbors and the second is the distance to the neighbors, if a putative TSS with much more putative TSSs close-by, it will have highest score. The third is the weight values which depend on sources of putative TSSs, the direct evidences from mRNA sequences, like CAGE tags and Solexa tags, are given higher weight.

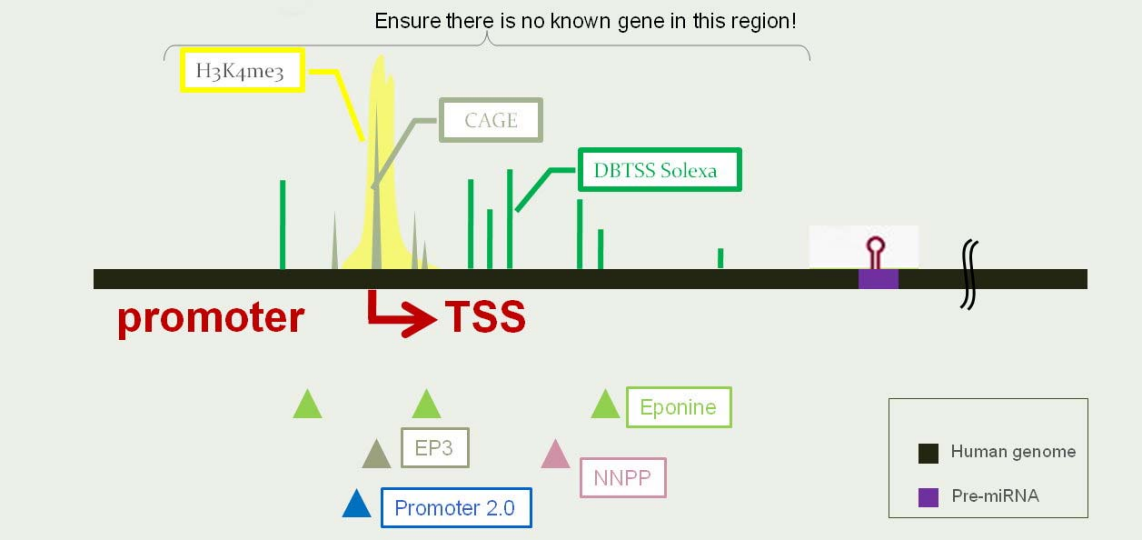


Figure 3.6 Predict TSS of intergenic miRNAs from computational and experimental data.

3.3.2 Identifying transcription factors control the syn-regulated expression patterns of miRNA and coding genes

We used Match^{TM} which is a position-specific weight matrices (PWMs) tool for searching putative transcription factor binding sites (TFBS) in DNA sequences. Match^{TM} uses the matrix library collected in TRANSFAC database. PWM method can scan genomic DNA sequences and identify potential TFBSs. But the majority of predicted sites are false positives that have no biological significance. We combined predictions with gene-expression data and used the knowledge of syn-regulation to reduce the false-positive rate.

3.3.2.1 Clustering the expression profiles to find the syn-expression of miRNA and coding genes

The tissue-specific pattern of mRNA expression can provide important clues

about gene function. The same expression pattern involves tissue-specific miRNA and coding gene has a close regulatory relationship. Gene expression profiles provide specific molecular signatures containing information able to explain the mechanisms of tumor development and progression.

Table 3.2 The source of gene and miRNA expression profiles among normal tissues.

Specimen	#Tissues	Method	Ref
coding genes	79	microarray	(64)
miRNAs	40	real time qPCR	(40)

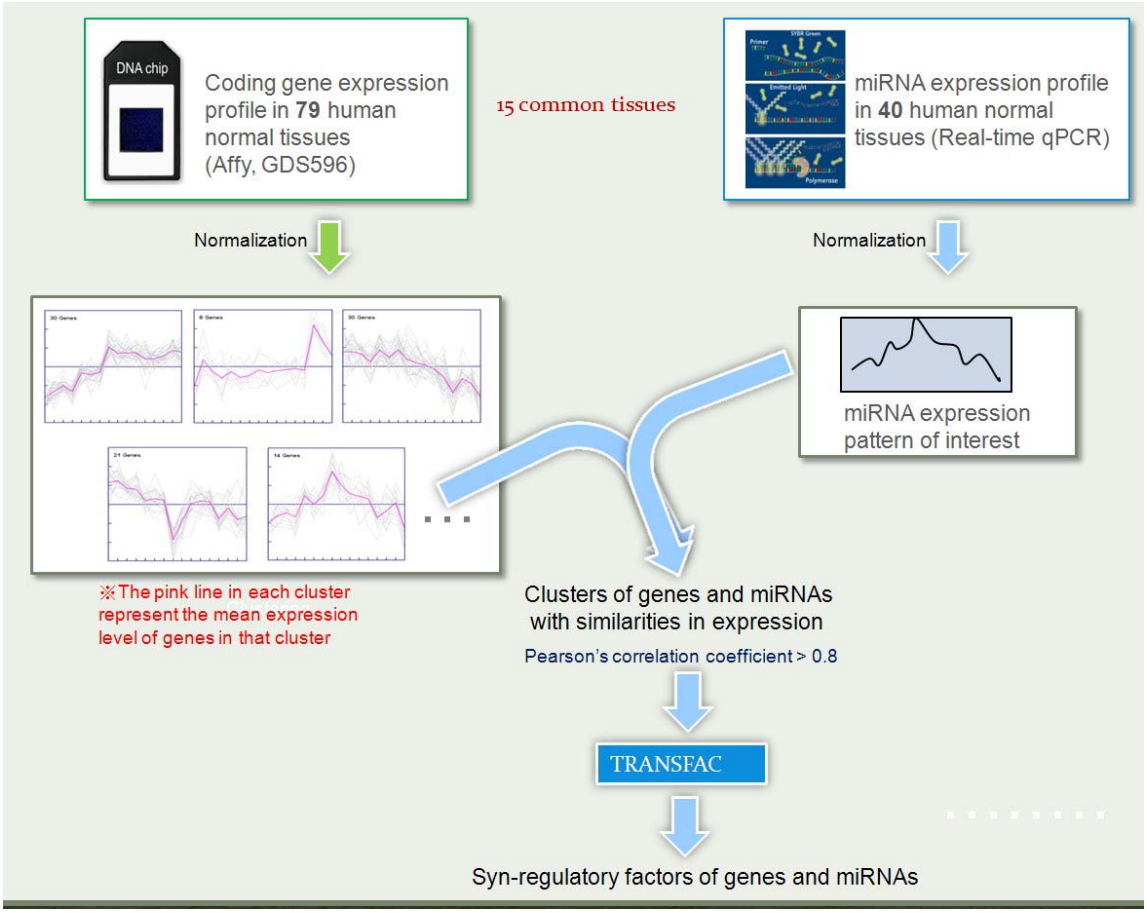


Figure 3.7 Analysis of syn-regulatory factors of genes and miRNAs.

The gene and miRNA expression profiles among human normal tissues have been used in Table 3.2. Fifteen normal tissues are identical between these two data sets:

brain, pancreas, ovary, placenta, uterus, prostate, heart, lymph node, lung, liver, skeletal_muscle, thymus, thyroid, trachea and kidney. The normalization and hierarchical clustering were reconstructed by software Hierarchical Clustering Explorer v3.5 (83).

The same expression patterns of coding genes and miRNAs are clustered with stronger correlation (r = 0.8). Pearson correlation coefficients were calculated using the equation:

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

Where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, s_x and s_y are the sample standard deviations of X and Y and the sum is from i = 1 to n.

3.3.2.2 Obtainment of miRNA and gene promoter sequences to identify transcription factor binding sites

The TSSs of intronic miRNAs are considered the same positions with their host genes. 3 kilobases upstream of TSS are obtained from human genome as the promoter regions; the whole genome sequences for human (hg17) were retrieved from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/).

The program Match™ was used for scanning upstream sequences of TSSs for potential TFBSs. Figure 3.7 shows an example of Match™, a PWM-based tool, from the TRANSFAC database. The sequences that have been shown experimentally to bind to the human transcription factor GATA-1 have 14 positions, among which only positions 6–10 are fully conserved. Abbreviations: R, G or A (purine); N, any; S, G or C (strong); D, G or A or T. Twelve sequences were used to build this matrix (84).

TRANSFAC accession number: M00127
TRANSFAC identifier: VSGATA1_03
Name: GATA-1
Description: GATA-binding factor 1

Position	A	C	G	T	Consensus sequence
1	4	1	2	0	R
2	1	1	3	2	N
3	1	2	4	0	S
4	2	2	2	1	N
5	3	0	2	2	D
6	0	0	12	0	G
7	12	0	0	0	A
8	0	0	0	12	T
9	12	0	0	0	A
10	8	1	3	0	A
11	1	4	4	3	N
12	3	4	3	2	N
13	3	1	7	1	G
14	2	4	4	2	N

Statistical basis: 12 selected binding sequences

Figure 3.8 An example of a position-specific weight matrix (PWM) adapted from the TRANSFAC database.

3.4 Results

3.4.1 TSS of intergenic and cancer-related miRNAs

The TSS of 78 intergenic and cancer-related miRNA are identified in this work (Figure 3.7). The heighest ranking score of putative TSS was selected from other candidates which are derived from experimental data and computational data. The majority of TSSs occur within 4 kb upstream of the pre-miRNA (Table 3.3).

Table 3.3 The locations of TSSs within upstream of pre-miRNAs.

TSS upstream to pre-miRNA	#TSSs
< 1 kb	10
1 kb ~ 2 kb	15
2 kb ~ 3 kb	16
3 kb ~ 4 kb	12
4 kb ~ 5 kb	20
5 kb ~ 6 kb	5

miRNA	Chr.	TSS	Strand
hsa-let-7a-1	9	-2832	+
hsa-let-7a-2	11	-1201	-
hsa-let-7f-1	9	-3222	+
hsa-let-7i	12	-1201	+
hsa-mir-100	11	-4701	-
hsa-mir-101-1	1	-2801	-
hsa-mir-106a	X	-3460	-
hsa-mir-10a	17	-3728	-
hsa-mir-106b	2	-2529	+
hsa-mir-122	18	-4811	+
hsa-mir-125a	19	-5150	+
hsa-mir-125b-1	11	-2001	-
hsa-mir-129-1	7	-2952	+
hsa-mir-129-2	11	-3001	+
hsa-mir-130a	11	-4145	+
hsa-mir-130b	22	-2527	+
hsa-mir-132	17	-2590	-
hsa-mir-136	14	-4753	+
hsa-mir-138-2	16	-3201	+
hsa-mir-141	12	-801	+
hsa-mir-142	17	-1201	-
hsa-mir-143	5	-4801	+
hsa-mir-145	5	-333	+
hsa-mir-146a	5	-5194	+
hsa-mir-146b	10	-801	+
hsa-mir-147	9	-2801	-
hsa-mir-148a	7	-2985	-
hsa-mir-181c	19	-1601	+
hsa-mir-181d	19	-2001	+
hsa-mir-182	7	-2601	-
hsa-mir-183	7	-5250	-
hsa-mir-184	15	-4783	+
hsa-mir-187	18	-4401	-
hsa-mir-192	11	-1901	-
hsa-mir-197	1	-1700	+
hsa-mir-200a	1	-4642	+
hsa-mir-200b	1	-3883	+
hsa-mir-200c	12	-401	+
hsa-mir-203	14	-1721	+

miRNA	Chr.	TSS	Strand
hsa-mir-206	6	-4001	+
hsa-mir-210	11	-1901	-
hsa-mir-212	17	-2218	-
hsa-mir-216a	2	-4801	-
hsa-mir-219-1	6	-258	+
hsa-mir-221	X	-801	-
hsa-mir-222	X	-4801	-
hsa-mir-223	X	-4774	+
hsa-mir-23a	19	-4001	-
hsa-mir-27a	19	-3201	-
hsa-mir-296	20	-401	-
hsa-mir-299	14	-4301	+
hsa-mir-29a	7	-3501	-
hsa-mir-29b-1	7	-2801	-
hsa-mir-29b-2	1	-3901	-
hsa-mir-29c	1	-4501	-
hsa-mir-30b	8	-4001	-
hsa-mir-30d	8	-3601	-
hsa-mir-320a	8	-1522	-
hsa-mir-329-1	14	-3601	+
hsa-mir-345	14	-2314	+
hsa-mir-34a	1	-4001	-
hsa-mir-34b	11	-1651	+
hsa-mir-372	19	-914	+
hsa-mir-373	19	-1729	+
hsa-mir-375	2	-1601	-
hsa-mir-376a-	14	-5140	+
hsa-mir-410	14	-4001	+
hsa-mir-422a	15	-4101	-
hsa-mir-487a	14	-1701	+
hsa-mir-494	14	-1501	+
hsa-mir-498	19	-294	+
hsa-mir-510	X	-2801	-
hsa-mir-513a-	X	-801	-
hsa-mir-520h	19	-2801	+
hsa-mir-7-2	15	-5139	+
hsa-mir-9-2	5	-1001	-
hsa-mir-96	7	-3401	-
hsa-mir-99b	19	-4508	+

Figure 3.9 The TSS of intergenic cancer-related miRNAs.

3.4.2 Syn-regulated expression patterns of miRNA and coding genes

We performed an unsupervised hierarchical clustering based on the variation of gene and miRNA expressions among 15 different tissue types, 84 groups of expression patterns are identified (Figure 3.10). The genes and miRNAs in the same groups are with the similar expression pattern ($r > 0.8$).

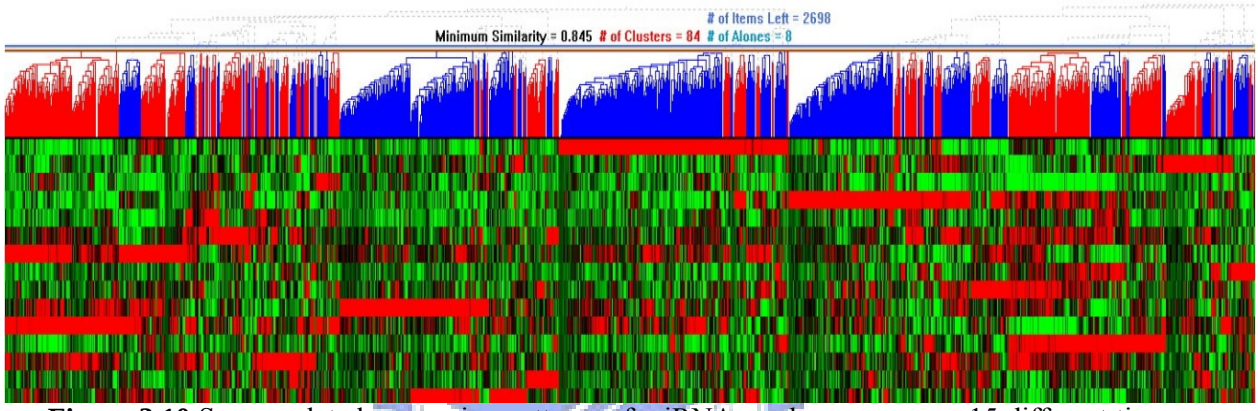


Figure 3.10 Syn-regulated expression patterns of miRNAs and genes among 15 different tissue types.

We used TRANSFAC to predict the transcription factors that can co-operate the cancer-related miRNAs and coding genes in the same expression patterns. We selected those transcription factors that are syn-regulated with cancer-related miRNAs by using tumor microarray data (See Table 2.5). A case study will illustrate the strategy in next section.

3.5 Case study

3.5.1 Cis-regulatory elements of miR-122 in hepatocellular carcinoma

3.5.1.1 Potential TSS of miR-122

Hsa-miR-122 is a liver specific miRNA located at 18q21.31, it acts as a key regulator of cholesterol and fatty acid metabolism in the adult liver. In this case study, the putative TSSs of miR-122 are showed in Figure 3.11. We obtained several candidates from experimental data and computational data. The putative TSSs was analyzed the density which depends on number of other putative TSSs and the distance to other putative TSSs. The putative was given a score as Table 3.4.

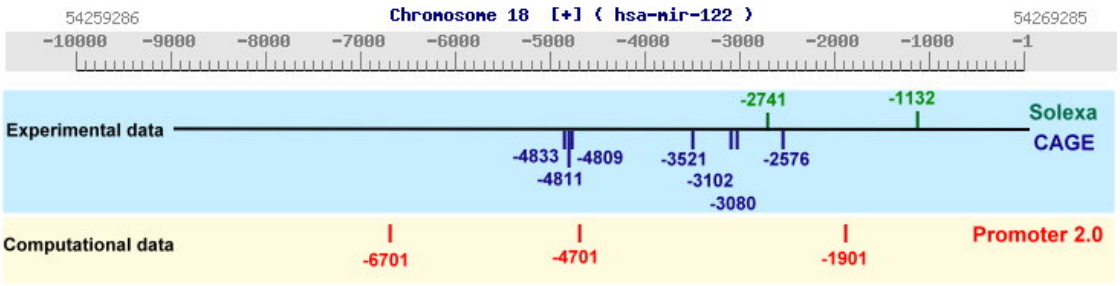


Figure 3.11 Putative TSSs of miR-122.

There are two cage tags (ID: H10BA06E2310 and H28BA88E0701 in FANTOM3 database) mapped the same position -4811 nt (negative number indicates the upstream size from the 5' end position of pre-miRNA which is given 0) and the position got the highest score. And in our analysis, the density shows that potential TSS should be around the position -4833 nt ~ -4809 nt. Using UCSC Genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway), we can find several ESTs are mapped to the upstream of miR-122, and the position of 5' end of EST R01863 is -4800 nt and it

should be very close to the TSS of miR-122 gene. With the support of experimental data, we believe that the position -4811 nt is the transcription start site of miR-122 gene.

Table 3.4 The score of putative TSSs of miR-122.

Data	Putative TSSs	Score
Solexa tags	-1132	2.000012386
Solexa tags	-2741	2.000017192
CAGE tags	-4833	5.006381298
CAGE tags	-4811	6.11300207
CAGE tags	-4811	6.11300207
CAGE tags	-4809	5.11271239
CAGE tags	-3521	5.000005669
CAGE tags	-3102	5.001890359
CAGE tags	-3080	5.000003921
CAGE tags	-2576	5.000013192
Promoter 2.0	-6701	1.00000025
Promoter 2.0	-4701	1.00013852
Promoter 2.0	-1901	1.000032276

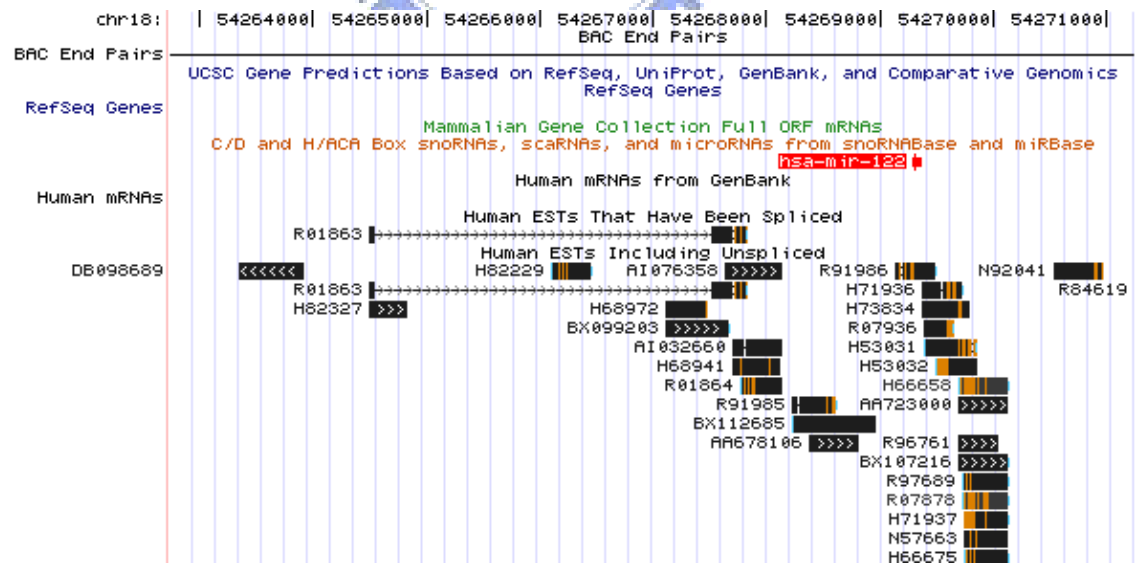


Figure 3.12 The 5' end of EST R01863 should be near the TSS of miR-122 gene.

3.5.1.2 Identifying transcription factors of miR-122

The liver-specific miR-122 and 159 genes are the same expression pattern among 15 normal tissues (Table 3.5). We obtained the promoter sequences of 159 genes by using BioMart (http://www.ensembl.org/biomart/martview/). The transcription factor binding sites are predicted through the 3000 bp upstream sequences. Only the TFs which regulate more than two genes are retained in our data. Table 3.6 shows the TFs of miR-122 and the number of genes were also regulated by the certain TF. It illustrates that the same expression pattern of genes may be co-regulated by certain TFs,

Table 3.5 Liver-specific miR-122 and 159 genes are the same expression pattern.

Gene	BRAIN	PANCREAS	OVARY	PLACENTA	UTERUS	PROSTATE	HEART	LYMPH_NODE	LUNG	LIVER	SKELETAL_MUSCLE	THYMUS	THYROID	TRACHEA	KIDNEY
hsa-miR-122	-0.239659	-0.260517	-0.260517	-0.260517	-0.260517	-0.260517	-0.260517	-0.260517	-0.260517	3.614721	-0.260517	-0.248861	-0.260517	-0.260517	-0.260517
ABCG5	-1.14087	0.602351	-1.85708	-1.205	-0.936516	-0.524885	1.08253	-0.922969	-0.545495	3.53278	-1.39765	-0.263557	-0.102194	0.573969	-0.049726
ACBD4	-0.218005	-0.49115	-0.400002	-0.895161	-0.580575	0.144566	0.359304	0.144566	0.04503	2.23266	0.027108	-0.68011	0.542583	-0.088155	-0.159661
ADH1A	-2.10267	-0.26862	0.024728	-0.844355	-0.524005	-0.788303	-0.021319	-0.273919	-1.0025	3.50181	0.530209	-1.28859	-0.581978	-0.225629	-0.511432
AGT	0.46044	-1.88027	-2.17194	-2.71251	-2.26505	-1.90551	0.894963	-2.31114	-2.1185	4.15208	-0.988467	-2.64247	-1.47255	-2.22088	-0.059868
AGXT	-3.00047	-2.62236	-2.86628	-2.35599	-2.35599	-2.98493	-1.22091	-1.83576	-1.52362	5.76841	-0.287663	-1.69425	-1.76548	-3.10859	-0.505957
AHSG	-3.84891	-4.53499	-3.25363	-4.71747	-4.26839	-2.94054	-2.4893	-4.50671	-5.46529	5.43848	-2.29687	-4.76694	-4.2332	-4.77535	-3.91708
AKR1C4	-1.91581	-1.22992	0.443272	-1.24955	0.750454	-1.11744	0.076329	-1.91581	-2.11744	3.27316	-0.428143	-2.04705	-0.02998	-1.76695	-1.50863
AKR1D1	-0.87907	-1.01805	-1.00996	-0.422317	-0.732228	-0.438497	-0.131836	-0.718992	-0.433084	3.58571	0.796239	-1.01805	-0.557142	-0.217655	-0.908817
ALB	-3.95245	0.019315	-2.23525	-3.97083	-3.66004	-1.89978	-2.65075	-3.51476	-2.88991	4.89349	-2.23386	-3.50803	-2.65075	-2.98944	-1.2854
AMBP	-2.96242	-1.42567	-2.44177	-3.26131	-2.9139	-2.36493	-2.25953	-1.90272	-2.31404	5.59341	-3.0898	-2.15631	-2.4058	-2.73961	-2.28653
ANG	-3.04857	-0.224999	0.592375	-1.0127	-0.985119	-0.033087	-1.33779	-2.4328	-0.04987	5.34658	-0.530889	-1.92434	-1.81879	0.746124	-0.737061
APCS	-1.15167	0.109455	-1.45025	-1.00425	-1.62645	-0.656823	-0.243592	-1.75198	-0.744286	5.07185	-0.158045	-1.75584	-0.779238	-1.5208	-1.40406
APOA1	-4.74849	-3.4378	-5.30362	-4.5108	-3.98422	-2.25081	-1.3048	-4.51812	-4.07581	5.60353	-3.53474	-4.55154	-3.49988	-3.62475	-4.04639
APOA2	-4.27435	-2.83234	-3.85931	-4.14844	-4.19921	-1.95515	-2.83904	-3.82236	-3.74488	5.40033	-2.24072	-4.34406	-4.35841	-3.31577	-3.83234
APOC1	-1.27737	-2.44351	-2.28171	-1.85434	-2.88505	-1.78182	0.042126	0.673352	1.3234	4.9379	-1.33621	-1.19154	-2.20138	-2.61806	-3.24819
APOC2	-2.65974	-3.0956	-3.17019	-1.92443	-3.69744	-2.4702	-3.17019	-2.24264	0.379001	5.57996	-3.21808	-3.0088	-2.74928	-3.16431	-2.71451
APOC3	-4.92549	-5.71295	-5.39498	-4.84749	-5.56292	-2.10673	-5.23035	-5.92549	-5.21627	5.78796	-3.72785	-4.36361	-5.99588	-5.39498	-3.87487
APOC4	-0.55146	-0.249516	-0.841306	-0.672418	-0.580762	0.329241	0.19918	-0.399457	0.001081	3.86779	0.48621	-0.098009	-0.103594	-0.504893	-0.3258
APOH	-5.2809	-2.60092	-2.77959	-4.97121	-5.24082	-3.08192	-4.66246	-4.63623	-4.70272	5.32133	-3.49401	-6.10872	-2.87222	-6.10872	-1.75469
AQP9	-3.97311	-1.93524	-1.78623	-1.10985	-3.27977	-2.30853	-2.73328	-2.34779	0.496055	2.8622	-1.82084	-0.816898	-3.5735	-2.51647	-2.85527
ARG1	-2.76364	-2.08916	-0.352793	-1.52632	-2.11331	-0.751751	-0.374516	-1.36905	-1.20763	2.36209	-0.319055	-1.63573	-1.20115	-1.55911	-0.510196
ASGR2	-3.06356	-1.63686	-3.74539	-3.04495	-3.02657	-2.38784	-1.87092	-2.79119	-2.90408	5.29083	-1.51657	-2.77576	-2.54899	-2.64385	-2.44771
ATF5	-0.972885	-0.159741	-0.651498	-0.56531	-1.22998	-0.090563	-0.106064	-0.445914	-0.462921	3.08669	0.621072	-0.797257	-0.2099	-0.891325	-0.221108
BPI	-2.33138	-1.98062	-0.544975	-1.87075	-2.34468	-0.880896	-1.60771	-1.80494	-0.728877	1.09519	-0.564289	-1.59177	-1.86116	-2.14537	-1.21692
C2	-0.875579	0.349584	-0.439818	0.82202	-1.2304	-0.498412	0.567342	0.626807	0.869712	2.95311	0.512757	-0.41703	-0.411389	0.150899	-0.530155
C4A	-1.04164	0.211983	0.078331	-1.32679	-0.448617	-0.323323	0.517798	0.042805	0.93483	4.21679	-0.939157	0.58085	1.10054	-1.11849	0.827998
C4B	-1.04164	0.211983	0.078331	-1.32679	-0.448617	-0.323323	0.517798	0.042805	0.93483	4.21679	-0.939157	0.58085	1.10054	-1.11849	0.827998
C4BPA	-0.583554	-0.106094	-0.431551	-0.001866	0.062264	0.187359	-0.702853	-0.01725	2.60479	3.61783	-0.338901	-0.182074	-0.01725	-0.1796	-0.461298
C4BPB	-1.83052	-0.515647	-0.446605	0.040746	-0.921286	0.311774	-0.446605	-0.544216	-1.30543	3.8611	0.050332	-0.884292	-0.233876	-0.795755	-1.23388
C6	-0.708023	1.3169	0.099332	-0.71021	-0.64808	-0.633476	-0.312793	-0.317788	-0.230424	3.67887	-0.552744	-0.48563	-0.040922	-0.761462	-0.008256
C8A	-1.96486	-1.23125	-1.1623	-1.83653	-1.06919	-0.261838	-0.484231	-0.915388	-1.2212	3.93868	0.512761	-0.806159	-1.21121	-0.346727	-0.303658
C8B	-0.75925	-1.64163	-1.09306	-0.842368	-0.993708	-0.017835	0.695617	-0.447383	-0.082571	3.25926	-0.472277	-0.651937	-0.210004	-0.09131	-0.053251
C8G	-3.00148	-3.00148	-3.00148	-3.00148	-2.54205	-3.00148	-2.01963	-3.00148	-3.00148	6.03249	-1.91402	-3.00148	-2.86398	-3.00148	-3.00148
CACNA1F	-0.028581	-0.124796	-0.338921	-0.358286	-0.685096	-0.709759	0.850113	-0.469707	0.200445	1.88141	-0.020845	-0.512457	-0.044178	0.321004	0.082173
CARKL	-1.23124	-0.471332	-0.856985	0.254908	-1.27202	-0.155353	0.293854	-0.43086	-0.67163	1.6537	-0.005975	-0.604981	-0.737933	-1.11074	0.274512
CES1	-3.0408	-1.29456	-1.65159	-1.00178	-0.212327	-1.10018	-0.283447	-2.53557	2.1096	5.18208	-0.795326	-3.36273	-1.39508	1.13737	-2.49675
CFHR2	-0.617671	-0.463781	0.046884	-0.489467	-0.657881	-0.650135	0.529414	-0.091812	-0.520907	2.872	-0.25699	-0.930657	0.156944	0.014354	-0.91665
CFHR4	-1.76625	-1.21371	1.29553	-2.35121	-1.44432	-0.684457	0.101646	-1.86579	-1.24688	3.82371	1.5913	-1.52339	-1.36936	-0.684457	-1.14958

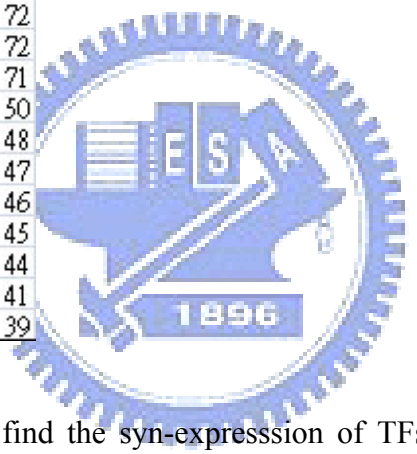
Gene	BRAIN	PANCREAS	OVARY	PLACENTA	UTERUS	PROSTATE	HEART	LYMPH_NODE	LUNG	LIVER	SKELETAL_MUSCLE	THYMUS	THYROID	TRACHEA	KIDNEY
CHST10	0.502907	-0.278704	-0.014941	-0.308163	-0.326131	0.440954	0.458116	-0.308907	0.28002	1.47205	-0.259154	-0.072545	0.281999	-0.646177	-0.222934
CLEC1B	-0.728986	-0.387752	-0.469694	-0.614084	-0.656107	-0.352776	0.658697	-0.543366	-1.40855	1.42171	-0.420576	-0.263451	0.313368	-0.817617	-0.318628
COLEC11	0.081284	0.2421	-0.075114	-0.245944	-0.27268	0.292075	0.794245	0.236277	0.673391	2.80018	-0.696741	0.536749	1.17747	-0.610262	0.972075
CPA4	-0.593722	-0.299056	0.249727	0.026344	-0.007575	0.37666	0.194213	-0.217399	-0.125916	0.83436	-0.213298	-0.028044	0.448301	-0.162343	-0.171595
CPB2	-1.58081	-0.964158	-0.381769	-0.92609	-0.874427	-0.265499	-0.384348	-0.878057	0.538797	4.187	-1.56854	-1.59816	-0.793376	-0.577361	-1.3016
CPN1	-0.248869	-0.119547	-0.318856	-0.304863	-0.908043	-0.277277	-0.196258	-0.205327	-0.02935	2.02927	1.00813	-0.193677	0.070646	-0.089741	0.340997
CPS1	-1.09352	-1.73242	-0.270183	-0.747963	-0.688028	-0.401249	-0.07707	-1.68678	-1.69178	3.27144	-1.46939	-1.61378	-0.408435	-0.551093	-0.914799
CTH	-0.362428	0.404595	0.722552	-1.23725	-0.661339	0.066777	-0.504156	-0.763472	-0.539489	2.32235	-0.706139	-1.63562	-0.132822	-0.735886	0.10389
CYP1A2	-1.91917	-0.637701	-0.455351	-0.837522	-0.425961	-0.611166	-1.06954	-0.508862	0.795898	4.78836	0.084503	-0.807086	-0.431941	-0.795475	-0.635974
CYP2A13	-0.54186	-0.575005	-0.328167	-0.583907	-0.319055	-0.279147	0.308491	-0.164164	0.05494	3.66365	-0.342365	-0.53513	-0.314932	-0.179826	-0.594863
CYP2A6	-0.790943	-0.906296	-0.657713	-0.763981	-0.851107	-0.68344	0.074098	-0.647053	-0.43633	4.78842	-0.308766	-0.79745	-0.44711	-0.589027	-0.785461
CYP2A7	-0.49644	-0.83629	-0.298101	-0.617122	-0.313709	-0.282661	0.373163	-0.224343	0.071352	3.89195	-0.451506	-0.671383	-0.167405	-0.307836	-0.642765
CYP2A7P1	-0.435858	-0.608679	-0.205772	-0.496253	-0.450505	-0.15974	0.381625	-0.205772	0.111178	3.5951	-0.23456	-0.544423	-0.274627	-0.209431	-0.601981
CYP2B6	-0.95259	0.31019	-0.022908	-0.889919	-0.750824	-0.003815	-0.395343	-1.08674	-0.403658	2.76279	0.31273	-0.802113	0.050457	0.099823	-0.663361
CYP2C18	-0.876738	0.051347	-0.309164	-0.323643	-0.342474	-0.280635	0.245994	-0.274595	-0.590007	1.89399	-0.474722	-0.477082	0.123262	0.480739	0.248793
CYP2C19	-0.674717	-0.367399	-0.212315	-0.699458	-0.541183	-0.0813	-0.114187	-0.760818	-0.370839	3.46975	0.341993	-1.20442	-0.376236	-0.515736	0.132638
CYP2C8	-1.81972	-1.34366	-1.49187	-1.77875	-1.28921	-0.393597	-0.009826	-0.544863	-2.22097	4.76845	0.075643	-1.07565	-1.13351	-1.44992	-1.42712
CYP2C9	-0.667324	-0.539125	-0.250516	-0.692821	-0.707007	-0.283205	-0.02348	-0.665476	-0.608127	3.8113	0.299158	-0.783778	-0.50045	-0.456913	-0.190876
CYP2D6	-1.1209	-1.16148	-1.66434	-1.24509	-1.16799	-1.03604	0.13201	-0.564293	-1.06614	4.97091	-0.313843	-0.994928	-0.536637	-1.02122	-0.456714
CYP2E1	-2.11056	-2.64444	-2.51716	-3.15352	-2.06885	-1.30927	-2.41955	-2.42857	-2.37675	5.92372	-2.2989	-3.17612	-2.49331	-3.27025	-1.85622
DNMT3L	-1.62934	-0.74482	-0.496892	-1.3525	-1.65736	-0.643282	-0.180085	-0.729869	-0.74482	2.30821	-0.410919	-0.307414	0.889386	-1.77519	0.528199
EPHX2	-2.4235	-0.957527	-1.59568	-2.08118	-1.83854	2.07395	-0.01072	-2.14552	-1.75674	3.75541	-1.3108	2.62907	1.07794	-2.37089	2.90428
ERCC8	-0.562797	-0.186546	-0.276535	-0.295575	-0.791845	-0.489216	-0.00822	-0.262844	-1.02617	1.32495	-0.244228	-0.026166	-0.08518	-0.004644	0.022165
ETV3	-0.466765	-0.689158	0.644743	-0.519233	-0.208031	0.011282	1.41238	-0.916568	-0.601695	1.9402	0.440125	-0.952192	-0.44123	-0.144837	-0.814688
F12	-0.787198	-0.657613	-2.05916	-0.793568	-2.09542	-1.36582	-2.14343	-0.89087	-2.584	5.42514	-0.696872	-1.09805	-1.23292	-2.78086	-2.29182
F2	-2.87739	-1.49888	-2.33513	-1.9705	-2.14403	-1.67184	-1.50992	-2.05466	-2.1225	5.73484	-2.90469	-2.07003	-1.48507	-1.75637	-2.64473
FADS2	0.129518	-1.10767	-1.30922	-1.1348	-0.942796	0.862632	-0.630186	-1.03082	-0.80162	2.00224	-2.22119	-0.718479	-1.87405	-1.19776	-0.768052
FAH	-0.835624	-1.36614	-0.79491	-0.67924	-0.845985	-0.145345	0.461838	-1.355	-0.192491	3.57961	0.259258	-0.291672	-0.456517	-1.21414	0.955789
FCN3	-2.39393	-0.358443	-1.65085	0.170619	-3.5392	-3.16445	0.450067	-1.30549	4.48294	4.51273	-1.91009	-1.24691	-1.25945	-1.37199	-0.195426
FETUB	-0.718657	-0.839482	0.226953	-1.12054	-0.454607	0.14048	0.718701	-0.873085	-0.595925	2.2427	0.388585	-0.87984	0.388585	-0.540959	-0.323583
FGA	-2.14628	-0.659888	-1.31438	-1.41789	-1.98713	-1.08568	-1.23684	-1.26711	-1.31185	4.12413	-0.610925	-1.6615	-1.75278	-1.04001	-0.4247
FGG	-4.11783	-1.87456	-0.959404	-1.4681	-2.70464	-2.0821	-1.72411	-3.24069	-2.07433	4.67436	-1.68815	-3.73254	-1.91556	-2.27311	-3.60826
FMO3	-0.594469	-0.311691	-0.169526	-1.10742	-1.07536	-0.588801	-0.560786	0.329735	1.21451	3.10247	0.488308	-0.225602	0.00556	0.794056	-0.225602
GALK1	-1.29872	-1.14985	-1.50085	-1.01492	-1.50085	1.08461	-0.108083	-0.822278	1.07859	4.19314	0.010612	-0.799911	-0.713754	-1.50085	-0.002099
GCKR	-0.190153	0.378918	-0.269084	-0.187897	-0.24773	0.329379	0.682468	-0.329916	0.39255	2.58788	0.36208	-0.63946	0.248398	-0.52114	-0.459977
GGCX	-0.270892	-0.12291	-0.148691	0.421667	-0.559049	0.161377	0.639657	-0.512166	0.077312	1.43346	0.14254	-0.151506	-0.132611	-0.723146	-0.297417
GJB1	0.642941	0.559639	-0.813636	-0.715147	-0.185759	0.578681	0.107126	-0.780054	-0.16061	3.20728	-0.36374	-0.905353	-0.577865	-0.028859	0.840221
GRHPR	0.016426	-0.798478	-1.22301	-0.49405	-0.845747	-0.598196	0.576435	-0.719735	-0.359825	3.33348	-0.654546	-0.519068	-0.467857	-0.866455	0.795399
GRPR	-0.116257	0.406256	-0.220516	-0.129942	-0.521202	-0.428527	0.154266	-0.008295	0.21715	0.989759	0.060056	0.207974	0.072979	0.459948	0.367822
GSTM1	-0.008743	-0.918712	0.224325	-0.697728	-0.19447	0.148434	0.805388	-0.277467	-0.610644	3.6295	-0.258936	-0.148631	0.152012	-0.527803	0.44875

Gene	BRAIN	PANCREAS	OVARY	PLACENTA	UTERUS	PROSTATE	HEART	LYMPH_NODE	LUNG	LIVER	SKELETAL_MUSCLE	THYMUS	THYROID	TRACHEA	KIDNEY
GSTM2	0.143954	-1.3577	0.045817	-0.654984	-0.284963	0.009958	0.950779	-0.2398	-0.976766	3.39623	0.116688	-0.267823	0.200564	-0.585043	0.554596
HAAO	-0.633581	-0.860415	-0.522659	-1.08684	-0.843219	0.332014	0.299868	-2.30698	0.08134	3.45996	0.173774	-0.381924	0.083734	-0.08656	0.696601
HABP2	-0.537359	0.058985	-0.164826	-0.520895	-0.418306	-0.021235	-0.084429	-0.514276	-0.318961	3.30308	-0.475119	-0.24743	-0.180488	-0.912617	0.341335
HAMP	-2.69517	-0.426391	-1.50829	-1.86174	-1.77808	-1.33388	0.933877	-1.97363	-3.14123	5.33904	-1.06985	-2.08989	-1.36102	-1.39801	-1.3309
HBBP1	-0.004474	-0.430635	0.183801	-0.138492	-0.527621	0.338573	0.676918	-0.501583	0.095061	1.00629	-0.153157	-0.453143	0.338573	0.163233	-0.107814
HGFAC	-1.09014	0.437112	-1.55311	-1.27071	-1.61281	-0.582652	-0.361681	-0.892196	-0.968145	4.7732	-0.09464	-1.59264	-0.831823	-0.942382	-0.602889
HP	-2.43546	-1.89767	-1.17447	-2.07817	-2.20544	-1.47501	-0.856635	-1.67633	-0.309201	5.13137	-1.41798	-1.63622	-1.48448	-1.27287	-2.06155
HPR	-3.89048	-2.19083	-1.48893	-4.3576	-3.43757	-3.88028	-0.172826	-2.24213	-0.660432	5.45327	-2.43012	-2.40432	-3.7081	-1.74001	-1.91629
HPX	-2.94708	-1.72549	-3.22863	-2.0471	-2.55146	-2.13416	-2.39111	-2.53795	-2.33554	5.64598	-1.45419	-2.55146	-1.34139	-0.91613	-1.61148
HRG	-1.69122	-1.20697	-0.779099	-1.0173	-1.99871	-1.0854	-1.32166	-0.955246	-1.12629	5.21091	-0.385296	-0.876772	-0.740126	-1.04564	0.627724
HSD17B8	-0.914954	-0.044011	-0.153081	-0.669287	-0.972571	0.364394	1.06145	-0.257606	0.547701	2.58212	0.218288	0.070448	1.16088	-0.402804	0.11143
HTR1A	-0.916454	-0.844901	0.614531	-0.833313	-1.24451	-1.49532	-0.64985	-0.221668	-0.690673	2.58353	0.288366	-1.00468	-0.659546	-0.941116	-0.441545
IFNA8	-1.04561	-1.04561	-1.04561	0.140255	-1.04561	-1.04561	-0.586179	-1.04561	-1.04561	1.35356	-0.100753	-1.04561	-0.560184	-1.04561	-1.04561
IGFALS	-1.27528	0.394737	-0.179938	-0.970015	-2.18395	-1.70467	-0.289633	-0.497194	-2.3129	4.27206	0.408687	-0.661092	-1.20285	-0.666702	-0.541172
INHBC	-0.639512	-0.444927	0.053844	-0.278524	-0.225122	-0.005784	-0.163357	-0.419154	-0.021716	0.707302	-0.261498	0.243285	0.417569	-0.228087	-0.133894
INHBE	-1.15332	-0.467599	-0.843108	-1.45563	-0.874643	-1.26396	-0.898756	-2.12656	-0.981546	3.19204	0.419927	-1.3858	-0.982172	-2.51651	-1.00817
ITIH1	-3.0495	-3.20997	-1.19589	-3.11421	-2.39858	-1.35883	-0.441522	-0.508013	-1.87116	5.45185	-2.28993	-3.29743	-2.8163	-2.78436	-1.6156
ITIH2	-1.14844	-1.38175	-0.486198	-1.19156	-1.37819	-1.04598	-0.14617	-1.00832	-0.884399	3.42433	-1.17756	-1.21044	-1.22476	-0.840847	-0.182988
ITIH3	-1.10496	-1.83315	-1.15271	-1.60049	-1.45227	-1.03297	0.3342	-1.19494	-0.325675	4.82088	-0.952285	-2.08497	-0.511077	-2.53122	-1.2871
ITIH4	-0.773532	-0.412382	-0.806049	-0.656765	-1.28491	-0.744469	-0.594935	-0.799237	-0.516802	4.64595	0.066209	-0.592466	-0.469609	-0.688044	-0.160878
JMJD5	-0.745628	-1.27342	-0.327524	0.152643	-1.19588	-0.867257	-0.834242	-0.639151	-1.67963	2.29523	-0.795596	0.092101	-1.19588	-1.51362	0.119322
KLKB1	-0.680345	-0.084459	-0.892592	-1.16108	-1.08446	-1.18089	-0.411059	0.229987	-0.273555	1.75453	-0.550123	-1.0598	-1.26296	-0.151277	-0.014662
KRT2	-0.857507	-0.375819	0.027765	-0.564292	-0.404178	-0.003531	-0.169056	-0.456424	0.134901	0.564458	-0.290629	0.106093	0.088391	-0.768795	0.207973
KRT34	-1.21738	-0.037673	-0.439771	-0.217379	-0.510161	-0.389145	0.086774	-0.682842	-0.405824	1.39673	0.133695	-0.60327	-0.262283	-0.232176	-0.340256
LBP	-0.641264	0.097252	-0.441205	-0.264544	-0.801045	-0.319336	0.19534	0.006929	-0.053719	2.79208	0.200599	-0.12153	-0.163615	0.125638	0.502571
LCAT	-1.51218	-0.86582	-1.66651	-0.906462	-0.826291	-1.12086	-0.906462	-0.852523	-0.490882	4.75995	-0.215566	-1.39189	-0.598339	-0.852523	-0.941227
LECT2	-2.31872	-1.76376	-1.09652	-1.40119	-2.44912	-2.28353	-0.429754	-0.687029	-1.95813	4.65629	-0.704377	-2.52929	-1.11941	-1.86416	-1.31872
LIMS1	-2.43395	-0.595109	-2.05454	-1.86867	-2.10084	-0.379654	-0.404925	0.267387	-1.14765	2.41108	-1.16702	1.17139	-0.545579	-1.68997	0.993813
LIPC	-2.67084	-1.73796	-1.82285	-3.02594	-2.32292	-2.49801	-1.96085	-2.24275	-2.28228	4.81019	-1.68402	-2.64485	-2.49801	-2.30245	-0.581837
LOC442271	-0.284204	-0.025326	0.249605	-0.569261	-0.1383	-0.001251	0.420591	-0.554172	-0.295313	1.00329	0.527688	-0.452747	0.143571	-0.078518	-0.317789
LOC55908	-1.04816	-1.29814	-0.075748	-0.087274	-1.13836	-1.12677	-0.779674	-1.37831	-1.96327	4.09015	0.675628	0.26059	-0.735004	-0.13836	0.136262
LOC728160	-0.289683	0.510485	0.332432	-0.25559	-0.299128	0.664108	-0.324602	-0.207396	-0.016232	2.58635	-1.16999	0.21134	-0.859049	-1.37962	-0.314995
MASP2	-0.777022	-0.194408	-0.39797	-1.22384	-0.601011	0.013546	-0.005291	-0.522222	-0.042029	3.20456	0.305713	-0.824284	-0.525765	-0.129175	0.463552
MAT1A	-2.86499	-2.33986	-2.88093	-1.98908	-1.91335	-1.3955	-2.22804	-2.18296	-2.20286	5.74636	-0.465892	-2.44204	-1.80298	-2.41859	-2.33986
MGC4859	-1.04341	-0.656392	0.460717	-1.22029	-0.342975	0.114127	0.76394	-1.98897	-0.342975	1.18392	-0.071429	-0.861211	-0.563945	0.416017	-0.996499
MTHFR	0.480708	-1.7994	-0.599931	-1.55613	-1.74168	0.313811	-0.402883	-1.11033	-1.43683	2.28633	-0.264363	-1.28483	0.19922	-1.55613	-1.95468
NNMT	-4.45752	-0.156728	-0.293081	0.082692	-0.763895	-1.60259	0.222099	-0.383984	1.44137	5.02966	0.899009	-1.35713	-1.10789	-0.928214	-3.16234
NPC1L1	-0.507131	-0.2134	-1.4107	-0.801077	-1.3029	-0.373865	1.74115	-1.0019	0.084464	2.89971	-0.416994	-0.760888	-0.435793	0.399095	-0.867808
NR1I3	0.091751	-0.300567	-0.086787	-0.654575	-0.629189	0.326089	0.5106	-0.600127	0.042578	2.54632	0.740586	0.114471	0.058288	-0.633389	0.692739
NR1N	-0.480854	-0.480854	-0.480854	-0.480854	-0.480854	-0.047895	2.09808	-0.480854	-0.480854	3.11906	-0.021422	-0.480854	-0.480854	-0.480854	-0.480854

Gene	BRAIN	PANCREAS	OVARY	PLACENTA	UTERUS	PROSTATE	HEART	LYMPH_NODE	LUNG	LIVER	SKELETAL_MUSCLE	THYMUS	THYROID	TRACHEA	KIDNEY
OPRS1	-0.622553	-0.564144	-0.752499	0.487318	-0.766943	-0.363396	-0.309543	-0.393249	-0.345777	1.84086	-0.516431	-0.625587	-0.053846	-0.769178	-0.628627
ORM1	-4.07869	-4.51439	-3.63992	-4.28409	-3.92328	1.00514	-2.87498	-3.72298	-3.53301	5.88764	-2.48464	-4.34858	-3.62487	-3.79979	-4.48691
ORM2	-3.97303	-4.53545	-3.07297	-3.76204	-3.48598	0.840801	-2.91187	-3.43033	-3.00752	5.85783	-1.89294	-3.87425	-3.62187	-3.69493	-4.31786
PECR	-0.077559	-0.814525	-0.077559	-0.538464	-0.969803	0.280424	0.104744	-0.824373	-0.226295	1.32565	-0.341772	-0.665469	0.294142	-0.313757	0.219423
PIPOX	0.158916	-0.534245	-0.325399	-0.056543	-0.498731	-0.459795	0.327914	-0.866321	-0.275712	3.3442	0.616072	-0.808106	-0.229497	-0.808106	1.30172
PKLR	-1.30924	-1.47917	-1.2179	-1.37681	-1.50054	-0.635895	-0.027473	-0.77873	-1.132	4.38902	0.123366	-1.38672	-0.549824	-1.08525	0.461274
POLR1C	-0.496764	-0.573557	-0.333551	-0.309757	-0.731187	-0.325146	-0.315509	-0.122784	-0.340049	0.642586	0.2259	-0.359063	-0.346577	-0.054397	-0.152216
PON1	-0.539374	-0.377677	-0.592575	-0.507313	-0.840736	-0.484592	-0.22508	-1.12704	-0.792715	3.5179	0.802942	0.006027	0.261017	-0.951413	-1.13518
PON3	-0.991656	0.031458	0.04761	-0.608522	-1.34648	-0.390608	0.119483	-1.36397	0.825926	3.74761	0.461939	-0.548735	0.212593	-0.256245	-0.453395
PTHLH	-0.883049	-0.257153	-0.278771	-0.163213	-0.324072	-0.170893	-0.129155	-1.17765	-0.131026	0.732593	-0.437589	-1.41234	-0.304935	-0.314472	-0.497882
PXMP2	0.100261	-0.890429	-1.12043	-1.16057	-0.423195	0.690027	1.49811	-2.06085	0.102806	3.59259	-0.349439	0.590169	-0.067967	-1.13836	1.49262
RDH16	-0.127801	-1.53993	-1.51159	-1.1278	-1.48437	0.539089	-0.816442	-1.9089	0.629086	5.29171	0.054938	-1.34319	-0.940431	-1.81644	-1.08601
RNASE4	-2.06678	-0.051358	0.948926	-0.743017	0.518564	0.92631	-1.0406	-0.555659	0.282454	4.29013	-0.505755	-1.123	-0.17096	1.17071	-0.191477
SAA4	-0.640452	-0.757135	-0.314913	-0.710842	-0.800917	-0.214608	0.124444	-1.15141	0.020799	3.89766	-0.319248	-0.796882	-0.358856	0.103007	-0.745422
SARDH	-0.485941	-0.628316	0.356149	-0.759328	-0.836144	0.137315	0.288839	-0.564681	-0.228813	2.44056	-0.340289	-0.409879	-0.289249	-0.382187	0.735609
SCLY	-0.362075	-0.388686	0.277909	-1.10933	-0.282685	-0.169037	0.056255	-0.295817	-0.015004	1.31057	-0.49895	-0.382942	0.355169	-0.27085	0.121468
SDS	-0.088968	-1.26944	-1.1527	-0.579915	-0.920816	-0.192652	0.067523	-0.646641	-0.414526	4.59904	-0.633893	-1.53973	-0.133136	-0.8163	-0.818702
SERPINA4	-0.880149	0.224582	-0.460452	-0.836681	-1.05305	-0.677778	-0.250621	-0.660184	-1.1582	4.36433	-0.484764	-1.0097	-0.730354	-1.12765	0.009668
SERPINC1	-4.31311	-3.79228	-4.08178	-4.02981	-3.68692	-3.15017	-2.99118	-4.19281	-3.55127	5.72671	-2.52379	-4.13623	-4.05531	-3.92978	-3.08178
SERPIND1	-3.06564	-3.34127	-2.05315	-3.16966	-4.00424	-2.26007	-3.34127	-2.51881	-1.19008	5.50841	-0.087759	-3.05315	-2.19008	-2.7412	-1.8026
SLC10A1	-0.944773	-0.057064	-0.118105	-0.668473	-1.55895	-2.37349	0.680453	0.068773	-0.275019	3.37196	-1.42008	-0.260124	-0.056145	0.124968	0.216049
SLC17A2	-1.81928	-1.5271	-1.97456	-1.0246	-1.55135	-0.56363	-1.04168	-0.9262	0.224217	3.55469	-0.19559	-2.11207	-0.56363	-2.11207	-1.62664
SLC22A1	-2.14571	-2.14571	-2.07948	-2.03397	-3.63696	-3.36488	-1.43415	-2.72103	-4.04295	5.6836	0.027435	-3.61	-3.90545	-3.25636	-2.87303
SLC27A5	0.494928	-3.05558	-1.37193	-2.12087	-2.1684	-0.141049	-2.52703	-1.88707	-1.40003	5.64464	-2.44949	-2.50944	-0.443217	-2.62783	-2.44949
SLC35D1	-0.767514	-0.308082	-0.596476	-0.16595	0.191844	0.406302	0.093918	-0.312644	-0.256201	1.21453	-0.613258	0.446265	0.090467	-0.235266	0.431067
SLC38A3	-0.341776	-0.105796	-0.936091	-0.911668	-0.607143	-0.19983	-1.92208	-0.864027	-0.593204	4.2256	0.215736	-0.406118	0.004648	-1.20194	-0.512284
SLCO2B1	-0.065756	-0.321542	-0.356956	1.44403	-0.508003	-0.8957	0.401645	0.742436	1.80834	3.13744	-0.432014	0.064164	-0.137229	-0.312475	-0.082257
SPP2	-2.372	-0.612207	-0.179635	0.07693	-1.03774	-0.290279	-0.253126	-0.477557	0.279269	4.06656	-0.521059	-0.599198	0.358753	-0.746317	-0.458636
SPTBN5	-0.071599	0.064192	-0.293795	-0.334163	-0.610243	-0.368446	0.308903	0.141349	0.219405	1.05452	-0.00379	-0.10734	-0.498452	0.547475	0.186327
SULT2A1	-1.21633	0.21135	-0.042348	-0.874226	-1.86783	-0.476119	-0.078828	-0.270154	-0.190155	2.4706	-0.368335	-0.028011	-0.026229	-0.038751	-0.877435
TAT	-1.9022	-0.01574	-0.440098	-0.952129	-1.40839	-0.844488	0.167251	-0.784463	-0.665164	4.10552	-0.011764	-0.970019	-1.35368	-0.788993	-0.73337
TDO2	-0.204337	-0.278646	-0.669161	-0.498636	-1.04686	-0.547463	0.08232	0.110921	0.145124	2.2042	0.596572	-0.95507	-0.102323	-0.453733	0.224339
TDRKH	-0.255095	-0.379696	0.05748	-0.011902	-0.220276	-0.016713	0.508231	-0.218058	-0.327362	0.752276	0.075673	-0.038072	0.160795	-0.019607	0.114909
TFR2	-0.918777	-1.16953	-2.33701	-1.87426	-1.21255	-0.891632	-1.05971	-2.5328	-1.15817	4.06828	-0.236259	-2.02243	-1.11515	-2.89837	-2.12302
TM4SF4	-1.77811	0.797659	-2.41554	-1.28027	-2.04942	-1.79887	-1.28291	-1.69146	-1.863	4.14924	-1.28027	-1.82702	-1.711	0.450154	-1.26836
TMPRSS6	-0.639508	-1.08825	-0.502753	-0.700755	-1.01778	0.099573	0.200235	-0.329623	-0.292173	3.16423	0.461516	-0.318285	-0.098719	0.252491	-0.320169
UGT2B15	-0.721292	0.46343	-0.018947	-0.402342	-0.249743	0.246117	-0.490426	-0.300688	-0.171458	2.57311	-0.448787	-0.219667	-1.04819	-0.542083	-0.363136
UGT2B17	-0.415973	0.363376	0.340573	-0.392273	-0.415973	0.643996	-0.389338	-0.329037	-0.127281	2.4176	-0.521281	0.05401	-0.918224	-1.44311	-0.383485
UGT2B4	-1.93737	-1.93737	-1.93737	-1.93737	-1.93737	-1.93737	1.37006	-1.93737	-0.020892	3.97372	-1.93737	-1.93737	-1.93737	-1.93737	-1.93737
UNC93A	-1.55819	-1.20918	-0.168041	-1.25768	-1.08365	-1.16226	0.820824	-0.978296	-1.4276	1.92452	-0.288839	-1.07276	0.255661	-0.448563	-0.710191
VTN	-3.4786	-2.37299	-2.38577	-0.897443	-2.90279	-1.66473	-0.022488	-3.04091	-2.68962	5.53974	-1.10995	-3.16399	-3.08889	-3.89363	-1.26471

Table 3.6 The TFs co-regulate miR-122 and other genes.

TFs of miR-122	#Genes are co-regulated
NR3C1	147
TEAD1	129
NR5A1	126
HMGA1	115
ETV4	112
ZBTB16	106
ZEB1	92
CYP27B1	91
KLF12	91
NFE2	86
FOS	78
FOSB	78
JUN	78
JUNB	78
JUND	78
CUX1	76
IRF1	76
EREG	72
ESR1	72
TFAP2A	71
PLAU	50
GTF2I	48
NCOA6	47
TCF7L2	46
NFE2L1	45
CREB1	44
GDNF	41
TRIM63	39



Then we tried to find the syn-expresssion of TFs and miR-122. miR-122 is downregulated in liver tumors of both human and rodents (85). Only 3 transcription factors: NR3C1, KLF12 and TCF7L2 are downregulated in hepatocellular carcinoma (the expression T1 stage > T3 stage). The three genes are most potential cis-regulatory elements of miR-122.

Chapter 4 Identification of target genes of cancer-related miRNAs

4.1 Introduction

In 2003, it was shown that the fly miRNA bantam targets and down regulates the pro-apoptotic gene hid (86). Using genetic approaches, other miRNA targets had been found in Caenorhabditis elegans before, but hid was the first target identified by performing a genome-wide, sequence-based bioinformatic screen for targets of a miRNA. To understand how many and what kind of genes are regulated by cancer-related miRNAs, the direct method is to predict the miRNA target sites through the base paring to the 3'-UTR of the mRNAs.

The basic compositions of algorithm to predict miRNA targets are as follows: first, searching the 3'UTR sequence to find the segments of perfect Watson-Crick complementarity to bases 2–8 of the miRNA from 5' end, the segments are called “miRNA seeds”. The seed correlated with both mRNA degradation and translational repression (12). Second, extending each seed match to a longer “target site” or using pairwise alignment to calculate the weight of base pairing. Third, calculating the free energy of the miRNA:target site interaction (kcal/mol); Figure 4.5 shows a miRNA target site which is based on seed region searching and minimum free energy calculation.

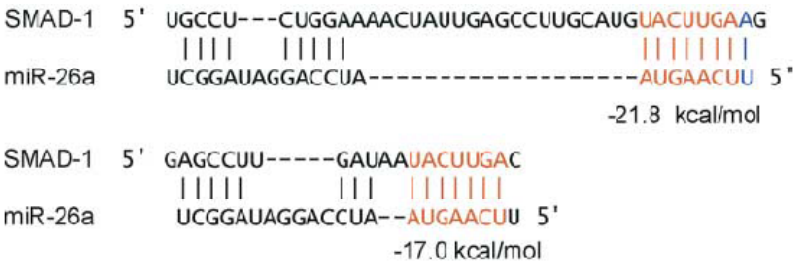


Figure 4.1 Structures and energies for predicted RNA duplexes involving human miR-26a and two target sites in the 3´ UTR of the human SMAD-1 gene, with seeds and seed matches in red and seed extension in blue (35).

In recent years, there are several useful tools to predict the miRNA targets, it makes easier to study the function of miRNA. But the defect is that a miRNA usually has thousands of putative target genes and hard to define which gene is the most significant regulated by miRNA, so it is crucial to reduce the false positive of putative miRNA targets. In this study, we combined with miRanda miRNA target prediction program and array chips to significantly reduce the number of candidate miRNA targets and had higher prediction rate.

4.2 Related works

Recently, many biologists have been paying much more attention to the functions of miRNAs in biological systems. Several miRNA target prediction tools were have been developed, such as miRanda (8), TargetScan (9) and RNAhybrid (10), for determining the energetically favored hybridization sites of small RNA to large RNAs. Lu et al. (11) developed an miRNA microarray to measure the expression profiles of all known miRNA in various normal tissues and tumors.

Many of the initial algorithms share a common approach of evaluating miRNA-target complementarity first, then move to evaluating the binding site thermodynamics to further prioritize. The putative target pool is then filtered, often by requiring species conservation. Some of the more widely used include:

miRanda

miRanda (http://www.microrna.org) was introduced in 2003 by researchers Rockefeller, Memorial Sloan-Kettering, and the Columbia Genome Center. A three step algorithm, miRanda performed the following tasks: (i) Searched for complementarity between miRNAs and 3' UTRs with to complementarity near the 5’ miRNA seed region (ii) Calculated the thermodynamics of the binding sites

When using these parameters miRanda was able to identify many known targets in Drosphilia and showed a false positive rate in the range of 24-39%, however, when researchers used the multiple sites miRNAs often exhibit in their mRNA targets, this rate improved. Additionally, updates to the original miRanda use a stricter seed pairing rules that improve the output and more recent updates, including the integration of a statistical model, further assist in improving specificity, making miRanda one of the more widely used programs. The program is available to download from website (Figure 4.1).

Figure 4.2 The website of miRanda program.

TargetScan and TargetScanS

TargetScan (http://www.targetscan.org/) deviates from miRanda in that it addresses some of the filtering at an earlier stage by requiring perfect complementarities to the seed region of the miRNA and by selecting for species conservation. TargetScan then follows a similar path as miRanda, evaluating the predicted targets by their thermodynamic stability using programs from the Vienna RNA Package.

The first algorithm to be applied to human target prediction, TargetScan showed a slightly improved false positive rate (22-31%) and could predict novel targets relatively well. Updates to the algorithm surfaced in TargetScanS (http://genes.mit.edu/tscan/targetscanS2005.html), a simplified version of TargetScan and today the miRNA-target complementarity is limited to six nucleotides of the seed

region (bases 2-7). Figure 4.2 shows the web applications of TargetScan and TargetScanS.

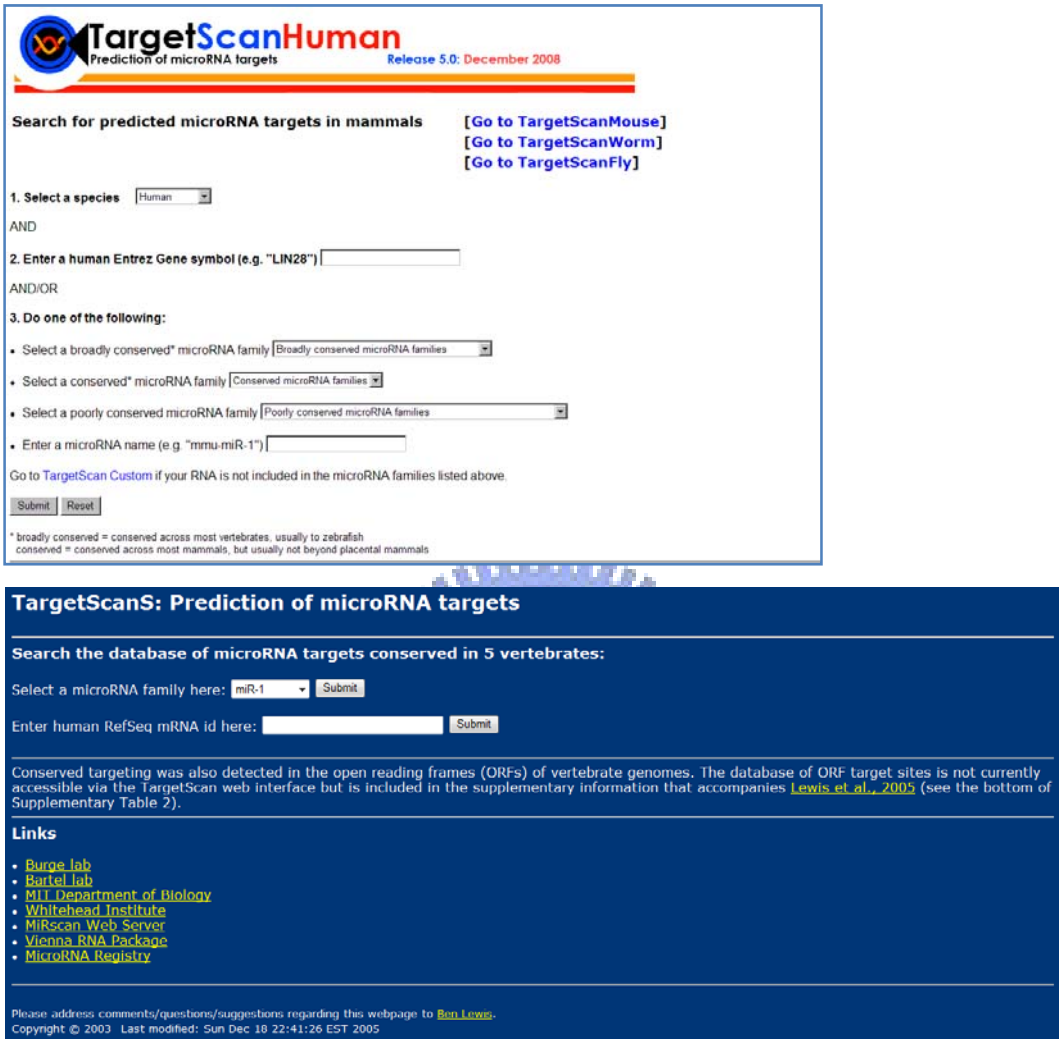


Figure 4.3 Web applications of TargetScan and TargetSanS.

RNAhybrid

RNAhybrid is a tool for finding the minimum free energy hybridization of a long and a short RNA. The hybridization is performed in a kind of domain mode, like the short sequence is hybridized to the best fitting part of the long one. The algorithmic core of RNAhybrid was modified in 2006, seed match function had been added. RNAhybrid can be used as webservice and the program is available at http://bibiserv.techfak.uni-bielefeld.de/rnahybrid (Figure 4.3).

Figure 4.4 RNAhybrid can be used as webservice and the program is available at website.

PicTar

PicTar (http://pictar.mdc-berlin.de/) (Figure 4.4) is an interesting algorithm that enables the prediction of miRNA targets by first aligning input orthologous 3' UTRs and a search set of co-expressed miRNAs, mapping target sites, then filtering them by their predicted free energy. The initial version of PicTar demonstrated similar prediction efficacy as miRanda and TargetScan(S) but an interesting added value PicTar brings to the table is the ability to identify targets that may be regulated by multiple miRNAs. The validation of PicTar actually proved very useful as it illustrated the coordinate regulation of Mtpn gene by three microRNAs.

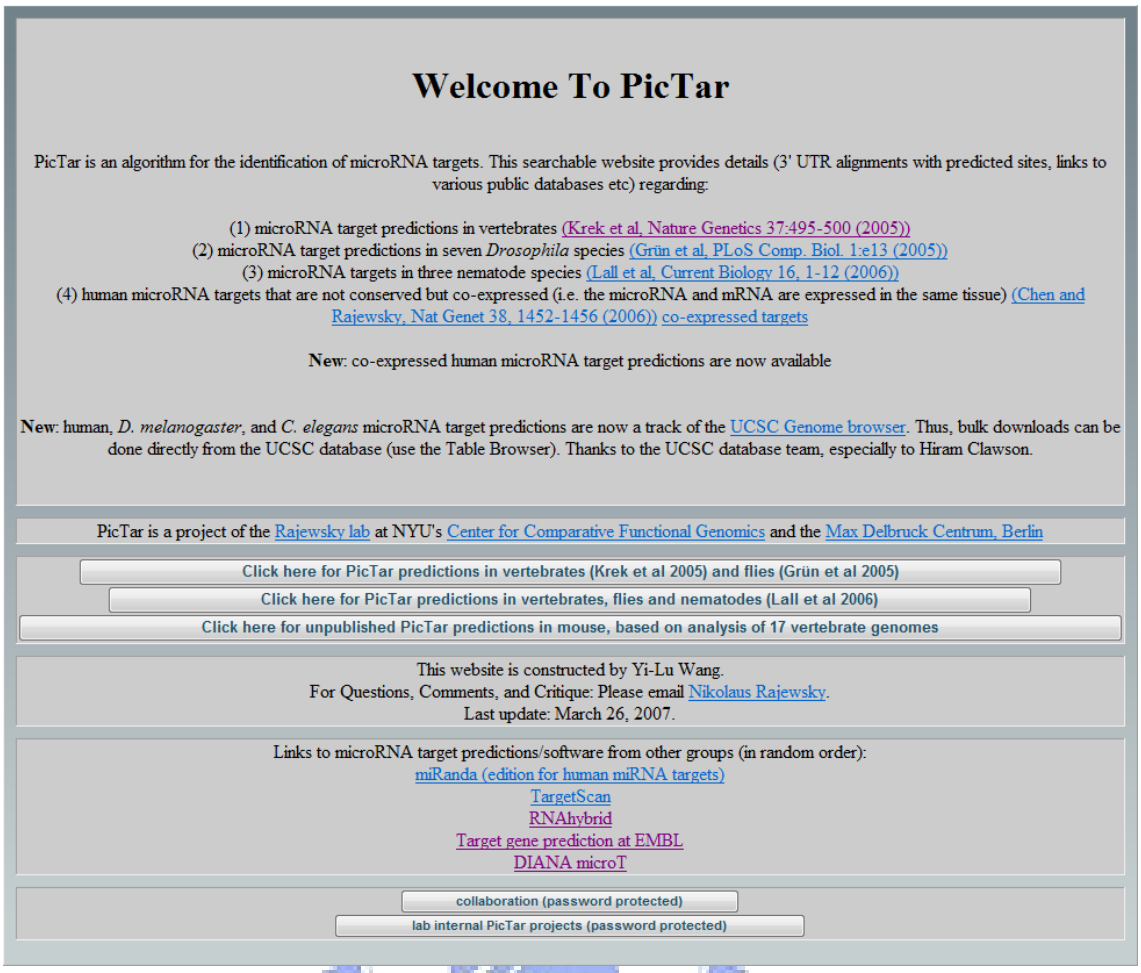


Figure 4.5 PicTar web interface.

4.3 Meterials and Method

To predict the target genes ofcancer-related miRNAs, we applied miRanda (8) to energetically detect the most probable targets of the miRNA against the 3’-UTR of all genes. The program have been explored the effect of the modified rules for scoring miRNA and target matches: reducing the number of G:U wobbles and increasing the high match scale factor from 2 to 4 at position 2-8 from 5' end of miRNA. The 3’-UTR sequences were obtained from the latest version of the human genome assembly (NCBI36), 19147 of 21227 mRNA transcripts were annotated the 3’-UTR region from BioMart system in Ensembl database (87). We adjusted the parameters of miRanda program from the real cases (49,52) and to fit the "seed match" which is the position 2-8

from 5' end of miRNA known to interact with perfect complementarity to its target genes(35).

The tumor microarray datasets are used (See Table 2.2) to identify the miRNA targets with contrary expression of miRNAs and their targets. Figure 4.6 shows the two steps of miRNA target prediction: Step 1 is to computational analysis and predicts the putative miRNA targets. Step 2 is to combine the tumor microarray and to find the target genes whish are expressed contrary to the miRNAs.

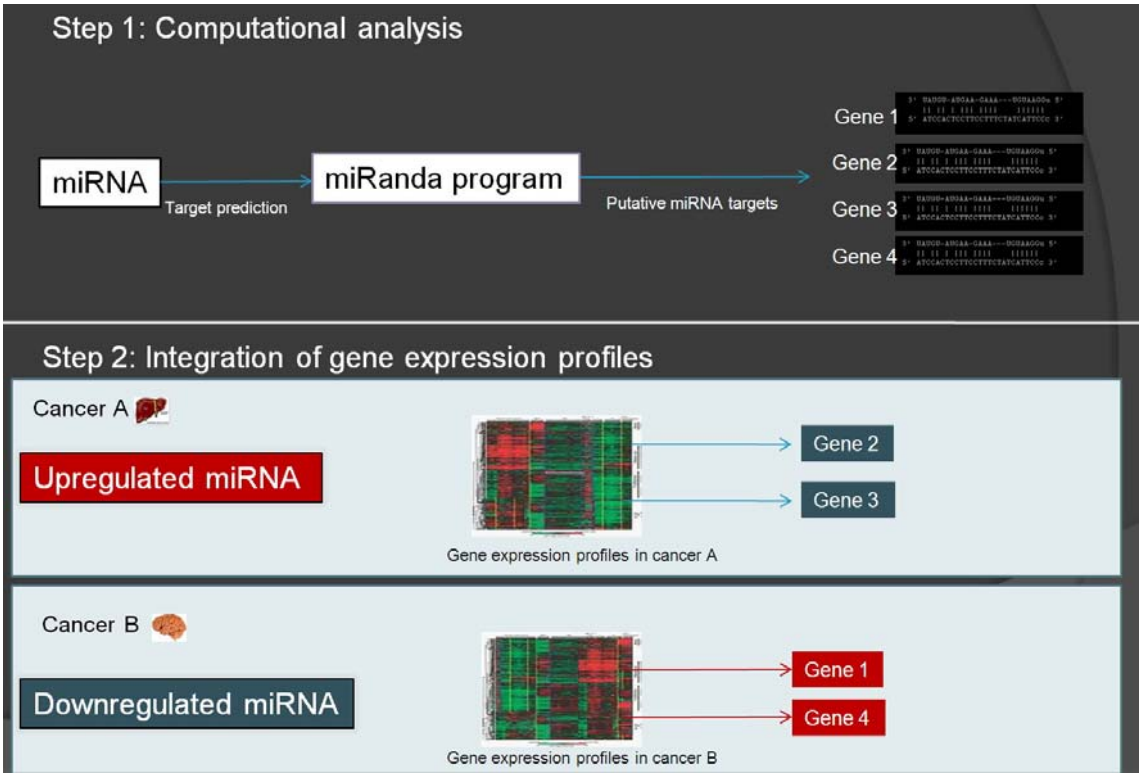


Figure 4.6 Refers to the gene expression profiles for miRNA target prediction.

4.4 Results

4.4.1 Prediction of the miRNA targets with high confidence

The parameters of miRanda program were adjusted to fit the real cases (the scores are more than 140 and minimum free energy (MFE) values are lower than 10 kcal/mol will be chosen). Most known miRNA targets can be detected with the parameters of miRanda,

Table 4.1 The list of the known miRNA targets and tested by miRanda.

miRNAs	target gene names	Reference	MFE (miRanda)	Score (miRanda)	Ensembl ID
hsa-let-7	NRAS	Johnson et al., 2005	-14.33	149	ENSG00000168638
hsa-let-7b	ACTG1	Kiriakidou et al., 2004	-24.2	143	ENSG00000184009
hsa-let-7b	RPIA	Kiriakidou et al., 2004	-23.47	136	ENSG00000153574
hsa-let-7e	SMC1L1	Kiriakidou et al., 2004	-23.33	169	ENSG00000072501
hsa-let-7e	EIF3S1	Kiriakidou et al., 2005	-18.9	147	ENSG00000104131
hsa-miR-101	N-MYC (NMI)	Lewis et al., 2003	-13.07	124	ENSG00000123609
hsa-miR-101	ENX-1 (EZH2)	Lewis et al., 2003	-22.41	166	ENSG00000106462
hsa-miR-103	FBXW1B (FBXW11)	Kiriakidou et al., 2004	-21.8	140	ENSG00000072803
hsa-miR-1	G6PD	Lewis et al., 2003	-14.4	153	ENSG00000160211
hsa-miR-1	BDNF	Lewis et al., 2003	-12.54	154	ENSG00000176697
hsa-miR-130a	MCSF (CSF1)	Lewis et al., 2003	-22.33	157	ENSG00000184371
hsa-miR-141	CLOCK	Kiriakidou et al., 2004	-17.89	152	ENSG00000134852
hsa-miR-145	FLJ21308	Kiriakidou et al., 2004	-24.56	155	ENSG00000151883

hsa-miR-15a	C22orf5	Kiriakidou et al., 2004	-23.69	167 ENSG00000198792
hsa-miR-15a	DMTF1	Kiriakidou et al., 2004	-21.98	175 ENSG00000135164
hsa-miR-23a	SDF-1 (CXCL12)	Lewis et al., 2003	-21.07	126 ENSG00000107562
hsa-miR-23a	BRN-3b (POU4F2)	Lewis et al., 2003	-13.7	149 ENSG00000151615
hsa-miR-24	MAPK14	Kiriakidou et al., 2004	-29.28	187 ENSG00000112062
hsa-miR-26a	SMAD1	Lewis et al., 2003	-13.96	149 ENSG00000170365
hsa-miR-34a	Delta1 (DLL1)	Lewis et al., 2003	-29.1	180 ENSG00000198719
hsa-miR-34a	HN1	Lewis et al., 2003	-18.69	151 ENSG00000189159
hsa-miR-103	GPD1	Kiriakidou et al., 2004	-23.3	158 ENSG00000167588
hsa-miR-141	KLF5	Kiriakidou et al., 2004	-22.22	169 ENSG00000102554
hsa-miR-141	STE20(STK3)	Kiriakidou et al., 2004	-21.35	164 ENSG00000104375
hsa-miR-145	TMOD3	Kiriakidou et al., 2004	-24.9	144 ENSG00000138594
hsa-miR-196a-1	HOXB8	Yekta et al., 2004	-36.01	186 ENSG00000120068
hsa-miR-375	MTPN	Poy et al., 2004	-12.53	120 ENSG00000105887

The "-" indicates that miRanda does not detect the miRNA target.

4.4.2 Case Study of miR-122 target prediction

In this study, we predicted miR-122 targets and using experimental verification to test our prediction rate. Several microarray datasets were used in the target prediction, human HCC arrays (GSE6222) and microarrays of mouse livers treated with A\antagomiR122 (GSM77216) or scramble antagomiR122 (GSM77217) (34). The predicted target genes with the scores more than 150 and minimum free energy (MFE) values less than -10 kcal/mol were chosen for further study. Figure 4.7 shows that the flowchart of miR-122 target prediction by using human and mouse microarrays. 45

miR-122 target genes are identified as Table 4.2.

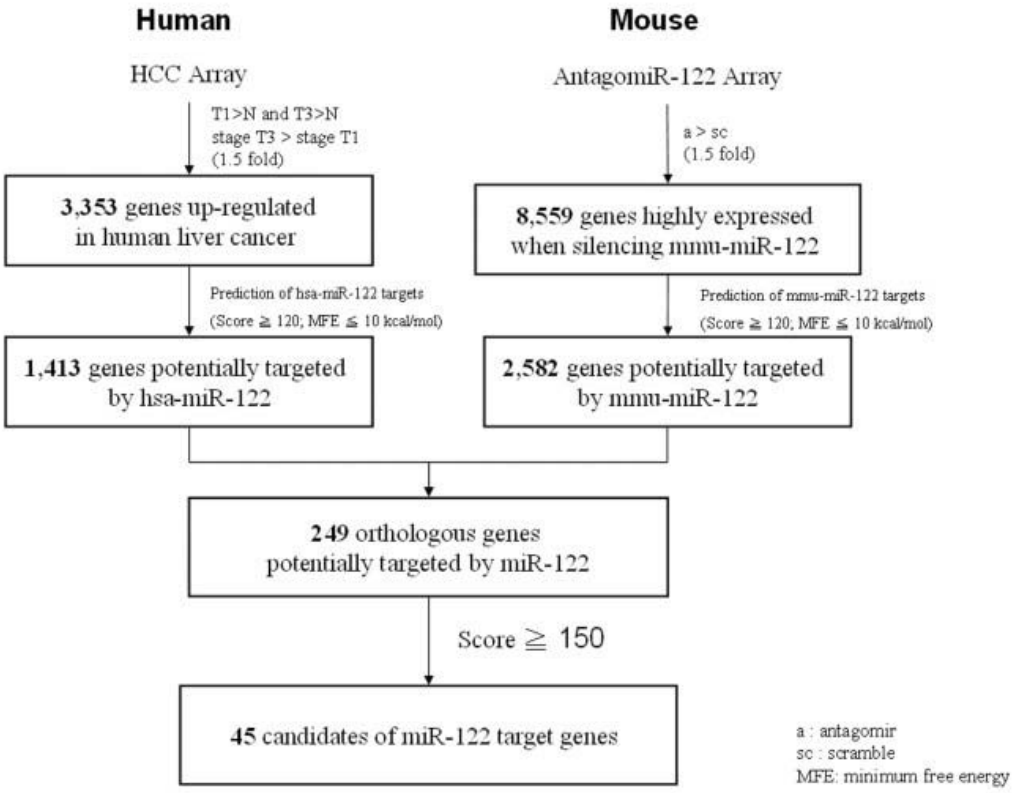


Figure 4.7 The flowchart of miR-122 target prediction in HCC.

32 genes are successful verified by 3' UTR luciferase reporter assay (88) (See Table 4.3) and 89 % prediction rate in our study. We provide an efficient strategy of miRNA target prediction in cancer study.

Table 4.2 miR-122 target genes in human and mouse.

	Human			Mouse				
Gene Symbol	[a]T3/T1	Sc-H	MFE-H	[b]an/c	Sc-H	MFE-M	Molecular Function	#Binding sites
NUMBL	3.1	181	-20.5	4.5	131	-14.5	Protein binding	8
FOXJ3	1.6	176	-21	1.7	133	-10.6	Transcription factor	3
XPO6	2.2	174	-22.3	1.7	125	-16.4	Nuclear protein transport	2
SLC7A1	2.6	173	-22.4	1.7	173	-23.6	Amino acid transport	4
STX6	1.6	169	-20.9	2.5	135	-17.2	Protein transport	3
AP3M2	1.6	168	-19.9	3.0	129	-15.9	Protein trafficking	3
G6PC3	2.1	167	-18.4	6.1	158	-13.1	Glucose-6-phospgatase	1
GALNT10	1.6	167	-20.4	1.9	152	-20.5	Calcium binding	6
ARHGAP19	1.7	166	-21.5	1.7	143	-17.2	GTPase	4
RIPK5	1.8	166	-18	1.6	150	-16.6	Kinase activity	4
TPD52L2	3.0	166	-23.8	16.3	140	-16.5	Cell proliferation	6
AKT3	2.0	165	-18.7	2.1	162	-15.6	Cell proliferation, differentiation, apoptosis	2
FUNDC2	2.5	165	-18	2.1	161	-15.8	HCV core binding protein	4
MAPK11	3.2	165	-20.2	1.6	123	-19.1	MAPK activity	2
ALS2CR13	2.0	162	-17.3	6.7	163	-17	Unknown	3
BACH2	1.8	162	-15	2.7	140	-13.5	Transcription factor	3
ATP11A	2.0	161	-12.9	1.9	162	-18.1	Transport ions	7
SORT1	2.2	161	-16.2	2.0	154	-15.3	Cell differentiation	2
ATP1A2	3.5	160	-17	1.5	132	-11.7	Ion concentration balance	7
ADAM17	1.7	159	-19	2.0	122	-17	Cell cell interaction	1
DUSP2	1.8	159	-13.4	1.6	123	-11.1	MAPK phosphatase	2
OSMR	2.5	159	-15.8	2.9	142	-15.5	Cell proliferation	4
RABIF	1.7	159	-20.1	2.4	139	-15.1	Small GPT regulator activity	2
PALM	2.3	156	-15.9	16.1	161	-18.7	Cell mobility and cell shape	3
SPRED1	2.2	156	-21.2	1.6	145	-19.8	Activate MAPK kinase	1
AACS	1.8	155	-13.8	1.8	163	-17.8	Isoprenoid biogenesis	2
TBX19	2.1	155	-17.2	4.3	131	-13.1	Transcription factor	5
UBAP2	2.2	155	-18.8	2.7	131	-23.3	Ubiquitin associated protein	2
EGLN3	3.9	154	-17	1.6	137	-15.7	Apoptosis	1
NCAM1	1.9	154	-15.2	3.8	154	-15.2	Cell differentiation	4
MECP2	2.2	153	-12.1	3.1	160	-26.2	Transcription	6
CS	1.6	152	-12.6	1.8	148	-11.9	Catalyze synthesis of citrate	3
FOXP1	1.6	152	-10.8	1.8	167	-17.7	Transcription factor	1
RAB11FIP1	1.5	152	-20.1	1.9	133	-13.3	Protein transport	4
RAB6B	1.5	152	-13.3	1.6	164	-19.6	GTPase	5
TRIB1	2.1	152	-17.4	1.8	134	-12.8	Kinase activity	4
TTYH3	1.8	152	-17.1	1.9	160	-16.6	Chlorode anion channel	2
ALDOA	3.7	151	-13.2	2.2	157	-16.7	Amino acid transport	2
ANXA11	2.0	151	-18.6	4.3	139	-17.2	Calcium binding	7
CLDN18	1.7	151	-18.5	1.9	130	-15.1	Cell and cell adhesion	4
ENTPD4	2.0	151	-18.5	2.0	131	-13.5	Calcium binding	2
NFATC2IP	2.7	151	-19.9	2.2	135	-15.2	Protein modification	4
ANK2	2.1	150	-12	3.1	137	-17.9	Cell proliferation	5
MEP1A	4.3	150	-17.8	1.6	132	-12.8	Peptidase	3
NFATC1	1.8	150	-13.1	2.3	148	-14.9	Transcription factor	3
SLC7A11	9.5	150	-19.5	1.6	164	-14.4	Amino acid transport	4

Table 4.3 Verification of *miR122* target genes using the 3'UTR reporter assay. *miR122*-directed repression of luciferase reporter genes bearing 3' UTR fragments of the candidate target genes was measured in 293T cells overexpressing wild-type *miR122* or mutant *miR122* (*miR-122M*, mutations in the seed region). Sc-H: prediction score of *miR122* "seed match" for human genes. Non-target genes: underlined and in bold-face (88).

Gene Symbol	Molecular Function	Sc-H	% of repression in miR-122-directed 3'UTR reporter assay (*p* value)	% of repression in miR122M-directed 3'UTR reporter assay (*p* value)
NUMBL	Protein binding	181	29**	40*
FOXJ3	Transcription factor	176	54**	0
XPO6	Nuclear protein transport	174	29*	15^{Δ}
SLC7A1	Amino acid transport	173	25***	0
STX6	Protein transport	169	35^{Δ}	ND
AP3M2	Protein trafficking	168	39***	ND
G6PC3	Glucose-6-phospgatase	167	22**	0
GALNT10	Calcium binding	167	19*	0
ARHGAP19	GTPase	166	↑17***	ND
RIPK5	Kinase activity	166	19*	5^{Δ}
TPD52L2	Cell proliferation	166	36**	5^{Δ}
AKT3	Cell proliferation, apoptosis	165	38*	0
FUNDC2	HCV core binding protein	165	27**	5^{Δ}
MAPK11	MAPK activity	165	22***	ND
ALS2CR13	Unknown	162	38***	10^{Δ}
SORT1	Cell differentiation	161	↑ 19**	ND
ATP1A2	Ion concentration balance	160	31**	5^{Δ}
ADAM17	Cell-cell interaction	159	28*	0
DUSP2	MAPK phosphatase	159	27**	0
SPRED1	Activate MAPK kinase	156	0	ND
AACS	isoprenoid biogenesis	155	14*	0
TBX19	Transcription factor	155	42*	0
UBAP2	Ubiquitin associated protein	155	32**	0
EGLN3	Apoptosis	154	38**	↑ 30%
NCAM1	Cell differentiation	154	27**	0
MECP2	Transcription	153	35***	0
FOXP1	Transcription factor	152	50***	5^{Δ}
RAB11FIP1	Protein transport	152	49*	5^{Δ}
RAB6B	GTPase	152	26***	5^{Δ}
TRIB1	Kinase activity	152	27**	25**
ALDOA	Amino acid transport	151	36*	0
ANXA11	Calcium binding	151	19**	0
ENTPD4	Calcium binding	151	42*	0
NFATC2IP	Protein modification	151	18**	0
ANK2	Cell proliferation	150	33**	0
SLC7A11	Amino acid transport	150	35**	15**

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.0001$; $^{\Delta}$, not significant; ↑, increased luciferase activity; ND, not done.

4.4.3 **Annotation of miRNA target genes**

All the targets of cancer-related miRNAs were annotated their molecular function by using Gene Ontology (59). Table 4.4 shows an example of oncomiR miR-221, which is overexpressed in brain cancer. The function of miRNA target gene is a important clue to discover the roles of miRNA in oncogenesis.

Table 4.4 miR-221 targets with gene expression profiles in brain cancer.

Gene	Score	EMBL_ID	Name	Expression	Function
BCL2	155	M13994	bcl-2-alpha	D	Cell Cycle Regulators
ERBB3	148	M34309	EGFR (epidermal growth factor receptor/HER3)	D	Growth Factors/Growth Factor Receptors
GFAP	154	S40719	GFAP (glial fibrillary acidic protein)	D	Cytoskeletal Elements
GRIN2A	146	U09002	NMDA receptor 2A (GRIN 2A)	D	Neurotransmitters
IGF1	157	X57025	IGF-I (insulin-like growth factor I)	D	Cytokines/Cytokines Receptors
MAP3K12	144	U07358	zipper protein kinase (zpk)	D	Kinases/Phosphatases
MYCBP	149	D50692	c-myc binding protein (AMY-1)	D	Proto-oncogene/ Oncogene
NEFH	155	X15306	NF-H 1 neurofilament protein	D	Miscellaneous
OLFM1	141	D82343	AMY (amylin)	D	Miscellaneous
PDGFRA	151	M21574	PDGFR-alpha?	D	Growth Factors/Growth Factor Receptors
RIMS3	151	D87074	KIAA0237	D	Unknown
SLC1A2	144	U01824	glutamate/aspartate transporter II	D	Solute Transporters

D = decreased expression in tumor compared to normal

Chapter 5 Discovery regulatory networks of oncomirs involved in different oncogenesis

5.1 Identification of tissue-specific miRNAs

To study the cancer-related miRNAs are with the characters of tissue-specific and abundant in particular tissues. We analyzed the expression of 345 human miRNAs in 40 normal human tissues. The average value of these miRNAs was used to normalize the input data between tissues. The sample input was also normalized by quantitating small nuclear RNAs using the TaqMan MicroRNA Assay Controls. The normalization and hierarchical clustering were reconstructed by software Hierarchical Clustering Explorer v3.5 (83) (Figure 5.1). An unsupervised hierarchical clustering was employed which based on the variation of expression for each miRNA across the tissues examined to explore the correlation between different tissue types.

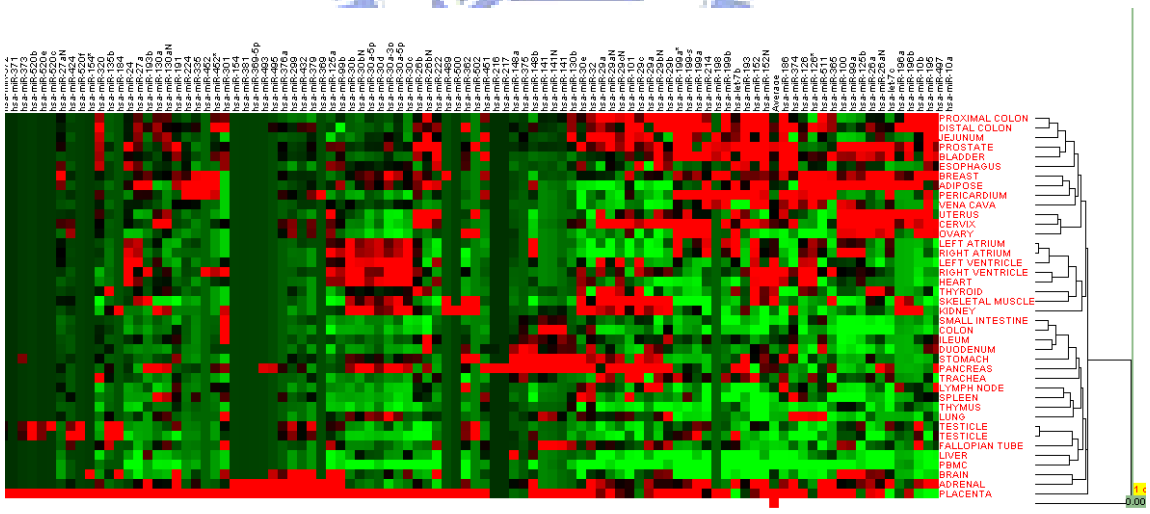


Figure 5.1 Unsupervised hierarchical clustering of the normal human tissues based on the variation of miRNA expression profiles

Tissue-specific miRNAs are as biomarkers which play important roles in differentiation during development and have been tied to the development of specific cancers (89), identifying the tissue-specific miRNAs involved in specific cancers could

provide useful diagnostic information as well as potential treatment targets.

Tissue-specific miRNAs

The tissue-specific expressions of miRNAs were defined as those miRNAs whose expressions are sufficient in specific tissues (Table 2.5) or 20-fold or higher in the specific tissues compared with the mean of the other tissues (Table 2.6). Tissues expressing the most specific miRNAs include brain, heart, skeletal muscle, and pancreas.

Table 5.1 Tissue-specific miRNAs are detectable at sufficient levels in the specific tissue but are undetectable in the rest of tissues.

miRNAs	Tissues
miR-122a	liver
miR-216	pancreas
miR-129, miR-219, miR-330	brain
miR-302a, miR-302c, miR-367	heart
miR-516, miR-518, miR-519d	placenta
miR-215	colon

Table 5.2 miRNA expressions are >= 20-fold higher in the specific tissues compared with the mean of the other tissues.

miRNAs	Tissues
miR-1	heart, skeletal muscle
miR-149, miR-153,miR-181a, miR-221	brain
miR-203	esophagus
miR-204	kidney
miR-206	skeletal muscle
miR-215	small intestine
miR-302d	heart
miR-371, miR-372	placenta
miR-375	pancreas

Tissue-enriched miRNAs

We selected tissue-enriched miRNAs depend on the expression is >= 5-fold higher in the specific tissue compared with the mean of the other tissues, 23 miRNAs were selected in Table 5.3.

Table 5.3 miRNA expression is >= 5-fold higher in the specific tissue compared with the mean of the other tissues.

miRNAs	**Tissues**
miR-205, miR-145	esophagus
let-7i	thyroid
miR-148a	liver, pancreas
miR-30e*	liver
miR-148b	pancreas
miR-150	spleen
miR-190	kidney
miR-192, miR-194	colon
miR-223	lung, spleen
miR-205	breast, prostate
miR-34a	prostate
miR-142	lung, spleen
miR-196b	cervix
miR-451	bladder
miR-202	testicle
miR-10b	uterus
miR-150, miR-142*, miR-155, miR-146b	lymph node
miR-212	stomach

Most differentially expressed miRNAs in each tissue

We normalized the miRNA expression profiles and transformed the distribution into a standard normal distribution with Z-score which n terms of the number of standard deviations from the mean value. The equation is as follows:

$$Z_{ij} = \frac{(Xij - \overline{Xi})}{Si}$$

The Z-score of *i* th miRNA in *j* th tissue was denoted Z_{ij}. There are 375 samples of miRNA expression data $i=\{1,2,\ldots,375\}$ among 40 tissues $j=\{1,2,\ldots,40\}$. The X*ij* denotes the *i* th miRNA expression data in *j* th tissue, and $\overline{Xi}$ is the mean of *i* th miRNA data in all tissues and S*i* is the standard deviation. The most differentially expressed miRNAs were identified with *p*-value <=0.05 (See Appendix II and Figure 5.2).

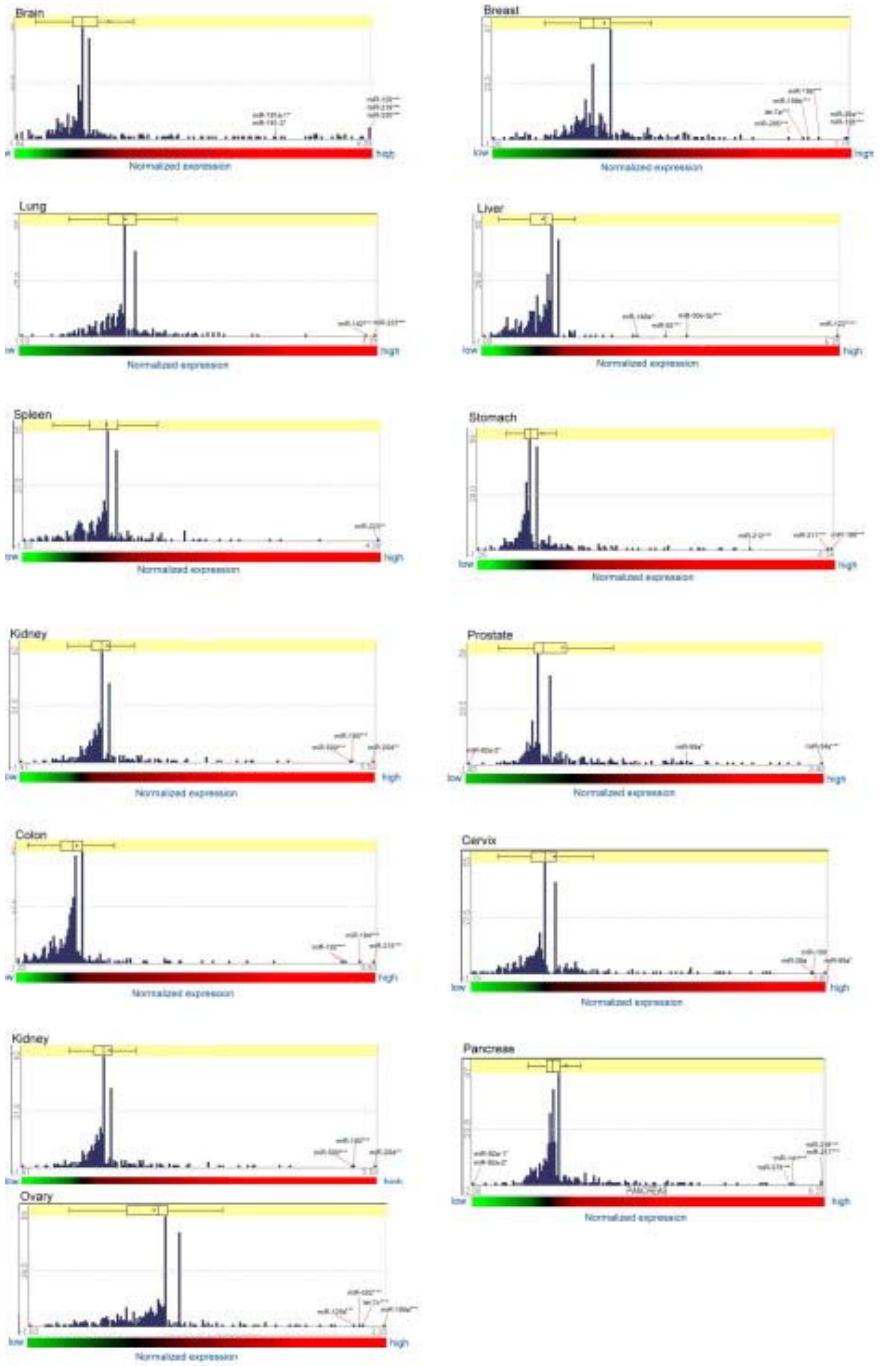


Figure 5.2 Normalization of each miRNA expressions across 40 tissues, p value <0.05 was considered statistically significant. *p<0.05; **p<0.01; ***p<0.001

5.2 Potential oncomir candidates

We combined the miRNA expression profile to identified human cancer-related miRNAs which are highly tissue-specifically expressed or differentially expressed in particular tissues (Table 5.4). Those miRNAs may play essential roles to maintain stability of human cellular mechanism in particular tissues.

Table 5.4 Human cancer-related miRNAs which are differentially expressed in certain tissues.

Cancer	miRNA	Up/down	Ref
Hepatocellular carcinoma	hsa-mir-122	Down	Kutay et al., 2006
Hepatocellular carcinoma	hsa-mir-30e*	Down	Budhu et al., 2008
Hepatocellular carcinoma	hsa-miR-92a-1	Down	Meng et al., 2007
Hepatocellular carcinoma	hsa-mir-148a	Down	Budhu et al., 2008
Breast cancer	hsa-miR-126	Down	Tavazoie et al., 2008
Breast cancer	hsa-miR-205	Down	Volinia et al., 2006
Colon cancer	hsa-miR-192	Down	Braun et al., 2008
Colon cancer	hsa-miR-194-2	Down	Braun et al., 2008
Colon cancer	hsa-miR-194-1	Down	Braun et al., 2008
Colon cancer	hsa-miR-215	Down	Braun et al., 2008
Brain cancer	hsa-miR-181a-1	Down	Ciafre et al., 2005
Brain cancer	hsa-miR-181a-2	Down	Ciafre et al., 2005
Brain cancer	hsa-miR-221	Up	Ciafre et al., 2005
Prostate cancer	hsa-miR-34a	Down	Ambs et al., 2008
Prostate cancer	hsa-miR-92a-2	Up	Volinia et al., 2006
Prostate cancer	hsa-miR-99a	Down	Porkka et al., 2007
Pancreas cancer	hsa-miR-375	Down	Lee et al., 2006
Pancreas cancer	hsa-miR-148a	Down	Bloomston et al., 2007
Pancreas cancer	hsa-miR-148b	Down	Bloomston et al., 2007
Pancreas cancer	hsa-miR-92a-1	Up	Volinia et al., 2006
Pancreas cancer	hsa-miR-92a-2	Up	Volinia et al., 2006
Ovarian cancer	hsa-miR-125b-1	Down	Iorio et al., 2008
Lung cancer	hsa-miR-126	Down	Volinia et al., 2006
Stomach cancer	hsa-miR-212	Down	Volinia et al., 2006

To understand the regulatory mechanism of miRNAs in cancer study, we identified complex interactions among transcription factors, miRNAs and their target genes. All the cis-regulatory elements and miRNA target genes had been annotated their molecular function by using Gene Ontology (59). Related information of gene function in different cancer are referred to the literatures.

5.3 Results

We identified 20 oncomirs of eight cancer types from 205 cancer-related miRNAs. Those oncomirs are most important than other cancer-related miRNAs because their features related to tissue specificity and can directly regulate downstream genes to result oncogenesis.

The regulatory networks were constructed by predicting the cis-regulatory elements and miRNA targets. Some experimental validations can support our predictions. The dotted red lines are our predictions in the networks and solid red lines indicate that putative regulatory relationships are verified by experimental data in published references. The solid black lines are the accomplished works in other studies and we integrated them in our networks.

5.3.1 Hepatocellular carcinoma

Four oncomirs: miR-122, miR-148a, miR-30e* and miR-92 are downregulated in hepatocellular carcinoma (HCC). Figure 5.3 shows the regulatory networks of those oncomirs. Several cis-regulatory elements of oncomirs are down-expressed (T1 stage> T3 stage) and may result their targets dowregulated. The gene NR3C1 is a receptor for glucocorticoids that can act as both a transcription factor and as a regulator of other transcription factors. The glucocorticoids can attenuate the response of estrogen which is a risk factor for breast cancer (90). Besides, it controls all the oncomiRs in HCC in our prediction and the dowregulated of NR3C1 may infect the expression of oncomiRs. Recent study indicates TCF7L2 which had been thought to boost malignant cell growth instead as a transcriptional repressor that restricts colorectal cancer (CRC) cell growth (13). It is one of putative target of transcription factor and tumor suppressor TP53 in TRANSFAC database.

The overexpression of transcription factors and nuclear factor of activated T-cells c1 (NFATC1) will activate its target gene autotaxin (ATX). ATX is known to stimulate migration of tumor cells (91).

One of the miR-122 targets in our prediction, ADAM17 (a disintegrin and metalloprotease 17) have been experimentally verified, miR-122 can silence ADAM17 and resulted reduction of migration, invasion of tumors in the livers of nude mice (88). Three oncomirs: miR-122, miR-148 and miR-30e*, can silence transcription factor SOX4 and miR-122 can also silence NRP1 which is one of SOX4 targets and plays an important role in tumor metastasis of mouse livers (88).

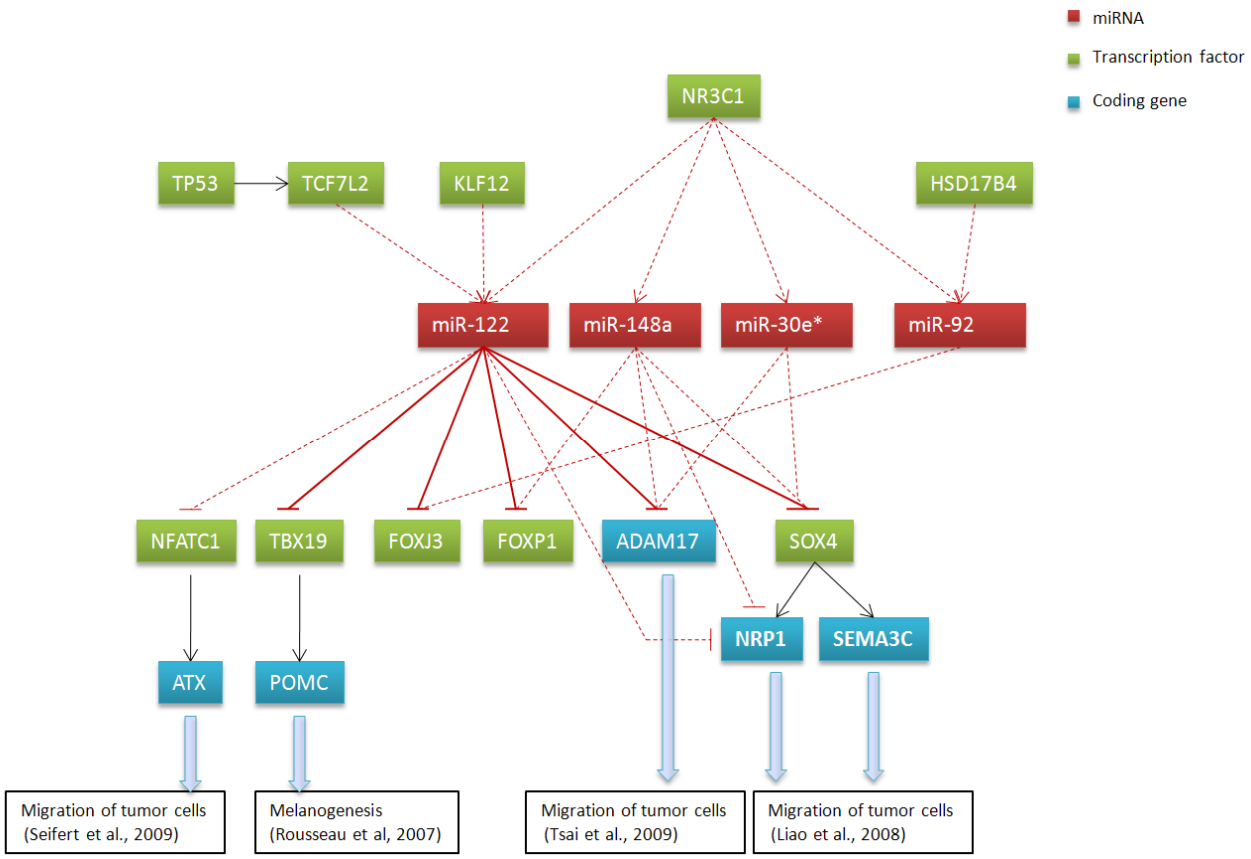


Figure 5.3 Oncomir regulatory network in HCC, all the oncomiRs and their cis-regulatory elements are downregulated and oncomir target genes are overexpressed in microarray data.

5.3.2 **Brain cancer**

Figure 5.4 shows the regulatory network in brain cancer, miR-181a-1 and

miR-181a-2 are down regulated and miR-221 is upregulated in this study. The human GSTP1 gene is frequently overexpressed in many human cancers and the expression increases with tumor progression, the presence of a cAMP response element (CRE) in the 5'-region of the human GSTP1 gene, raising the possibility that the cAMP signaling pathway to active CRE (Ser133) binding protein-1 (CREB1) (92). CREB1 is one of the cis-regulatory elements of miR-221 in our prediction. Overexpression of miR-221 was shown to cause downregulation of cyclin-dependent kinase (CDK) inhibitors CDKN1C/p57 and promoted cell cycle progression (93).

hsa-mir-181a functions as tumor suppressors in human glioma cells (94). Downexpression of hsa-mir-181a may cause upregulation of oncogenes like CKS1B, FOS and ADM.

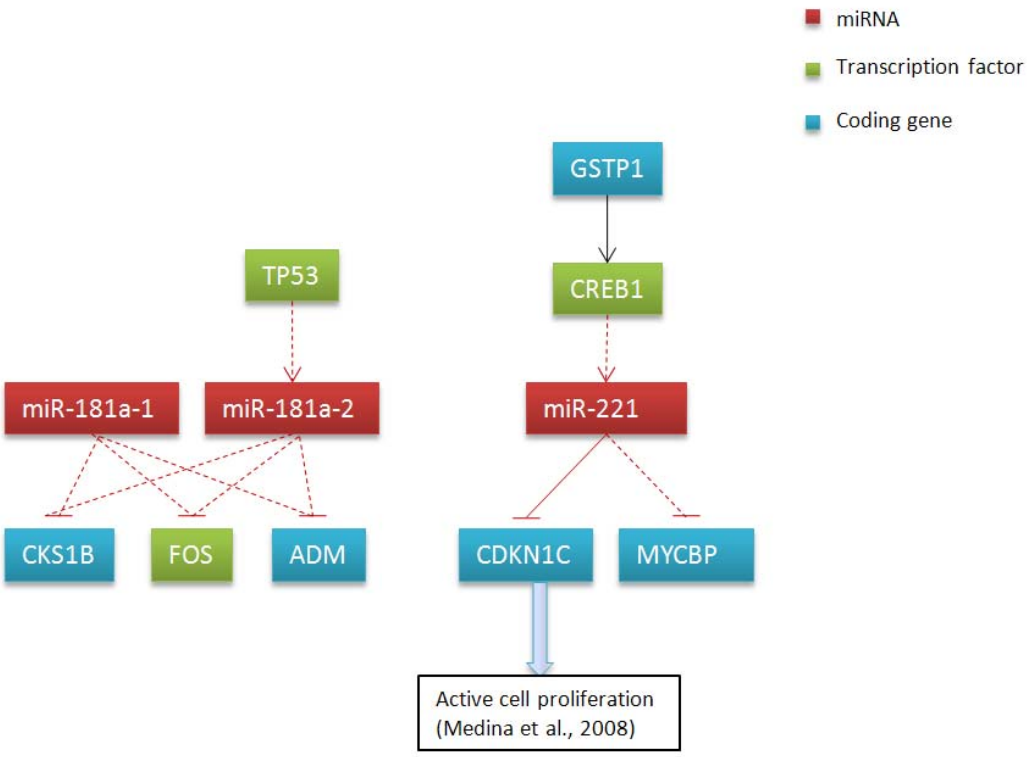


Figure 5.4 miR-181a-1 and -2 are downregulated and miR-221 is upregulated in brain cancer.

5.3.3 Colon cancer

The miR-200 family of miRNAs plays a major role in specifying the epithelial phenotype by preventing expression of the transcription repressors, ZEB1(95). ZEB1 is a putative cis-regulatory element of miR-215, miR-194-1, miR-194-2 and miR-192 (Figure 5.5). ZEB1 and their target oncomirs are downregulated in colon cancer.

CD47 and MYLK are common targets of 4 oncomirs in colon cancer. Myosin light chain kinase 2, skeletal muscle (MYLK2) encodes a calcium/calmodulin-dependent serine/threonine kinase. In a recent study, MYLK2 gene was somatically mutated in colorectal carcinomas. The aim of this study was to explore the possibility that other common human carcinomas besides colorectal carcinomas harbored MYLK2 mutations in the kinase domain. (96). MYLK2 is also specifically related to prostate carcinoma (97). This enzyme catalyzes the phosphorylation of a specific serine residue on the 20 kD light chain of myosin II (MCL20), consequently regulating the actin-myosin II interaction. since MLC20 phosphorylation is necessary for cell motility, MYLK inhibition blocks cancer cell invasion and adhesion in vitro. As a result, some reports described the use of MYLK inhibitors as anti-cancer agents since they prevent cancer cells migration (97).

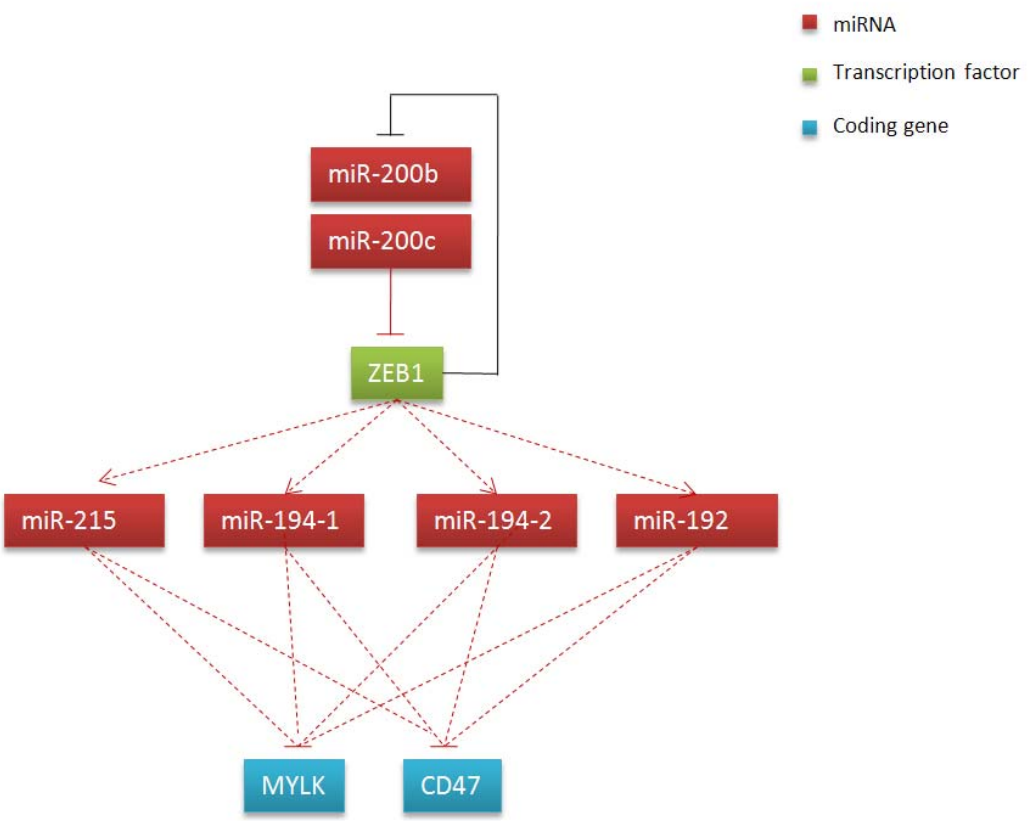


Figure 5.5 Oncomirs miR-215,mIR-194-1,miR-194-2 and miR-192 are downregulated in colon cancer, the miR-200 family may be involved in their upstream regulation.

5.3.4 **Breast cancer**

TFAP4 is the cis-regulatory element of miR-205 and miR-126 which are downregulated in breast cancer (Figure 5.6). The interesting is that the oncogene c-FOS or the name FOS in HGNC symbol was downregulated in the network of breast cancer. It plays a role in a variety of physiological functions including cell proliferation and differentiation, but in recent studies have raised the idea that c-Fos may also have tumour-suppressor activity and might have a function in apoptosis(98).

In recent studies, the team led by Joan Massagué (23) found that miR-335, miR-126, and miR-206 are metastasis-suppressor miRNAs. To identify these miRNAs, they compared miRNA expression of the metastatic nodules versus the unselected breast cancer parental cells. These miRNAs were consistently downregulated in metastatic

tumors. Moreover, the authors found that restoring the expression of miR-335, miR-126, and miR-106 significantly decreased the number of metastatic foci. The cytometry analysis from Zhang et al. (99) showed that mir-126 targeted IRS-1 and inhibited cell cycle progression from G1/G0 to S. RPS6KB1 is amplified and overexpressed in 10-30% of primary breast cancers and breast cancer cell lines (100-102). This gene is a serine/threonine kinase and plays a crucial role in control of cell cycle, growth and survival.

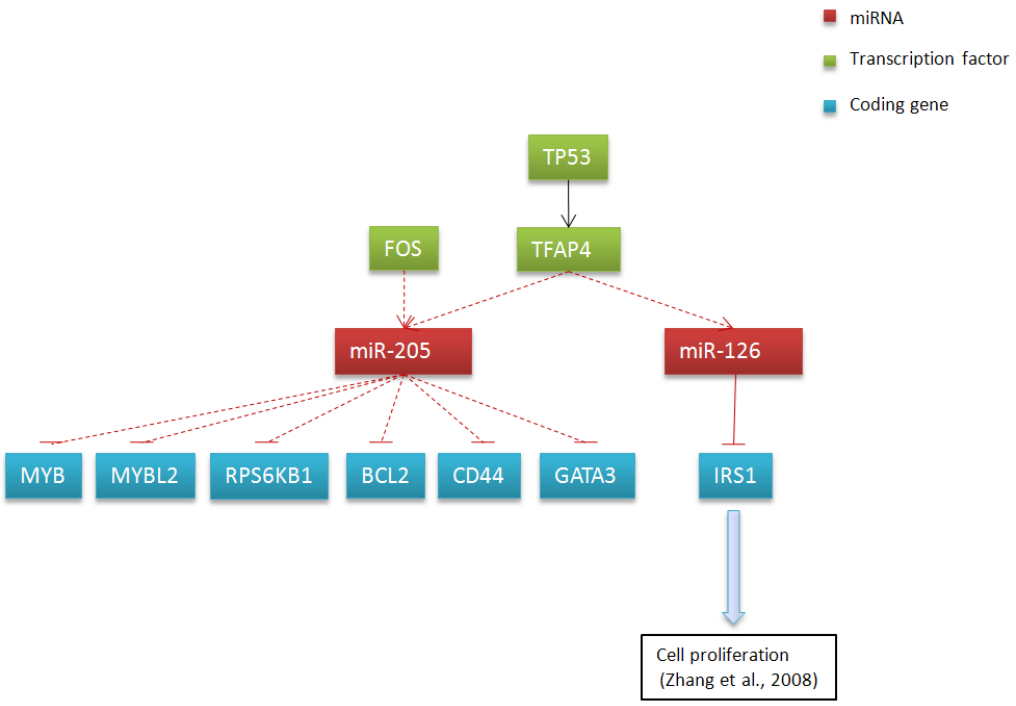


Figure 5.6 miR-205 and miR-126 are downregulated in breast cancer.

5.3.5 **Lung cancer**

The oncomir miR-142 was upregulated in lung cancer. The regulatory network is shown in Figure 5.7. TFAP2A and CREB1 can activate miR-142 expression in our prediction. Recent studies have shown that AP-2alpha (TFAP2A) and AP-2gamma (TFAP2C) have tumor-suppressive activity in breast cancer, melanoma, and prostate cancer cells. Decreased expression of TFAP2A or TFAP2A has been found in these

tumor cells and is associated with disease progression and the metastatic capabilities of the tumors (103). But in lung cancer, upregulated expression of TFAP2A activates human telomerase reverse transcriptase (hTERT) which is hallmark of tumorigenesis in lung cancer (104).

The miR-142 target, LATS2, is one of LATS tumour suppressor family which plays an important role in the control of tumour development and cell cycles (105). LATS1 and LATS2 are functionally conserved and regulate the cell cycle progression and apoptosis. LATS1 is implicated in the regulation of the cell cycle at the G2/M and LATS2 at the G1/S phases (106).

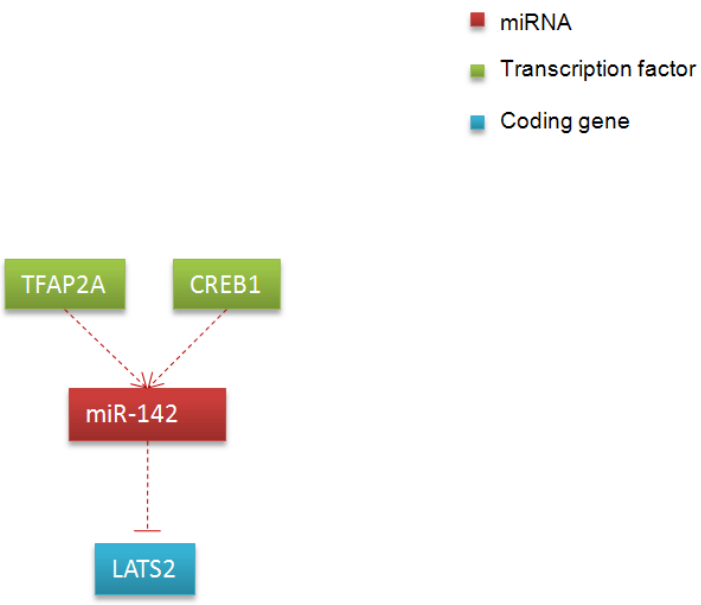


Figure 5.7 miR-142 is upregulated In lung cancer.

5.3.6 Ovarian cancer

FOS (c-Fos) is the putative cis-regulatory element of oncomir miR-125b in ovarian cancer. Mahner found that loss of FOS expression is associated with tumour progression in ovarian carcinoma and that FOS may be a prognostic factor (107). The downregulation of miR-125b may cause upregulation of oncogenes like BCL2, VEGFA

and MUC1 in our prediction (Figure 5.8).

The 3 potential miR-125b targets are involved oncogenesis. BCL2 (for B cell lymphoma gene-2) proteins are associated with membranes and membrane activity. The BCL2 protein is a part of a complex system of signaling that controls apoptosis. Bufalo et al. shows that overexpression of BCL2 enhances the metastatic potential of a human breast cancer (94). Vascular endothelial growth factor A (VEGFA) is critical to angiogenesis. Evidence shows that VEGFA stimulates angiogenesis in tumor growth and mediates neuroprotection to prevent an apoptotic cell death (108). Mucin core protein 1 (MUC1) is associated with invasive growth of neoplasms (109). It is emphasized in most cases of carcinoma and high expression of MUC1 is closely associated with cancer progression and metastasis (110).

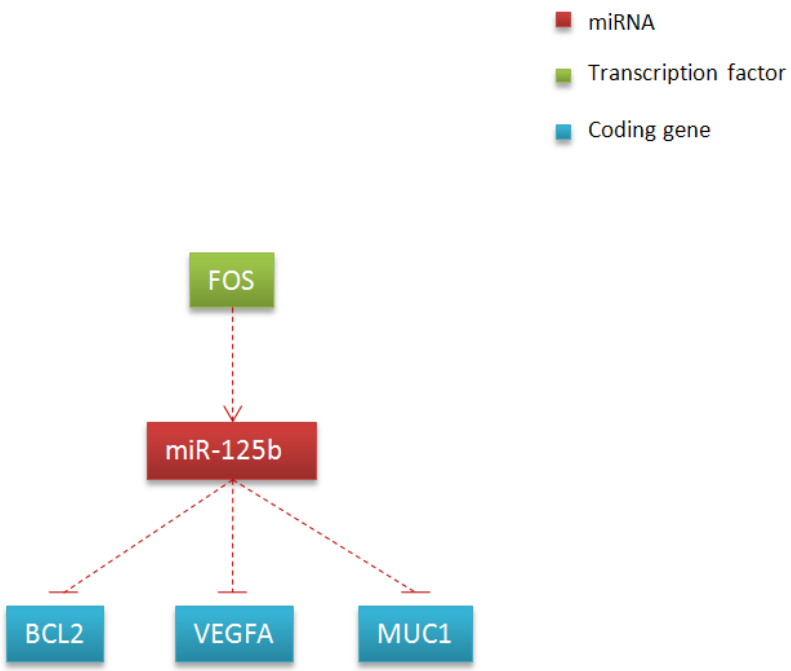


Figure 5.8 miR-125b is downregulated in ovarian cancer.

5.3.7 **Prostate cancer**

In prostate cancer, miR34a and miR99a are down regulated and miR-92a is upregulated (Figure 5.9). One of the cis-regulatory element of miR-34a and miR-99a, JUNB is an essential upstream regulator of p16 which is a tumor suppressor protein (111), and contributes to maintain cell senescence that blocks malignant transformation. Gene expression profiling showed the cis-regulatory element of miR-92a, HSD17B4, to be significantly overexpressed in prostate cancer compared to matched-benign epithelium (112). This gene also overexpressed in metastatic samples compared to primary tumor samples (69).

Gene IGF1R, the type 1 insulin-like growth factor receptor, is the target of both miR-99a and miR34a. IGF1R is overexpressed in prostate cancer, and mediates proliferation, motility, and survival, silencing of the IGF1R gene enhances sensitivity to DNA-damaging agents in both PTEN wild-type and mutant human prostate cancer (113). E2F transcription factors, including miR-34a target gene E2F3, directly modulate expression of EZH2. Recently, overexpression of the EZH2 gene has been implicated in the development of human prostate cancer (114). Another miR-34a target, DDX17 (p72) may affect the transcriptional activity of ß-catenin in colon cancer cell and promote *c-Myc, cyclin D1, c-jun*, and *fra-1*, all of which are proto-oncogenes (19). DUSP5 is a direct target of p53, it represents a novel mechanism by which p53 might negatively regulate cell-cycle progression (115).

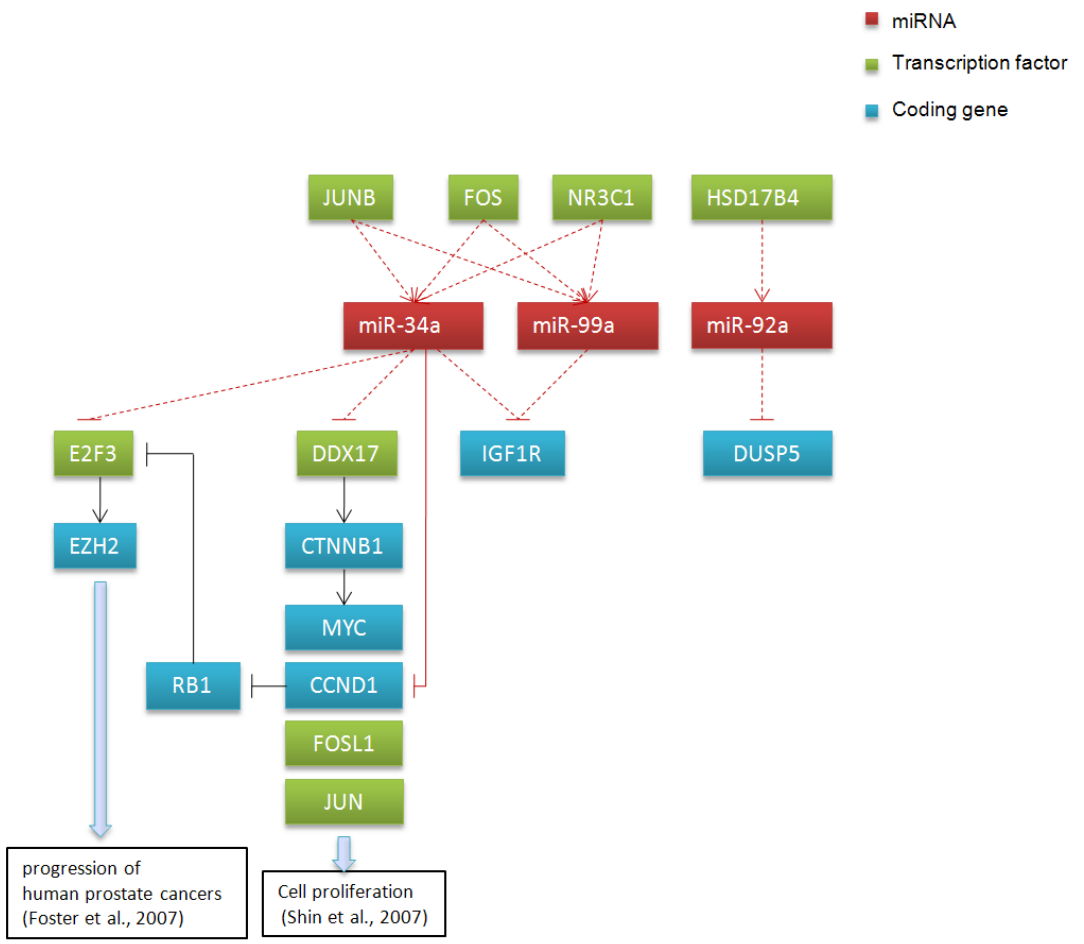


Figure 5.9 miR34a and miR99a are down regulated and miR-92a is upregulated in prostate cancer.

5.3.8 Pancreatic cancer

Figure 5.10 shows the regulatory network in pancreatic cancer, miR-92a is upregulated and miR-375 is downregulated. E2F1 and ETV4 are two cis-regulatory elements of miR-92a. E2F1 is a transcriptional factor that mediates cell cycle progression from G1 to S phase, thereby influencing tumor progression. ETV4 is one of ETS-family factors play major roles in development and cancer, notably as critical targets for extra-cellular signaling pathways, including MAPK-signaling, ETV4 also involves in pancreatic development (116).

Most pancreatic cancers correspond to ductal adenocarcinoma (DAC), which

develops from epithelium in a multistep process(117), the miR-375 target gene MUC4 is a transmembrane mucin expressed in pancreatic ductal adenocarcinoma (DAC) in contrast to normal pancreas, and is an independent predictor of poor prognosis in patients with invasive DAC (118). Recent study shows transcription factor TFAP2A can repress MUC4 (118) and TFAP2A is also cis-regulatory element of miR-375 gene.

The miR-92 target gene FBN2, a large modular extracellular matrix glycoprotein, is known to be a key component of human elastic fiber. A loss of FBN2 expression due to promoter methylation was recently identified in pancreatic cancer (119).

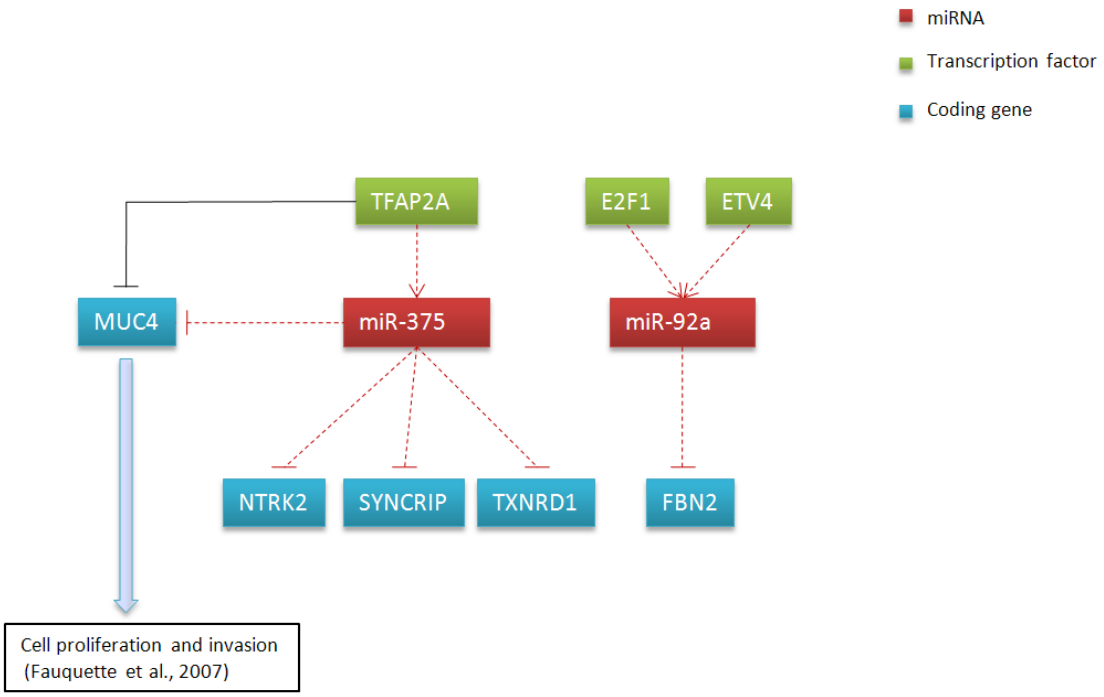


Figure 5.10 Upregulation of miR-92a and downregulation of miR-375 in pancreatic cancer.

Chapter 6 Discussions

Selbach et al. suggest that a miRNA can tune protein synthesis from thousands of genes by direct or indirect effects (12), in our study, we found many miRNAs are uniquely and differentially expressed in certain cancer tissues as compared with normal tissues. Some miRNAs may be directly involved in cancer development by controlling cell differentiation and apoptosis, while others may be involved in cancers by targeting cancer oncogenes and/or tumor suppressors.

We discovered several novel regulatory pathways among miRNAs and coding genes or transcription factors in different cancer types. Some predicted regulatory pathways were verified by experiment in our previous study. In some cases we found the same miRNAs may be involved in different tumorgenesis, for example, miR-92a was both upregulated in prostate and pancreatic cancer but played different roles, in prostate cancer, it inhibits Dual-specificity phosphatase 5 (DUSP5) which can suppress the growth of several types of human cancer cells (115) but in pancreatic cancer, miR-92 regulated epithelial membrane protein 1 (EMP1) which is implicated in promoting tumorigenesis (65).

Some data in the literature support our computational predictions. For example, the recent study indicates that miR-181a functions as tumor suppressors in human brain glioma cells (94). In our study, we found that miR-181a is enriched in normal brain tissue but downregulated in most brain cancer cells. We integrated the gene expression data in brain cancer (120) and predicted c-fos (FOS) and adrenomedullin (ADM) may be the target gene of miR-181a. ADM is highly expressed in brain cancer and it is an angiogenic factor that has also been shown to be a mitogen and a hypoxia survival factor for tumour cells. Overexpression of FOS was found to inhibit cell cycle

progression, stimulated murine hepatocyte cell death and strongly suppressed tumour formation in vivo (121).

FOS plays dual roles in different networks; it can be a oncogene or a tumor suppressor gene. Mahner found that loss of FOS expression is associated with tumour progression in ovarian carcinoma and that FOS may be a prognostic factor (107). In our study, downregulation of FOS is the cis-regulatory element of miR-205 in breast cancer and miR-125b in ovarian cancer and FOS may play tumor suppressor function.

Chapter 7 Conclusion

Cancer remains the No.1 killer in Taiwan, according to the report from Department of Health, Executive Yuan, R.O.C. (Taiwan), 40,305 people were killed in 2007 (Data source: http://www.doh.gov.tw/CHT2006/index_populace.aspx). On average, cancer claims a life in Taiwan about every 13 minutes.

There are five major steps for cancer development: initiation, promotion, malignant conversion, progression, and metastasis. The formation of cancer is the combined interaction of tumor suppressor genes and cancer inducer genes (oncogenes). Although several genes, including oncogenes and tumor suppressor genes, have been identified in human genomes, most of the mechanism of cancer formation is still unknown. Recent studies indicate miRNAs may provide new insight in cancer research. Using up- or downregulated expression of the cancer-related miRNAs is a good approach to study the function of miRNAs in cancer pathogenesis.

We developed a systematic analysis method by to identify the roles of miRNAs in cancers. 205 cancer-related miRNAs were collected in the first step, and then their transcription start sites (TSS) were identified by experimental and computational data. We analyzed cis-regulatory elements from promoter regions as 3-kb segment upstreams of TSS by using Match program with TRANSFAC database. All the putative transcription factors were involved in syn-regulated expression patterns of miRNA and coding genes among 15 tissues. After understanding the potential elements result up or downregulation of miRNAs, we analyzed the cancer-related miRNA functions. The cancer-related miRNA targets were predicted by miRanda program and the expression of miRNA targets were checked by tumor microarray data. We selected the tissue-specific miRNAs which were most differentially expressed in certain tissues. All

the cis-regulatory elements and miRNA target genes were annotated the molecular function by Gene Ontology database. Finally the miRNA regulatory networks were identified in 8 cancers. Understanding of the function of miRNAs can provide the new insights on the mechanism of tumorgenesis, and new biomarkers for cancer.

References

1. Ruvkun, G. (2001) Molecular biology. Glimpses of a tiny RNA world. *Science*, **294**, 797-799.
2. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843-854.
3. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, **34**, D140-144.
4. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. and Hannon, G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231-235.
5. Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M. and Lai, E.C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, **130**, 89-100.
6. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev*, **15**, 2654-2659.
7. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*, **15**, 185-197.
8. Ohrt, T., Mutze, J., Staroske, W., Weinmann, L., Hock, J., Crell, K., Meister, G. and Schwille, P. (2008) Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res*, **36**, 6439-6449.
9. Diederichs, S. and Haber, D.A. (2007) Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*, **131**, 1097-1108.
10. McManus, M.T., Petersen, C.P., Haines, B.B., Chen, J. and Sharp, P.A. (2002) Gene silencing using micro-RNA designed hairpins. *Rna*, **8**, 842-850.
11. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*, **27**, 91-105.
12. Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58-63.
13. Wang, X., Tang, S., Le, S.Y., Lu, R., Rader, J.S., Meyers, C. and Zheng, Z.M.

(2008) Aberrant expression of oncogenic and tumor-suppressive microRNAs in cervical cancer is required for cancer cell growth. *PLoS ONE*, **3**, e2557.

14. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932-935.
15. Cai, X., Hagedorn, C.H. and Cullen, B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna*, **10**, 1957-1966.
16. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res*, **14**, 1902-1910.
17. Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna*, **11**, 241-247.
18. Garzon, R., Fabbri, M., Cimmino, A., Calin, G.A. and Croce, C.M. (2006) MicroRNA expression and function in cancer. *Trends Mol Med*, **12**, 580-587.
19. Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J. (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635-647.
20. Calin, G.A. and Croce, C.M. (2006) MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res*, **66**, 7390-7394.
21. Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T. *et al.* (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, 189-198.
22. Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y. *et al.* (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res*, **64**, 3753-3756.
23. Tavazoie, S.F., Alarcon, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L. and Massague, J. (2008) Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*, **451**, 147-152.
24. Huang, Q., Gumireddy, K., Schrier, M., le Sage, C., Nagel, R., Nair, S., Egan, D.A., Li, A., Huang, G., Klein-Szanto, A.J. *et al.* (2008) The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol*, **10**, 202-210.
25. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) 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*, **99**, 15524-15529.

26. Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M. *et al.* (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*, **102**, 13944-13949.
27. Bottoni, A., Piccin, D., Tagliati, F., Luchin, A., Zatelli, M.C. and degli Uberti, E.C. (2005) miR-15a and miR-16-1 down-regulation in pituitary adenomas. *J Cell Physiol*, **204**, 280-285.
28. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*, **65**, 7065-7070.
29. Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A*, **103**, 2257-2261.
30. Voorhoeve, P.M., le Sage, C., Schrier, M., Gillis, A.J., Stoop, H., Nagel, R., Liu, Y.P., van Duijse, J., Drost, J., Griekspoor, A. *et al.* (2006) A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*, **124**, 1169-1181.
31. Tsujimoto, Y., Finger, L.R., Yunis, J., Nowell, P.C. and Croce, C.M. (1984) Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science*, **226**, 1097-1099.
32. Hutvagner, G., Simard, M.J., Mello, C.C. and Zamore, P.D. (2004) Sequence-specific inhibition of small RNA function. *PLoS Biol*, **2**, E98.
33. Meister, G., Landthaler, M., Dorsett, Y. and Tuschl, T. (2004) Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *Rna*, **10**, 544-550.
34. Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685-689.
35. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787-798.
36. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, **75**, 855-862.
37. Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., Yatabe, Y., Kawahara, K., Sekido, Y. and Takahashi, T. (2005) A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and

enhances cell proliferation. *Cancer Res*, **65**, 9628-9632.

38. Gusev, Y., Schmittgen, T.D., Lerner, M., Postier, R. and Brackett, D. (2007) Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer. *BMC Bioinformatics*, **8 Suppl 7**, S16.
39. Jiang, J., Lee, E.J., Gusev, Y. and Schmittgen, T.D. (2005) Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res*, **33**, 5394-5403.
40. Liang, Y., Ridzon, D., Wong, L. and Chen, C. (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, 166.
41. Thomson, J.M., Parker, J., Perou, C.M. and Hammond, S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nat Methods*, **1**, 47-53.
42. Babak, T., Zhang, W., Morris, Q., Blencowe, B.J. and Hughes, T.R. (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. *Rna*, **10**, 1813-1819.
43. Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y. *et al.* (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res*, **14**, 2486-2494.
44. Liang, R.Q., Li, W., Li, Y., Tan, C.Y., Li, J.X., Jin, Y.X. and Ruan, K.C. (2005) An oligonucleotide microarray for microRNA expression analysis based on labeling RNA with quantum dot and nanogold probe. *Nucleic Acids Res*, **33**, e17.
45. Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A*, **101**, 9740-9744.
46. Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W. and Mourelatos, Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods*, **1**, 155-161.
47. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834-838.
48. Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A. (2007) microRNAs as oncogenes and tumor suppressors. *Dev Biol*, **302**, 1-12.
49. Hsu, P.W., Huang, H.D., Hsu, S.D., Lin, L.Z., Tsou, A.P., Tseng, C.P., Stadler, P.F., Washietl, S. and Hofacker, I.L. (2006) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids*

Res, **34**, D135-139.

50. Hsu, S.D., Chu, C.H., Tsou, A.P., Chen, S.J., Chen, H.C., Hsu, P.W., Wong, Y.H., Chen, Y.H., Chen, G.H. and Huang, H.D. (2008) miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res*, **36**, D165-169.
51. Shahi, P., Loukianiouk, S., Bohne-Lang, A., Kenzelmann, M., Kuffer, S., Maertens, S., Eils, R., Grone, H.J., Gretz, N. and Brors, B. (2006) Argonaute--a database for gene regulation by mammalian microRNAs. *Nucleic Acids Res*, **34**, D115-118.
52. Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna*, **12**, 192-197.
53. Papadopoulos, G.L., Reczko, M., Simossis, V.A., Sethupathy, P. and Hatzigeorgiou, A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res*, **37**, D155-158.
54. Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res*, **31**, 51-54.
55. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol*, **2**, e363.
56. Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res*, **31**, 7280-7301.
57. Edgar, R. and Barrett, T. (2006) NCBI GEO standards and services for microarray data. *Nat Biotechnol*, **24**, 1471-1472.
58. Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F. *et al.* (2005) Ensembl 2005. *Nucleic Acids Res*, **33**, D447-453.
59. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, **32 Database issue**, D258-261.
60. Cotton, R.G., McKusick, V. and Scriver, C.R. (1998) The HUGO Mutation Database Initiative. *Science*, **279**, 10-11.
61. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res*, **32**, D109-111.
62. Kruger, J. and Rehmsmeier, M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res*, **34**, W451-454.
63. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

64. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A*, **101**, 6062-6067.

65. Bredel, M., Bredel, C., Juric, D., Harsh, G.R., Vogel, H., Recht, L.D. and Sikic, B.I. (2005) Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. *Cancer Res*, **65**, 8679-8689.

66. Birkenkamp-Demtroder, K., Christensen, L.L., Olesen, S.H., Frederiksen, C.M., Laiho, P., Aaltonen, L.A., Laurberg, S., Sorensen, F.B., Hagemann, R. and TF, O.R. (2002) Gene expression in colorectal cancer. *Cancer Res*, **62**, 4352-4363.

67. Bertucci, F., Houlgatte, R., Granjeaud, S., Nasser, V., Loriod, B., Beaudoing, E., Hingamp, P., Jacquemier, J., Viens, P., Birnbaum, D. *et al.* (2002) Prognosis of breast cancer and gene expression profiling using DNA arrays. *Ann N Y Acad Sci*, **975**, 217-231.

68. Bignotti, E., Tassi, R.A., Calza, S., Ravaggi, A., Bandiera, E., Rossi, E., Donzelli, C., Pasinetti, B., Pecorelli, S. and Santin, A.D. (2007) Gene expression profile of ovarian serous papillary carcinomas: identification of metastasis-associated genes. *Am J Obstet Gynecol*, **196**, 245 e241-211.

69. Chandran, U.R., Ma, C., Dhir, R., Bisceglia, M., Lyons-Weiler, M., Liang, W., Michalopoulos, G., Becich, M. and Monzon, F.A. (2007) Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer*, **7**, 64.

70. Nakamura, T., Fidler, I.J. and Coombes, K.R. (2007) Gene expression profile of metastatic human pancreatic cancer cells depends on the organ microenvironment. *Cancer Res*, **67**, 139-148.

71. Nakamura, T., Kuwai, T., Kitadai, Y., Sasaki, T., Fan, D., Coombes, K.R., Kim, S.J. and Fidler, I.J. (2007) Zonal heterogeneity for gene expression in human pancreatic carcinoma. *Cancer Res*, **67**, 7597-7604.

72. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*, **101**, 2999-3004.

73. Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S. and Nakai, K. (2008) DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res*, **36**, D97-101.

74. Schmid, C.D., Perier, R., Praz, V. and Bucher, P. (2006) EPD in its twentieth year: towards complete promoter coverage of selected model organisms. *Nucleic*

Acids Res, **34**, D82-85.

75. Abeel, T., Saeys, Y., Bonnet, E., Rouze, P. and Van de Peer, Y. (2008) Generic eukaryotic core promoter prediction using structural features of DNA. *Genome Res*, **18**, 310-323.
76. Reese, M.G. (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem*, **26**, 51-56.
77. Knudsen, S. (1999) Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics*, **15**, 356-361.
78. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, **38**, 626-635.
79. Bentley, D.R. (2006) Whole-genome re-sequencing. *Curr Opin Genet Dev*, **16**, 545-552.
80. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823-837.
81. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. and Young, R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **130**, 77-88.
82. Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J. *et al.* (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, **134**, 521-533.
83. Seo, J., Gordish-Dressman, H. and Hoffman, E.P. (2006) An interactive power analysis tool for microarray hypothesis testing and generation. *Bioinformatics*, **22**, 808-814.
84. Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res*, **29**, 281-283.
85. Kutay, H., Bai, S., Datta, J., Motiwala, T., Pogribny, I., Frankel, W., Jacob, S.T. and Ghoshal, K. (2006) Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. *J Cell Biochem*, **99**, 671-678.
86. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell*, **113**, 25-36.
87. Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y.,

Clarke, L., Coates, G., Cunningham, F., Cutts, T. *et al.* (2007) Ensembl 2007. *Nucleic Acids Res*, **35**, D610-617.

88. Tsai, W.C., Hsu, P.W., Lai, T.C., Chau, G.Y., Lin, C.W., Chen, C.M., Lin, C.D., Liao, Y.L., Wang, J.L., Chau, Y.P. *et al.* (2008) MicroRNA-122, a tumor suppressor microRNA that regulates intrahepatic metastasis of hepatocellular carcinoma. *Hepatology*, **49**, 1571-1582.
89. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401-1414.
90. Gong, H., Jarzynka, M.J., Cole, T.J., Lee, J.H., Wada, T., Zhang, B., Gao, J., Song, W.C., DeFranco, D.B., Cheng, S.Y. *et al.* (2008) Glucocorticoids antagonize estrogens by glucocorticoid receptor-mediated activation of estrogen sulfotransferase. *Cancer Res*, **68**, 7386-7393.
91. Seifert, A., Rau, S., Kullertz, G., Fischer, B. and Santos, A.N. (2009) TCDD induces cell migration via NFATc1/ATX-signaling in MCF-7 cells. *Toxicol Lett*, **184**, 26-32.
92. Lo, H.W. and Ali-Osman, F. (2002) Cyclic AMP mediated GSTP1 gene activation in tumor cells involves the interaction of activated CREB-1 with the GSTP1 CRE: a novel mechanism of cellular GSTP1 gene regulation. *J Cell Biochem*, **87**, 103-116.
93. Zhang, L., Huang, J., Yang, N., Greshock, J., Megraw, M.S., Giannakakis, A., Liang, S., Naylor, T.L., Barchetti, A., Ward, M.R. *et al.* (2006) microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A*, **103**, 9136-9141.
94. Shi, L., Cheng, Z., Zhang, J., Li, R., Zhao, P., Fu, Z. and You, Y. (2008) hsa-mir-181a and hsa-mir-181b function as tumor suppressors in human glioma cells. *Brain Res*, **1236**, 185-193.
95. Bracken, C.P., Gregory, P.A., Kolesnikoff, N., Bert, A.G., Wang, J., Shannon, M.F. and Goodall, G.J. (2008) A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res*, **68**, 7846-7854.
96. Soung, Y.H., Lee, J.W., Kim, S.Y., Nam, S.W., Park, W.S., Lee, J.Y., Yoo, N.J. and Lee, S.H. (2006) Mutational analysis of the kinase domain of MYLK2 gene in common human cancers. *Pathol Res Pract*, **202**, 137-140.
97. Fujita, A., Gomes, L.R., Sato, J.R., Yamaguchi, R., Thomaz, C.E., Sogayar, M.C. and Miyano, S. (2008) Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates. *BMC Syst Biol*, **2**, 106.

98. Teng, C.S. (2000) Protooncogenes as mediators of apoptosis. *Int Rev Cytol*, **197**, 137-202.

99. Zhang, J., Du, Y.Y., Lin, Y.F., Chen, Y.T., Yang, L., Wang, H.J. and Ma, D. (2008) The cell growth suppressor, mir-126, targets IRS-1. *Biochem Biophys Res Commun*, **377**, 136-140.

100. Couch, F.J., Wang, X.Y., Wu, G.J., Qian, J., Jenkins, R.B. and James, C.D. (1999) Localization of PS6K to chromosomal region 17q23 and determination of its amplification in breast cancer. *Cancer Res*, **59**, 1408-1411.

101. Barlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M.L., Torhorst, J., Haas, P., Bucher, C., Sauter, G. *et al.* (2000) Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J Natl Cancer Inst*, **92**, 1252-1259.

102. Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M.L. *et al.* (2001) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A*, **98**, 5711-5716.

103. Ruiz, M., Pettaway, C., Song, R., Stoeltzing, O., Ellis, L. and Bar-Eli, M. (2004) Activator protein 2alpha inhibits tumorigenicity and represses vascular endothelial growth factor transcription in prostate cancer cells. *Cancer Res*, **64**, 631-638.

104. Deng, W.G., Jayachandran, G., Wu, G., Xu, K., Roth, J.A. and Ji, L. (2007) Tumor-specific activation of human telomerase reverses transcriptase promoter activity by activating enhancer-binding protein-2beta in human lung cancer cells. *J Biol Chem*, **282**, 26460-26470.

105. Tao, W., Zhang, S., Turenchalk, G.S., Stewart, R.A., St John, M.A., Chen, W. and Xu, T. (1999) Human homologue of the Drosophila melanogaster lats tumour suppressor modulates CDC2 activity. *Nat Genet*, **21**, 177-181.

106. Li, Y., Pei, J., Xia, H., Ke, H., Wang, H. and Tao, W. (2003) Lats2, a putative tumor suppressor, inhibits G1/S transition. *Oncogene*, **22**, 4398-4405.

107. Mahner, S., Baasch, C., Schwarz, J., Hein, S., Wolber, L., Janicke, F. and Milde-Langosch, K. (2008) C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. *Br J Cancer*, **99**, 1269-1275.

108. Wang, D., Weng, Q., Zhang, L., He, Q. and Yang, B. (2008) VEGF and Bcl-2 Interact Via MAPKs Signaling Pathway in the Response to Hypoxia in Neuroblastoma. *Cell Mol Neurobiol*.

109. Yonezawa, S. and Sato, E. (1997) Expression of mucin antigens in human cancers and its relationship with malignancy potential. *Pathol Int*, **47**, 813-830.

110. Mikami, Y., Hisatsune, A., Tashiro, T., Isohama, Y. and Katsuki, H. (2009) Hypoxia enhances MUC1 expression in a lung adenocarcinoma cell line. *Biochem Biophys Res Commun*, **379**, 1060-1065.
111. Konishi, N., Shimada, K., Nakamura, M., Ishida, E., Ota, I., Tanaka, N. and Fujimoto, K. (2008) Function of JunB in transient amplifying cell senescence and progression of human prostate cancer. *Clin Cancer Res*, **14**, 4408-4416.
112. Rasiah, K.K., Gardiner-Garden, M., Padilla, E.J., Moller, G., Kench, J.G., Alles, M.C., Eggleton, S.A., Stricker, P.D., Adamski, J., Sutherland, R.L. *et al.* (2008) HSD17B4 overexpression, an independent biomarker of poor patient outcome in prostate cancer. *Mol Cell Endocrinol*.
113. Rochester, M.A., Riedemann, J., Hellawell, G.O., Brewster, S.F. and Macaulay, V.M. (2005) Silencing of the IGF1R gene enhances sensitivity to DNA-damaging agents in both PTEN wild-type and mutant human prostate cancer. *Cancer Gene Ther*, **12**, 90-100.
114. Foster, C.S., Falconer, A., Dodson, A.R., Norman, A.R., Dennis, N., Fletcher, A., Southgate, C., Dowe, A., Dearnaley, D., Jhavar, S. *et al.* (2004) Transcription factor E2F3 overexpressed in prostate cancer independently predicts clinical outcome. *Oncogene*, **23**, 5871-5879.
115. Ueda, K., Arakawa, H. and Nakamura, Y. (2003) Dual-specificity phosphatase 5 (DUSP5) as a direct transcriptional target of tumor suppressor p53. *Oncogene*, **22**, 5586-5591.
116. Kobberup, S., Nyeng, P., Juhl, K., Hutton, J. and Jensen, J. (2007) ETS-family genes in pancreatic development. *Dev Dyn*, **236**, 3100-3110.
117. Takaori, K., Hruban, R.H., Maitra, A. and Tanigawa, N. (2004) Pancreatic intraepithelial neoplasia. *Pancreas*, **28**, 257-262.
118. Fauquette, V., Aubert, S., Groux-Degroote, S., Hemon, B., Porchet, N., Van Seuningen, I. and Pigny, P. (2007) Transcription factor AP-2alpha represses both the mucin MUC4 expression and pancreatic cancer cell proliferation. *Carcinogenesis*, **28**, 2305-2312.
119. Hagihara, A., Miyamoto, K., Furuta, J., Hiraoka, N., Wakazono, K., Seki, S., Fukushima, S., Tsao, M.S., Sugimura, T. and Ushijima, T. (2004) Identification of 27 5' CpG islands aberrantly methylated and 13 genes silenced in human pancreatic cancers. *Oncogene*, **23**, 8705-8710.
120. Markert, J.M., Fuller, C.M., Gillespie, G.Y., Bubien, J.K., McLean, L.A., Hong, R.L., Lee, K., Gullans, S.R., Mapstone, T.B. and Benos, D.J. (2001) Differential gene expression profiling in human brain tumors. *Physiol Genomics*, **5**, 21-33.
121. Tingulstad, S., Skjeldestad, F.E., Halvorsen, T.B. and Hagen, B. (2003) Survival and prognostic factors in patients with ovarian cancer. *Obstet Gynecol*, **101**,

885-891.

Appendix I Cancer-related miRNAs

Table A1 List of up or down regulated miRNAs that are reported in the literature in different human cancers or cancer cell lines

Cancer	miRNA	Up/Down	Reference
B-cell lymphoma	hsa-miR-16-1	Down	Eis et al., 2005
	hsa-miR-143	Down	He et al., 2005
	hsa-mir-17	Down	He et al., 2005
	hsa-miR-145	Down	O'Donnell et al., 2005
	hsa-miR-19a	Up	Eis et al., 2005
	hsa-miR-155	Up	Esquela-Kerscher and Slack, 2006
	hsa-miR-221	Up	Esquela-Kerscher and Slack, 2006
	hsa-miR-222	Up	Esquela-Kerscher and Slack, 2006
	hsa-miR-92a-1	Up	He et al., 2005
	hsa-miR-142	Up	O'Donnell et al., 2005
Bladder cancer	hsa-miR-21	Up	Neely et al., 2008
	hsa-miR-205	Up	Neely et al., 2008
	hsa-miR-223	Up	Gottardo et al., 2007
	hsa-miR-26b	Up	Gottardo et al., 2007
	hsa-miR-221	Up	Gottardo et al., 2007
	hsa-miR-103-1	Up	Gottardo et al., 2007
	hsa-miR-185	Up	Gottardo et al., 2007
	hsa-miR-23b	Up	Gottardo et al., 2007
	hsa-miR-203	Up	Gottardo et al., 2007
	hsa-miR-17	Up	Gottardo et al., 2007
	hsa-miR-23a	Up	Gottardo et al., 2007
	hsa-miR-205	Up	Gottardo et al., 2007
	hsa-miR-205	Up	Gottardo et al., 2007
Brain cancer	hsa-miR-181a-1	Down	Ciafre et al., 2005
	hsa-miR-21	Up	Ciafre et al., 2005
	hsa-miR-221	Up	Ciafre et al., 2005
Breast cancer	hsa-miR-10b	Down	Iorio et al., 2005
	hsa-miR-125b-1	Down	Iorio et al., 2005
	hsa-miR-145	Down	Iorio et al., 2005
	hsa-miR-21	Up	Chan et al., 2005
	hsa-miR-155	Up	Iorio et al., 2005
	hsa-miR-21	Up	Iorio et al., 2005
	hsa-miR-21	Up	Volinia et al., 2006
	hsa-miR-29b-2	Up	Volinia et al., 2006
	hsa-miR-146	Up	Volinia et al., 2006
	hsa-miR-125b-2	Down	Volinia et al., 2006
	hsa-miR-125b-1	Down	Volinia et al., 2006
	hsa-miR-10b	Down	Volinia et al., 2006
	hsa-miR-145	Down	Volinia et al., 2006
	hsa-miR-181a-1	Up	Volinia et al., 2006
	hsa-miR-140	Down	Volinia et al., 2006
	hsa-miR-213	Up	Volinia et al., 2006
	hsa-miR-29a	Up	Volinia et al., 2006

	hsa-miR-181b-1	Up	Volinia et al., 2006
	hsa-miR-199b	Up	Volinia et al., 2006
	hsa-miR-29b-1	Up	Volinia et al., 2006
	hsa-miR-130a	Down	Volinia et al., 2006
	hsa-miR-155	Up	Volinia et al., 2006
	hsa-let-7a-2	Down	Volinia et al., 2006
	hsa-miR-205	Down	Volinia et al., 2006
	hsa-miR-29c	Up	Volinia et al., 2006
	hsa-miR-224	Down	Volinia et al., 2006
	hsa-miR-100	Down	Volinia et al., 2006
	hsa-miR-31	Up	Volinia et al., 2006
	hsa-miR-30c	Down	Volinia et al., 2006
	hsa-miR-17-5p	Up	Volinia et al., 2006
	hsa-miR-210	Up	Volinia et al., 2006
	hsa-miR-122	Up	Volinia et al., 2006
	hsa-miR-16-2	Down	Volinia et al., 2006
	hsa-miR-373	Up	Huang et al., 2008
	hsa-miR-335	Down	Tavazoie et al.,2008
	hsa-miR-126	Down	Tavazoie et al.,2008
Cervix cancer	hsa-miR-145	Down	Esquela-Kerscher and Slack, 2006
	hsa-miR-143	Down	Esquela-Kerscher and Slack, 2006
	hsa-miR-143	Down	Lui et al., 2007
	hsa-miR-21	Up	Lui et al., 2007
Cholangiocarcinoma cell line	hsa-miR-141	Up	Meng et al., 2006
	hsa-miR-200b	Up	Meng et al., 2006
	hsa-miR-21	Up	Meng et al., 2006
	hsa-miR-15a	Down	Calin et al., 2004
	hsa-miR-16-1	Down	Calin et al., 2004
	hsa-miR-15a	Down	Cimmino et al., 2005
	hsa-miR-16-1	Down	Cimmino et al., 2005
CNS Tumor-derived cell line	hsa-let-7d	Down	Gaur et al, 2007
	hsa-let-7g	Down	Gaur et al, 2007
	hsa-let-7i	Down	Gaur et al, 2007
	hsa-miR-103-1	Down	Gaur et al, 2007
	hsa-miR-107	Down	Gaur et al, 2007
	hsa-miR-128-1	Down	Gaur et al, 2007
	hsa-miR-128-2	Down	Gaur et al, 2007
	hsa-miR-129-1	Down	Gaur et al, 2007
	hsa-miR-134	Down	Gaur et al, 2007
	hsa-miR-135a-1	Down	Gaur et al, 2007
	hsa-miR-137	Down	Gaur et al, 2007
	hsa-miR-138-1	Down	Gaur et al, 2007
	hsa-miR-138-2	Down	Gaur et al, 2007
	hsa-miR-148b	Down	Gaur et al, 2007
	hsa-miR-149	Down	Gaur et al, 2007
	hsa-miR-153-1	Down	Gaur et al, 2007
	hsa-mir-17	Down	Gaur et al, 2007
	hsa-miR-181a-1	Down	Gaur et al, 2007
	hsa-miR-181a-1	Down	Gaur et al, 2007
	hsa-miR-181b-1	Down	Gaur et al, 2007
	hsa-miR-181c	Down	Gaur et al, 2007

	hsa-miR-187	Down	Gaur et al, 2007
	hsa-miR-191	Down	Gaur et al, 2007
	hsa-miR-192	Down	Gaur et al, 2007
	hsa-miR-194-1	Down	Gaur et al, 2007
	hsa-miR-197	Down	Gaur et al, 2007
	hsa-miR-203	Down	Gaur et al, 2007
	hsa-miR-20a	Down	Gaur et al, 2007
	hsa-miR-212	Down	Gaur et al, 2007
	hsa-miR-219-1	Down	Gaur et al, 2007
	hsa-miR-26b	Down	Gaur et al, 2007
	hsa-miR-30b	Down	Gaur et al, 2007
	hsa-miR-30b	Down	Gaur et al, 2007
	hsa-miR-30c-1	Down	Gaur et al, 2007
	hsa-miR-30d	Down	Gaur et al, 2007
	hsa-miR-32	Down	Gaur et al, 2007
	hsa-miR-323	Down	Gaur et al, 2007
	hsa-miR-324	Down	Gaur et al, 2007
	hsa-miR-324	Down	Gaur et al, 2007
	hsa-miR-328	Down	Gaur et al, 2007
	hsa-miR-330	Down	Gaur et al, 2007
	hsa-miR-331	Down	Gaur et al, 2007
	hsa-miR-338	Down	Gaur et al, 2007
	hsa-miR-340	Down	Gaur et al, 2007
	hsa-miR-345	Down	Gaur et al, 2007
	hsa-miR-346	Down	Gaur et al, 2007
	hsa-miR-34a	Down	Gaur et al, 2007
	hsa-miR-34a	Down	Gaur et al, 2007
	hsa-miR-361	Down	Gaur et al, 2007
	hsa-miR-370	Down	Gaur et al, 2007
	hsa-miR-382	Down	Gaur et al, 2007
	hsa-miR-383	Down	Gaur et al, 2007
	hsa-miR-425	Down	Gaur et al, 2007
	hsa-miR-7-1	Down	Gaur et al, 2007
	hsa-miR-7-1	Down	Gaur et al, 2007
	hsa-miR-98	Down	Gaur et al, 2007
	hsa-miR-10a	Up	Gaur et al, 2007
	hsa-miR-196a-1	Up	Gaur et al, 2007
	hsa-miR-196a-2	Up	Gaur et al, 2007
	hsa-miR-196b	Up	Gaur et al, 2007
Colon cancer	hsa-miR-130a	Down	Lu et al, 2005
	hsa-miR-181a-1	Down	Lu et al, 2005
	hsa-miR-21	Up	Chan et al., 2005
	hsa-miR-145	Down	Schepeler et al., 2008
	hsa-miR-455	Down	Schepeler et al., 2008
	hsa-miR-484	Down	Schepeler et al., 2008
	hsa-miR-101-1	Down	Schepeler et al., 2008
	hsa-miR-101-2	Down	Schepeler et al., 2008
	hsa-miR-20a	Up	Schepeler et al., 2008
	hsa-miR-510	Up	Schepeler et al., 2008
	hsa-miR-92a-1	Up	Schepeler et al., 2008
	hsa-miR-513a-1	Up	Schepeler et al., 2008
	hsa-miR-24-1	Up	Volinia et al., 2006
	hsa-miR-29b-2	Up	Volinia et al., 2006

	hsa-miR-20a	Up	Volinia et al., 2006
	hsa-miR-10a	Up	Volinia et al., 2006
	hsa-miR-32	Up	Volinia et al., 2006
	hsa-miR-203	Up	Volinia et al., 2006
	hsa-miR-106a	Up	Volinia et al., 2006
	hsa-miR-17-5p	Up	Volinia et al., 2006
	hsa-miR-30c	Up	Volinia et al., 2006
	hsa-miR-223	Up	Volinia et al., 2006
	hsa-miR-126	Up	Volinia et al., 2006
	hsa-miR-128b	Up	Volinia et al., 2006
	hsa-miR-21	Up	Volinia et al., 2006
	hsa-miR-24-2	Up	Volinia et al., 2006
	hsa-miR-99b prec	Up	Volinia et al., 2006
	hsa-miR-155	Up	Volinia et al., 2006
	hsa-miR-213	Up	Volinia et al., 2006
	hsa-miR-150	Up	Volinia et al., 2006
	hsa-miR-107	Up	Volinia et al., 2006
	hsa-miR-191	Up	Volinia et al., 2006
	hsa-miR-221	Up	Volinia et al., 2006
	hsa-miR-9-3	Down	Volinia et al., 2006
ColonTumor-derived cell line	hsa-miR-130a	Down	Gaur et al, 2007
	hsa-miR-148a	Down	Gaur et al, 2007
	hsa-miR-15a	Down	Gaur et al, 2007
	hsa-miR-214	Down	Gaur et al, 2007
	hsa-miR-378	Down	Gaur et al, 2007
	hsa-miR-422a	Down	Gaur et al, 2007
	hsa-miR-424	Down	Gaur et al, 2007
Colorectal neoplasia	hsa-miR-133b	Down	Bandres et al., 2006
	hsa-miR-145	Down	Bandres et al., 2006
	hsa-miR-143	Down	Michael et al., 2003
	hsa-miR-145	Down	Michael et al., 2003
	hsa-miR-135b	Up	Bandres et al., 2006
	hsa-miR-183	Up	Bandres et al., 2006
	hsa-miR-31	Up	Bandres et al., 2006
	hsa-miR-96	Up	Bandres et al., 2006
Esophageal cancer	hsa-miR-103-1	Up	Guo et al., 2008
	hsa-miR-107	Up	Guo et al., 2008
Glioblastoma	hsa-miR-181a-1	Down	Ciafre et al., 2005
	hsa-miR-30c-1	Down	Ciafre et al., 2005
	hsa-miR-10b	Up	Chan et al., 2005
	hsa-miR-21	Up	Chan et al., 2005
Head and neck cancer cell lines	hsa-miR-127	Down	Tran et al., 2007
	hsa-miR-133a-1	Down	Tran et al., 2007
	hsa-miR-133b	Down	Tran et al., 2007
	hsa-miR-154	Down	Tran et al., 2007
	hsa-miR-200c	Down	Tran et al., 2007
	hsa-miR-212	Down	Tran et al., 2007
	hsa-miR-302b	Down	Tran et al., 2007
	hsa-miR-302c	Down	Tran et al., 2007
	hsa-miR-302d	Down	Tran et al., 2007
	hsa-miR-328	Down	Tran et al., 2007
	hsa-miR-340	Down	Tran et al., 2007
	hsa-miR-342	Down	Tran et al., 2007

	hsa-miR-345	Down	Tran et al., 2007
	hsa-miR-346	Down	Tran et al., 2007
	hsa-miR-371	Down	Tran et al., 2007
	hsa-miR-373	Down	Tran et al., 2007
	hsa-miR-375	Down	Tran et al., 2007
	hsa-miR-378	Down	Tran et al., 2007
	hsa-miR-382	Down	Tran et al., 2007
	hsa-miR-449a	Down	Tran et al., 2007
	hsa-let-7a-1	Up	Tran et al., 2007
	hsa-let-7b	Up	Tran et al., 2007
	hsa-let-7c	Up	Tran et al., 2007
	hsa-let-7d	Up	Tran et al., 2007
	hsa-let-7f-1	Up	Tran et al., 2007
	hsa-miR-100	Up	Tran et al., 2007
	hsa-miR-103-1	Up	Tran et al., 2007
	hsa-miR-107	Up	Tran et al., 2007
	hsa-miR-125b-1	Up	Tran et al., 2007
	hsa-miR-15a	Up	Tran et al., 2007
	hsa-miR-15b	Up	Tran et al., 2007
	hsa-miR-16-1	Up	Tran et al., 2007
	hsa-miR-16-2	Up	Tran et al., 2007
	hsa-miR-18a	Up	Tran et al., 2007
	hsa-miR-19a	Up	Tran et al., 2007
	hsa-miR-200a	Up	Tran et al., 2007
	hsa-miR-200b	Up	Tran et al., 2007
	hsa-miR-205	Up	Tran et al., 2007
	hsa-miR-21	Up	Tran et al., 2007
	hsa-miR-22	Up	Tran et al., 2007
	hsa-miR-221	Up	Tran et al., 2007
	hsa-miR-23a	Up	Tran et al., 2007
	hsa-miR-23b	Up	Tran et al., 2007
	hsa-miR-24-1	Up	Tran et al., 2007
	hsa-miR-27a	Up	Tran et al., 2007
	hsa-miR-28	Up	Tran et al., 2007
	hsa-miR-29b-2	Up	Tran et al., 2007
	hsa-miR-30b	Up	Tran et al., 2007
	hsa-miR-31	Up	Tran et al., 2007
	hsa-miR-320a	Up	Tran et al., 2007
	hsa-miR-361	Up	Tran et al., 2007
	hsa-miR-98	Up	Tran et al., 2007
Hematologic Tumor-derived cell line	hsa-miR-10a	Down	Gaur et al, 2007
	hsa-miR-196b	Down	Gaur et al, 2007
	hsa-miR-27b	Down	Gaur et al, 2007
	hsa-miR-28	Down	Gaur et al, 2007
Hepatocellular carcinoma	hsa-miR-125a	Down	Murakami et al., 2006
	hsa-miR-18a	Down	Murakami et al., 2006
	hsa-miR-195	Down	Murakami et al., 2006
	hsa-miR-199b	Down	Murakami et al., 2006
	hsa-miR-199a-1	Down	Murakami et al., 2006
	hsa-miR-199a-2	Down	Murakami et al., 2006
	hsa-miR-200a	Down	Murakami et al., 2006
	hsa-miR-224	Down	Murakami et al., 2006

hsa-mir-199a-1	Down	Gramantieri et al., 2007
hsa-mir-199a-2	Down	Gramantieri et al., 2007
hsa-mir-150	Down	Gramantieri et al., 2007
hsa-mir-200b	Down	Gramantieri et al., 2007
hsa-mir-214	Down	Gramantieri et al., 2007
hsa-mir-223	Down	Gramantieri et al., 2007
hsa-mir-145	Down	Gramantieri et al., 2007
hsa-mir-130a	Down	Gramantieri et al., 2007
hsa-miR-181a-1	Down	Gramantieri et al., 2007
hsa-mir-136	Down	Gramantieri et al., 2007
hsa-mir-141	Down	Gramantieri et al., 2007
hsa-mir-142	Down	Gramantieri et al., 2007
hsa-mir-143	Down	Gramantieri et al., 2007
hsa-mir-195	Down	Gramantieri et al., 2007
hsa-mir-9-1	Down	Varnholt et al., 2008
hsa-mir-29c	Down	Varnholt et al., 2008
hsa-mir-95	Down	Varnholt et al., 2008
hsa-mir-137	Down	Varnholt et al., 2008
hsa-mir-147	Down	Varnholt et al., 2008
hsa-mir-185	Down	Varnholt et al., 2008
hsa-mir-198	Down	Varnholt et al., 2008
hsa-mir-204	Down	Varnholt et al., 2008
hsa-mir-218-2	Down	Varnholt et al., 2008
hsa-mir-302b	Down	Varnholt et al., 2008
hsa-mir-145	Down	Varnholt et al., 2008
hsa-mir-145	Down	Wang et al., 2008
hsa-mir-139	Down	Wang et al., 2008
hsa-mir-214	Down	Wang et al., 2008
hsa-mir-215	Down	Kutay et al., 2006
hsa-mir-122	Down	Kutay et al., 2006
hsa-mir-122	Down	Laderio et al., 2008
hsa-mir-199a-1	Down	Laderio et al., 2008
hsa-mir-199a-2	Down	Laderio et al., 2008
hsa-mir-122	Down	Meng et al., 2007
hsa-miR-92a-1	Down	Meng et al., 2007
hsa-mir-125b-2	Down	Meng et al., 2007
hsa-mir-125a	Down	Meng et al., 2007
hsa-mir-101-1	Down	Jiang et al., 2008
hsa-mir-199a-1	Down	Jiang et al., 2008
hsa-mir-199a-2	Down	Jiang et al., 2008
hsa-mir-139	Down	Jiang et al., 2008
hsa-mir-214	Down	Jiang et al., 2008
hsa-mir-200b	Down	Jiang et al., 2008
hsa-mir-223	Down	Jiang et al., 2008
hsa-mir-150	Down	Jiang et al., 2008
hsa-mir-214	Down	Jiang et al., 2008
hsa-mir-30c-1	Down	Budhu et al., 2008
hsa-mir-1-2	Down	Budhu et al., 2008
hsa-mir-34a	Down	Budhu et al., 2008
hsa-mir-19a	Down	Budhu et al., 2008
hsa-mir-148a	Down	Budhu et al., 2008
hsa-mir-148b	Down	Budhu et al., 2008
hsa-mir-9-2	Down	Budhu et al., 2008

hsa-miR-194-1	Down	Budhu et al., 2008
hsa-mir-30a	Down	Budhu et al., 2008
hsa-mir-126	Down	Budhu et al., 2008
hsa-let-7g	Down	Budhu et al., 2008
hsa-mir-15a	Down	Budhu et al., 2008
hsa-mir-30e	Down	Budhu et al., 2008
hsa-mir-223	Down	Queenie et al., 2008
hsa-mir-126	Down	Queenie et al., 2008
hsa-mir-122	Down	Queenie et al., 2008
hsa-mir-222	Up	Queenie et al., 2008
hsa-mir-221	Up	Queenie et al., 2008
hsa-mir-31	Up	Queenie et al., 2008
hsa-mir-338	Up	Budhu et al., 2008
hsa-mir-219-1	Up	Budhu et al., 2008
hsa-mir-185	Up	Budhu et al., 2008
hsa-miR-224	Up	Murakami et al., 2006
hsa-let-7a-1	Up	Huang et al., 2008
hsa-let-7b	Up	Huang et al., 2008
hsa-let-7c	Up	Huang et al., 2008
hsa-let-7g	Up	Huang et al., 2008
hsa-let-7i	Up	Huang et al., 2008
hsa-mir-22	Up	Huang et al., 2008
hsa-mir-98	Up	Huang et al., 2008
hsa-mir-126	Up	Huang et al., 2008
hsa-mir-195	Up	Huang et al., 2008
hsa-mir-21	Up	Huang et al., 2008
hsa-mir-10a	Up	Varnholt et al., 2008
hsa-mir-15a	Up	Varnholt et al., 2008
hsa-mir-16-1	Up	Varnholt et al., 2008
hsa-mir-16-2	Up	Varnholt et al., 2008
hsa-mir-100	Up	Varnholt et al., 2008
hsa-mir-122	Up	Varnholt et al., 2008
hsa-mir-299	Up	Varnholt et al., 2008
hsa-mir-326	Up	Varnholt et al., 2008
hsa-mir-370	Up	Varnholt et al., 2008
hsa-mir-21	Up	Varnholt et al., 2008
hsa-mir-9-1	Up	Wang et al., 2008
hsa-mir-25	Up	Wang et al., 2008
hsa-mir-96	Up	Wang et al., 2008
hsa-mir-301a	Up	Wang et al., 2008
hsa-mir-21	Up	Wang et al., 2008
hsa-mir-221	Up	Wang et al., 2008
hsa-mir-137	Up	Wang et al., 2008
hsa-mir-151	Up	Wang et al., 2008
hsa-mir-155	Up	Wang et al., 2008
hsa-mir-182	Up	Wang et al., 2008
hsa-mir-183	Up	Wang et al., 2008
hsa-mir-186	Up	Wang et al., 2008
hsa-mir-216a	Up	Wang et al., 2008
hsa-mir-222	Up	Wang et al., 2008
hsa-mir-324	Up	Wang et al., 2008
hsa-mir-374a	Up	Wang et al., 2008
hsa-mir-224	Up	Wang et al., 2008

	hsa-mir-21	Up	Kutay et al., 2006
	hsa-mir-20a	Up	Kutay et al., 2006
	hsa-mir-23a	Up	Kutay et al., 2006
	hsa-mir-23b	Up	Kutay et al., 2006
	hsa-mir-24-1	Up	Kutay et al., 2006
	hsa-mir-93	Up	Kutay et al., 2006
	hsa-mir-99b	Up	Kutay et al., 2006
	hsa-miR-103-1	Up	Kutay et al., 2006
	hsa-mir-106a	Up	Kutay et al., 2006
	hsa-mir-106b	Up	Kutay et al., 2006
	hsa-mir-130a	Up	Kutay et al., 2006
	hsa-mir-219-1	Up	Kutay et al., 2006
	hsa-mir-320a	Up	Kutay et al., 2006
	hsa-mir-328	Up	Kutay et al., 2006
	hsa-let-7a-1	Up	Kutay et al., 2006
	hsa-mir-21	Up	Meng et al., 2007
	hsa-mir-221	Up	Meng et al., 2007
	hsa-mir-222	Up	Meng et al., 2007
	hsa-mir-34a	Up	Meng et al., 2007
	hsa-mir-210	Up	Meng et al., 2007
	hsa-mir-373	Up	Meng et al., 2007
	hsa-mir-376a-1	Up	Meng et al., 2007
	hsa-mir-18a	Up	Jiang et al., 2008
	hsa-mir-33a	Up	Jiang et al., 2008
	hsa-mir-130b	Up	Jiang et al., 2008
	hsa-miR-135a-1	Up	Jiang et al., 2008
	hsa-mir-301a	Up	Jiang et al., 2008
	hsa-mir-21	Up	Jiang et al., 2008
	hsa-mir-221	Up	Jiang et al., 2008
Lung cancer	hsa-let-7a-1	Down	Takamizawa et al., 2004
	hsa-let-7a-1	Down	Johnson et al., 2005
	hsa-let-7a-1	Down	Hayashita et al., 2005
	hsa-let-7a-1	Down	O'Donnell et al., 2005
	hsa-let-7a-1	Down	Takamizawa et al., 2004
	hsa-let-7a-1	Down	Johnson et al., 2005
	hsa-let-7a-1	Down	Hayashita et al., 2005
	hsa-let-7a-1	Down	O'Donnell et al., 2005
	hsa-miR-21	Up	Chan et al., 2005
	hsa-miR-205	Up	Michael et al., 2003
	hsa-miR-17	Up	Takamizawa et al., 2004
	hsa-miR-18a	Up	Takamizawa et al., 2004
	hsa-miR-19a	Up	Takamizawa et al., 2004
	hsa-miR-19b-1	Up	Takamizawa et al., 2004
	hsa-miR-20a	Up	Takamizawa et al., 2004
	hsa-miR-92a-1	Up	Takamizawa et al., 2004
	hsa-miR-17	Up	Johnson et al., 2005
	hsa-miR-18a	Up	Johnson et al., 2005
	hsa-miR-19a	Up	Johnson et al., 2005
	hsa-miR-19b-1	Up	Johnson et al., 2005
	hsa-miR-20a	Up	Johnson et al., 2005
	hsa-miR-92a-1	Up	Johnson et al., 2005
	hsa-miR-17	Up	Hayashita et al., 2005
	hsa-miR-18a	Up	Hayashita et al., 2005

hsa-miR-19a	Up	Hayashita et al., 2005
hsa-miR-19b-1	Up	Hayashita et al., 2005
hsa-miR-20a	Up	Hayashita et al., 2005
hsa-miR-92a-1	Up	Hayashita et al., 2005
hsa-miR-17	Up	O'Donnell et al., 2005
hsa-miR-18a	Up	O'Donnell et al., 2005
hsa-miR-19a	Up	O'Donnell et al., 2005
hsa-miR-19b-1	Up	O'Donnell et al., 2005
hsa-miR-20a	Up	O'Donnell et al., 2005
hsa-miR-92a-1	Up	O'Donnell et al., 2005
hsa-miR-17	Up	Takamizawa et al., 2004
hsa-miR-18a	Up	Takamizawa et al., 2004
hsa-miR-19a	Up	Takamizawa et al., 2004
hsa-miR-19b-1	Up	Takamizawa et al., 2004
hsa-miR-20a	Up	Takamizawa et al., 2004
hsa-miR-92a-1	Up	Takamizawa et al., 2004
hsa-miR-17	Up	Johnson et al., 2005
hsa-miR-18a	Up	Johnson et al., 2005
hsa-miR-19a	Up	Johnson et al., 2005
hsa-miR-19b-1	Up	Johnson et al., 2005
hsa-miR-20a	Up	Johnson et al., 2005
hsa-miR-92a-1	Up	Johnson et al., 2005
hsa-miR-17	Up	Hayashita et al., 2005
hsa-miR-18a	Up	Hayashita et al., 2005
hsa-miR-19a	Up	Hayashita et al., 2005
hsa-miR-19b-1	Up	Hayashita et al., 2005
hsa-miR-20a	Up	Hayashita et al., 2005
hsa-miR-92a-1	Up	Hayashita et al., 2005
hsa-miR-17	Up	O'Donnell et al., 2005
hsa-miR-18a	Up	O'Donnell et al., 2005
hsa-miR-19a	Up	O'Donnell et al., 2005
hsa-miR-19b-1	Up	O'Donnell et al., 2005
hsa-miR-20a	Up	O'Donnell et al., 2005
hsa-miR-92a-1	Up	O'Donnell et al., 2005
hsa-miR-1-1	Down	Nasser et al., 2008
hsa-miR-1-2	Down	Nasser et al., 2008
hsa-miR-21	Up	Volinia et al., 2006
hsa-miR-205	Up	Volinia et al., 2006
hsa-miR-200b	Up	Volinia et al., 2006
hsa-miR-9-1	Up	Volinia et al., 2006
hsa-miR-210	Up	Volinia et al., 2006
hsa-miR-148	Up	Volinia et al., 2006
hsa-miR-141	Up	Volinia et al., 2006
hsa-miR-132	Up	Volinia et al., 2006
hsa-miR-215	Up	Volinia et al., 2006
hsa-miR-128b	Up	Volinia et al., 2006
hsa-let-7g	Up	Volinia et al., 2006
hsa-miR-16-2	Up	Volinia et al., 2006
hsa-miR-129-1	Up	Volinia et al., 2006
hsa-miR-129-2	Up	Volinia et al., 2006
hsa-miR-126	Down	Volinia et al., 2006
hsa-miR-142	Up	Volinia et al., 2006
hsa-miR-30d	Down	Volinia et al., 2006

	hsa-miR-30a	Down	Volinia et al., 2006
	hsa-miR-7-2	Up	Volinia et al., 2006
	hsa-miR-199a-1	Up	Volinia et al., 2006
	hsa-miR-127	Up	Volinia et al., 2006
	hsa-miR-34a	Up	Volinia et al., 2006
	hsa-miR-136	Up	Volinia et al., 2006
	hsa-miR-202	Up	Volinia et al., 2006
	hsa-miR-196-2	Up	Volinia et al., 2006
	hsa-miR-199a-2	Up	Volinia et al., 2006
	hsa-let-7a-2	Up	Volinia et al., 2006
	hsa-miR-124a-1	Up	Volinia et al., 2006
	hsa-miR-149	Up	Volinia et al., 2006
	hsa-miR-17	Up	Volinia et al., 2006
	hsa-miR-196-1	Up	Volinia et al., 2006
	hsa-miR-10a	Up	Volinia et al., 2006
	hsa-miR-99b	Up	Volinia et al., 2006
	hsa-miR-196-1	Up	Volinia et al., 2006
	hsa-miR-199b	Up	Volinia et al., 2006
	hsa-miR-191	Up	Volinia et al., 2006
	hsa-miR-195	Up	Volinia et al., 2006
	hsa-miR-155	Up	Volinia et al., 2006
Lymphomas	hsa-miR-155	Up	Eis et al., 2005
	hsa-miR-17	Up	Eis et al., 2005
	hsa-miR-18a	Up	Eis et al., 2005
	hsa-miR-19a	Up	Eis et al., 2005
	hsa-miR-19b-1	Up	Eis et al., 2005
	hsa-miR-20a	Up	Eis et al., 2005
	hsa-miR-92a-1	Up	Eis et al., 2005
	hsa-miR-155	Up	Metzler et al., 2004
	hsa-miR-17	Up	Metzler et al., 2004
	hsa-miR-18a	Up	Metzler et al., 2004
	hsa-miR-19a	Up	Metzler et al., 2004
	hsa-miR-19b-1	Up	Metzler et al., 2004
	hsa-miR-20a	Up	Metzler et al., 2004
	hsa-miR-92a-1	Up	Metzler et al., 2004
	hsa-miR-155	Up	He et al., 2005
	hsa-miR-17	Up	He et al., 2005
	hsa-miR-18a	Up	He et al., 2005
	hsa-miR-19a	Up	He et al., 2005
	hsa-miR-19b-1	Up	He et al., 2005
	hsa-miR-20a	Up	He et al., 2005
	hsa-miR-92a-1	Up	He et al., 2005
Ovarian cancer	hsa-miR-210	Up	Yanaihara et al., 2006
	hsa-miR-21	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-141	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-200a	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-200c	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-200b	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-203	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-205	Up	Taylor and Gercel-Taylor, 2008
	hsa-miR-214	Up	Taylor and Gercel-Taylor, 2008
Pancreas cancer	hsa-miR-22	Up	Sun et al., 2008
	hsa-miR-22	Up	Sun et al., 2008

hsa-miR-375	Down	Lee et al., 2006
hsa-miR-345	Down	Lee et al., 2006
hsa-miR-142	Down	Lee et al., 2006
hsa-miR-139	Down	Lee et al., 2006
hsa-miR-221	Up	Lee et al., 2006
hsa-miR-424	Up	Lee et al., 2006
hsa-miR-301a	Up	Lee et al., 2006
hsa-miR-100	Up	Lee et al., 2006
hsa-miR-376a-1	Up	Lee et al., 2006
hsa-miR-125b-1	Up	Lee et al., 2006
hsa-miR-21	Up	Lee et al., 2006
hsa-miR-16-1	Up	Lee et al., 2006
hsa-miR-181a-1	Up	Lee et al., 2006
hsa-miR-181c	Up	Lee et al., 2006
hsa-miR-92a-1	Up	Lee et al., 2006
hsa-miR-15b	Up	Lee et al., 2006
hsa-miR-155	Up	Lee et al., 2006
hsa-let-7f-1	Up	Lee et al., 2006
hsa-miR-212	Up	Lee et al., 2006
hsa-miR-107	Up	Lee et al., 2006
hsa-miR-24-1	Up	Lee et al., 2006
hsa-miR-24-2	Up	Lee et al., 2006
hsa-let-7d	Up	Lee et al., 2006
hsa-miR-221	Up	Bloomston et al., 2007
hsa-miR-181a-1	Up	Bloomston et al., 2007
hsa-miR-155	Up	Bloomston et al., 2007
hsa-miR-210	Up	Bloomston et al., 2007
hsa-miR-222	Up	Bloomston et al., 2007
hsa-miR-181b-2	Up	Bloomston et al., 2007
hsa-miR-21	Up	Bloomston et al., 2007
hsa-miR-181b-1	Up	Bloomston et al., 2007
hsa-miR-181c	Up	Bloomston et al., 2007
hsa-miR-220a	Up	Bloomston et al., 2007
hsa-miR-181d	Up	Bloomston et al., 2007
hsa-miR-223	Up	Bloomston et al., 2007
hsa-miR-100	Up	Bloomston et al., 2007
hsa-miR-125a	Up	Bloomston et al., 2007
hsa-miR-143	Up	Bloomston et al., 2007
hsa-miR-10a	Up	Bloomston et al., 2007
hsa-miR-146a	Up	Bloomston et al., 2007
hsa-miR-99a	Up	Bloomston et al., 2007
hsa-miR-100	Up	Bloomston et al., 2007
hsa-miR-199a-1	Up	Bloomston et al., 2007
hsa-miR-10b	Up	Bloomston et al., 2007
hsa-miR-199a-2	Up	Bloomston et al., 2007
hsa-miR-107	Up	Bloomston et al., 2007
hsa-miR-103-2	Up	Bloomston et al., 2007
hsa-miR-125b-1	Up	Bloomston et al., 2007
hsa-miR-205	Up	Bloomston et al., 2007
hsa-miR-23b	Up	Bloomston et al., 2007
hsa-miR-23a	Up	Bloomston et al., 2007
hsa-miR-148a	Down	Bloomston et al., 2007
hsa-miR-148b	Down	Bloomston et al., 2007

hsa-miR-375	Down	Bloomston et al., 2007
hsa-miR-103-2	Up	Volinia et al., 2006
hsa-miR-103-1	Up	Volinia et al., 2006
hsa-miR-24-2	Up	Volinia et al., 2006
hsa-miR-107	Up	Volinia et al., 2006
hsa-miR-100	Up	Volinia et al., 2006
hsa-miR-125b-2	Up	Volinia et al., 2006
hsa-miR-125b-1	Up	Volinia et al., 2006
hsa-miR-24-1	Up	Volinia et al., 2006
hsa-miR-191	Up	Volinia et al., 2006
hsa-miR-23a	Up	Volinia et al., 2006
hsa-miR-26a-1	Up	Volinia et al., 2006
hsa-miR-125a	Up	Volinia et al., 2006
hsa-miR-130a	Up	Volinia et al., 2006
hsa-miR-26b	Up	Volinia et al., 2006
hsa-miR-145	Up	Volinia et al., 2006
hsa-miR-221	Up	Volinia et al., 2006
hsa-miR-126	Up	Volinia et al., 2006
hsa-miR-16-2	Up	Volinia et al., 2006
hsa-miR-146	Up	Volinia et al., 2006
hsa-miR-214	Up	Volinia et al., 2006
hsa-miR-99b	Up	Volinia et al., 2006
hsa-miR-128b	Up	Volinia et al., 2006
hsa-miR-155	Down	Volinia et al., 2006
hsa-miR-29b-2	Up	Volinia et al., 2006
hsa-miR-29a	Up	Volinia et al., 2006
hsa-miR-25	Up	Volinia et al., 2006
hsa-miR-16-1	Up	Volinia et al., 2006
hsa-miR-99a	Up	Volinia et al., 2006
hsa-miR-224	Up	Volinia et al., 2006
hsa-miR-30d	Up	Volinia et al., 2006
hsa-miR-92-2	Up	Volinia et al., 2006
hsa-miR-199a-1	Up	Volinia et al., 2006
hsa-miR-223	Up	Volinia et al., 2006
hsa-miR-29c	Up	Volinia et al., 2006
hsa-miR-30b	Up	Volinia et al., 2006
hsa-miR-129-1	Up	Volinia et al., 2006
hsa-miR-129-2	Up	Volinia et al., 2006
hsa-miR-197	Up	Volinia et al., 2006
hsa-miR-17	Up	Volinia et al., 2006
hsa-miR-30c	Up	Volinia et al., 2006
hsa-miR-7-1	Up	Volinia et al., 2006
hsa-miR-93-1	Up	Volinia et al., 2006
hsa-miR-140	Up	Volinia et al., 2006
hsa-miR-30a	Up	Volinia et al., 2006
hsa-miR-132	Up	Volinia et al., 2006
hsa-miR-181b-1	Up	Volinia et al., 2006
hsa-miR-152	Down	Volinia et al., 2006
hsa-miR-23b	Up	Volinia et al., 2006
hsa-miR-20a	Up	Volinia et al., 2006
hsa-miR-222	Up	Volinia et al., 2006
hsa-miR-27a	Up	Volinia et al., 2006
hsa-miR-92-1	Up	Volinia et al., 2006

	hsa-miR-21	Up	Volinia et al., 2006
	hsa-miR-129-1	Up	Volinia et al., 2006
	hsa-miR-129-2	Up	Volinia et al., 2006
	hsa-miR-150	Up	Volinia et al., 2006
	hsa-miR-32	Up	Volinia et al., 2006
	hsa-miR-106a	Up	Volinia et al., 2006
	hsa-miR-29b-1	Up	Volinia et al., 2006
	hsa-miR-21	Up	Chan et al., 2005
Papillary thyroid carcinoma	hsa-miR-146a	Up	He et al., 2005
	hsa-miR-181a-1	Up	He et al., 2005
	hsa-miR-221	Up	He et al., 2005
	hsa-miR-222	Up	He et al., 2005
	hsa-miR-146a	Up	Pallante et al., 2006
	hsa-miR-181a-1	Up	Pallante et al., 2006
	hsa-miR-221	Up	Pallante et al., 2006
	hsa-miR-222	Up	Pallante et al., 2006
Prostate cancer	hsa-let-7a-1	Down	Porkka et al., 2007
	hsa-let-7b	Down	Porkka et al., 2007
	hsa-let-7c	Down	Porkka et al., 2007
	hsa-let-7d	Down	Porkka et al., 2007
	hsa-let-7f-1	Down	Porkka et al., 2007
	hsa-let-7g	Down	Porkka et al., 2007
	hsa-miR-100	Down	Porkka et al., 2007
	hsa-miR-103-1	Down	Porkka et al., 2007
	hsa-miR-125a	Down	Porkka et al., 2007
	hsa-miR-125b-1	Down	Porkka et al., 2007
	hsa-miR-141	Down	Porkka et al., 2007
	hsa-miR-143	Down	Porkka et al., 2007
	hsa-miR-145	Down	Porkka et al., 2007
	hsa-miR-148a	Down	Porkka et al., 2007
	hsa-miR-16-1	Down	Porkka et al., 2007
	hsa-miR-195	Down	Porkka et al., 2007
	hsa-miR-199a-1	Down	Porkka et al., 2007
	hsa-miR-199a-1	Down	Porkka et al., 2007
	hsa-miR-19b-1	Down	Porkka et al., 2007
	hsa-miR-205	Down	Porkka et al., 2007
	hsa-miR-22	Down	Porkka et al., 2007
	hsa-miR-221	Down	Porkka et al., 2007
	hsa-miR-222	Down	Porkka et al., 2007
	hsa-miR-23a	Down	Porkka et al., 2007
	hsa-miR-23b	Down	Porkka et al., 2007
	hsa-miR-26b	Down	Porkka et al., 2007
	hsa-miR-27a	Down	Porkka et al., 2007
	hsa-miR-27b	Down	Porkka et al., 2007
	hsa-miR-29a	Down	Porkka et al., 2007
	hsa-miR-30a	Down	Porkka et al., 2007
	hsa-miR-30b	Down	Porkka et al., 2007
	hsa-miR-30c-1	Down	Porkka et al., 2007
	hsa-miR-497	Down	Porkka et al., 2007
	hsa-miR-92a-1	Down	Porkka et al., 2007
	hsa-miR-99a	Down	Porkka et al., 2007
	hsa-miR-21	Up	Chan et al., 2005
	hsa-miR-184	Up	Porkka et al., 2007

hsa-miR-198	Up	Porkka et al., 2007
hsa-miR-202	Up	Porkka et al., 2007
hsa-miR-210	Up	Porkka et al., 2007
hsa-miR-296	Up	Porkka et al., 2007
hsa-miR-302c	Up	Porkka et al., 2007
hsa-miR-320a	Up	Porkka et al., 2007
hsa-miR-345	Up	Porkka et al., 2007
hsa-miR-370	Up	Porkka et al., 2007
hsa-miR-373	Up	Porkka et al., 2007
hsa-miR-491	Up	Porkka et al., 2007
hsa-miR-498	Up	Porkka et al., 2007
hsa-miR-503	Up	Porkka et al., 2007
hsa-miR-513a-1	Up	Porkka et al., 2007
hsa-miR-32	Up	Ambs et al., 2008
hsa-miR-182	Up	Ambs et al., 2008
hsa-miR-31	Up	Ambs et al., 2008
hsa-miR-26a-1	Up	Ambs et al., 2008
hsa-miR-26a-2	Up	Ambs et al., 2008
hsa-miR-200c	Up	Ambs et al., 2008
hsa-miR-375	Up	Ambs et al., 2008
hsa-miR-196a-1	Up	Ambs et al., 2008
hsa-miR-196a-2	Up	Ambs et al., 2008
hsa-miR-370	Up	Ambs et al., 2008
hsa-miR-425	Up	Ambs et al., 2008
hsa-miR-194-1	Up	Ambs et al., 2008
hsa-miR-194-2	Up	Ambs et al., 2008
hsa-miR-181a-1	Up	Ambs et al., 2008
hsa-miR-181a-2	Up	Ambs et al., 2008
hsa-miR-34b	Up	Ambs et al., 2008
hsa-let-7i	Up	Ambs et al., 2008
hsa-miR-188	Up	Ambs et al., 2008
hsa-miR-25	Up	Ambs et al., 2008
hsa-miR-106b	Up	Ambs et al., 2008
hsa-miR-449a	Up	Ambs et al., 2008
hsa-miR-99b	Up	Ambs et al., 2008
hsa-miR-93	Up	Ambs et al., 2008
hsa-miR-92a-1	Up	Ambs et al., 2008
hsa-miR-92a-2	Up	Ambs et al., 2008
hsa-miR-125a	Up	Ambs et al., 2008
hsa-miR-520h	Down	Ambs et al., 2008
hsa-miR-494	Down	Ambs et al., 2008
hsa-miR-490	Down	Ambs et al., 2008
hsa-miR-133a-1	Down	Ambs et al., 2008
hsa-miR-1-2	Down	Ambs et al., 2008
hsa-miR-218-2	Down	Ambs et al., 2008
hsa-miR-220a	Down	Ambs et al., 2008
hsa-miR-128-1	Down	Ambs et al., 2008
hsa-miR-221	Down	Ambs et al., 2008
hsa-miR-499	Down	Ambs et al., 2008
hsa-miR-329-1	Down	Ambs et al., 2008
hsa-miR-340	Down	Ambs et al., 2008
hsa-miR-345	Down	Ambs et al., 2008
hsa-miR-410	Down	Ambs et al., 2008

hsa-miR-126	Down	Ambs et al., 2008
hsa-miR-205	Down	Ambs et al., 2008
hsa-miR-7-1	Down	Ambs et al., 2008
hsa-miR-7-2	Down	Ambs et al., 2008
hsa-miR-145	Down	Ambs et al., 2008
hsa-miR-34a	Down	Ambs et al., 2008
hsa-miR-487a	Down	Ambs et al., 2008
hsa-let-7b	Down	Ambs et al., 2008
hsa-let-7d	Up	Volinia et al., 2006
hsa-miR-128a	Down	Volinia et al., 2006
hsa-miR-195	Up	Volinia et al., 2006
hsa-miR-203	Up	Volinia et al., 2006
hsa-let-7a-2	Down	Volinia et al., 2006
hsa-miR-34a	Up	Volinia et al., 2006
hsa-miR-20a	Up	Volinia et al., 2006
hsa-miR-218-2	Down	Volinia et al., 2006
hsa-miR-29a	Up	Volinia et al., 2006
hsa-miR-25	Up	Volinia et al., 2006
hsa-miR-95	Up	Volinia et al., 2006
hsa-miR-197	Up	Volinia et al., 2006
hsa-miR-135-2	Up	Volinia et al., 2006
hsa-miR-187	Up	Volinia et al., 2006
hsa-miR-196-1	Up	Volinia et al., 2006
hsa-miR-148	Up	Volinia et al., 2006
hsa-miR-191	Up	Volinia et al., 2006
hsa-miR-21	Up	Volinia et al., 2006
hsa-let-7i	Up	Volinia et al., 2006
hsa-miR-198	Up	Volinia et al., 2006
hsa-miR-199a-2	Up	Volinia et al., 2006
hsa-miR-30c	Up	Volinia et al., 2006
hsa-miR-17	Up	Volinia et al., 2006
hsa-miR-92-2	Up	Volinia et al., 2006
hsa-miR-146	Up	Volinia et al., 2006
hsa-miR-181b-1	Up	Volinia et al., 2006
hsa-miR-32	Up	Volinia et al., 2006
hsa-miR-206	Up	Volinia et al., 2006
hsa-miR-184	Up	Volinia et al., 2006
hsa-miR-29a	Down	Volinia et al., 2006
hsa-miR-29b-2	Up	Volinia et al., 2006
hsa-miR-149	Down	Volinia et al., 2006
hsa-miR-181b-1	Up	Volinia et al., 2006
hsa-miR-196-1	Up	Volinia et al., 2006
hsa-miR-93-1	Up	Volinia et al., 2006
hsa-miR-223	Up	Volinia et al., 2006
hsa-miR-16-1	Up	Volinia et al., 2006
hsa-miR-101-1	Up	Volinia et al., 2006
hsa-miR-124a-1	Up	Volinia et al., 2006
hsa-miR-26a-1	Up	Volinia et al., 2006
hsa-miR-214	Up	Volinia et al., 2006
hsa-miR-27a	Up	Volinia et al., 2006
hsa-miR-24-1	Down	Volinia et al., 2006
hsa-miR-106a	Up	Volinia et al., 2006
hsa-miR-199a-1	Up	Volinia et al., 2006

Renal cancer	hsa-let-7f-2	Up	Gottardo et al., 2007
	hsa-miR-26b	Down	Gottardo et al., 2007
	hsa-miR-185	Up	Gottardo et al., 2007
	hsa-miR-141	Down	Nakada et al., 2008
	hsa-miR-200c	Down	Nakada et al., 2008
	hsa-miR-27a	Up	Gottardo et al., 2007
	hsa-miR-23b	Down	O'Rourke et al., 2006
	hsa-miR-24-1	Down	O'Rourke et al., 2006
Stomach cancer	hsa-miR-21	Up	Chan et al., 2005
	hsa-miR-223	Up	Volinia et al., 2006
	hsa-miR-21	Up	Volinia et al., 2006
	hsa-miR-218-2	Down	Volinia et al., 2006
	hsa-miR-103-2	Up	Volinia et al., 2006
	hsa-miR-92-2	Up	Volinia et al., 2006
	hsa-miR-25	Up	Volinia et al., 2006
	hsa-miR-136	Down	Volinia et al., 2006
	hsa-miR-191	Up	Volinia et al., 2006
	hsa-miR-221	Up	Volinia et al., 2006
	hsa-miR-125b-2	Up	Volinia et al., 2006
	hsa-miR-103-1	Up	Volinia et al., 2006
	hsa-miR-214	Up	Volinia et al., 2006
	hsa-miR-222	Up	Volinia et al., 2006
	hsa-miR-212	Down	Volinia et al., 2006
	hsa-miR-125b-1	Up	Volinia et al., 2006
	hsa-miR-100	Up	Volinia et al., 2006
	hsa-miR-107	Up	Volinia et al., 2006
	hsa-miR-92-1	Up	Volinia et al., 2006
	hsa-miR-96	Down	Volinia et al., 2006
	hsa-miR-192	Up	Volinia et al., 2006
	hsa-miR-23a	Up	Volinia et al., 2006
	hsa-miR-215	Up	Volinia et al., 2006
	hsa-miR-7-2	Up	Volinia et al., 2006
	hsa-miR-138-2	Down	Volinia et al., 2006
	hsa-miR-24-1	Up	Volinia et al., 2006
	hsa-miR-99b	Up	Volinia et al., 2006
	hsa-miR-33b	Down	Volinia et al., 2006
	hsa-miR-24-2	Up	Volinia et al., 2006
Testicular germ cell tumors	hsa-miR-372	Up	Voorhoeve et al., 2006
	hsa-miR-373	Up	Voorhoeve et al., 2006
Thyroid cancer	hsa-miR-146a	Up	He et al., 2005
	hsa-miR-221	Up	He et al., 2005
	hsa-miR-187	Up	Nikiforova et al., 2008
	hsa-miR-221	Up	Nikiforova et al., 2008
	hsa-miR-222	Up	Nikiforova et al., 2008
	hsa-miR-146b	Up	Nikiforova et al., 2008
	hsa-miR-155	Up	Nikiforova et al., 2008
	hsa-miR-224	Up	Nikiforova et al., 2008
	hsa-miR-197	Up	Nikiforova et al., 2008
	hsa-miR-30d	Down	Visone et al., 2007
	hsa-miR-125b-1	Down	Visone et al., 2007
	hsa-miR-26a-1	Down	Visone et al., 2007
	hsa-miR-30a	Down	Visone et al., 2007
	hsa-miR-221	Up	Pallante et al., 2006

hsa-miR-222	Up	Pallante et al., 2006
hsa-miR-181b-1	Up	Pallante et al., 2006
hsa-miR-197	Up	Weber et al., 2006
hsa-miR-346	Up	Weber et al., 2006

Appendix II Most differentially expressed miRNAs in each tissue

Table A2 The most differentially expressed miRNAs were identified from standard normal distribution

Tissue	miRNA	z-score	p-value
Brain	hsa-miR-330	6.166441437	0
Brain	hsa-miR-219	6.155599979	0
Brain	hsa-miR-9	6.149303519	0
Brain	hsa-miR-9*	6.146810392	0
Brain	cel-miR-124	6.135189458	0
Brain	hsa-miR-124a	5.973091852	0
Brain	hsa-miR-129	5.964733449	0
Brain	hsa-miR-340	5.759107905	0
Brain	hsa-miR-137	5.60770678	0
Brain	hsa-miR-153	5.579729094	0
Brain	hsa-miR-181b	5.53358058	0
Brain	hsa-miR-132	5.510212006	0
Brain	hsa-miR-383	5.426392084	0
Brain	hsa-miR-124b	5.426139319	0
Brain	hsa-miR-181b	5.318182541	0
Brain	hsa-miR-181d	5.082984252	0
Brain	hsa-miR-328	4.969354982	0
Brain	hsa-miR-324-5p	4.927015863	0
Brain	hsa-miR-346	4.91988791	0
Brain	hsa-miR-338	4.854180702	0
Brain	hsa-miR-433	4.806035691	0
Brain	hsa-miR-149	4.741905348	0
Brain	hsa-miR-323	4.279315562	0.0001
Brain	hsa-miR-128b	4.18399347	0.0001
Brain	hsa-miR-181c	4.060188598	0.0002
Brain	hsa-miR-128a	4.056137304	0.0002
Brain	hsa-miR-213	3.841746965	0.0004
Brain	hsa-miR-342	3.717774665	0.0006
Brain	hsa-miR-425	3.701753621	0.0007
Brain	hsa-miR-98	3.673701099	0.0008
Brain	hsa-miR-181a	3.531392669	0.0013
Brain	hsa-miR-331	3.501545743	0.0014
Brain	hsa-miR-151	3.409238699	0.002

Brain	hsa-miR-491	3.339842439	0.0026
Brain	hsa-miR-485-3p	3.252181581	0.0035
Brain	hsa-miR-218	3.221900518	0.0038
Brain	hsa-miR-485-5p	3.132713491	0.0051
Brain	hsa-miR-370	2.987345293	0.008
Brain	hsa-miR-139	2.821251445	0.0131
Brain	hsa-miR-324-3p	2.651673217	0.0208
Brain	hsa-miR-103	2.611845794	0.0231
Brain	hsa-miR-139	2.535638979	0.028
Brain	hsa-miR-125b	2.507243218	0.03
Brain	hsa-miR-361	2.33081001	0.0454
Liver	hsa-mir-122	6.166324971	0
Liver	hsa-mir-30e*	4.927492447	0
Liver	hsa-miR-92a-1	3.250755287	0
Liver	hsa-mir-148a	2.117824773	0.017
Pancreas	hsa-miR-217	6.166441437	0
Pancreas	hsa-miR-216	6.166086652	0
Pancreas	hsa-miR-141	5.487784263	0
Pancreas	hsa-miR-375	5.386166147	0
Pancreas	hsa-miR-148b	4.869396489	0
Pancreas	hsa-miR-148a	4.803975947	0
Pancreas	hsa-miR-141	4.151145152	0.0003
Pancreas	hsa-miR-130b	3.894458618	0.0057
Pancreas	hsa-miR-96	3.098338093	0.0131
Pancreas	hsa-miR-182	2.918124458	0.0217
Pancreas	hsa-miR-30e	2.82041734	0.0485
Breast	hsa-miR-126	3.192557833	0.0014
Breast	hsa-miR-26a	3.146502907	0.0017
Breast	hsa-miR-195	2.852579996	0.0043
Breast	hsa-miR-199b	2.774658926	0.0055
Breast	hsa-miR-205	2.58277472	0.0098
Heart	hsa-miR-302b	6.166441437	0
Heart	hsa-miR-302a	5.406344146	0
Heart	hsa-miR-302d	5.335713449	0
Heart	hsa-miR-189	5.301829902	0
Heart	hsa-miR-30e-3p	4.963839092	0
Heart	hsa-miR-302c	4.591792659	0
Heart	hsa-miR-367	3.631917733	0.0009
Heart	hsa-miR-422a	3.558232803	0.0012
Heart	hsa-miR-30c	2.816870236	0.0147
Heart	hsa-miR-221	2.778935839	0.0293

Ovary	hsa-let-7c	2.925654551	0.0034
Ovary	hsa-miR-502	2.216210531	0.0267
Ovary	hsa-miR-125b	2.205465596	0.0274
Kidney	hsa-miR-204	5.454413876	0
Kidney	hsa-miR-190	5.005567426	0
Kidney	hsa-miR-500	4.964273955	0
Kidney	hsa-miR-501	3.659108858	0.0008
Kidney	hsa-miR-30e	3.433586413	0.0018
Kidney	hsa-miR-196a	3.022291201	0.0072
Kidney	hsa-miR-489	2.961327888	0.0087
Kidney	hsa-miR-30a-5p	2.810569783	0.0135
Kidney	hsa-miR-363	2.608224598	0.0233
Kidney	hsa-miR-107	2.292953197	0.0494
Lung	hsa-miR-223	3.336488497	0.0026
Lung	hsa-miR-142-5p	3.187486859	0.0043
Lung	hsa-miR-181c	2.35385202	0.0431
Colon	hsa-miR-192	5.853878243	0
Colon	hsa-miR-194	5.558348545	0
Colon	hsa-miR-215	5.247238008	0
Colon	hsa-miR-31	3.2825819	0.0031
Colon	hsa-miR-338	2.936739188	0.0094
Colon	hsa-miR-18	2.607308026	0.0233
Colon	hsa-miR-200a	2.389625227	0.0397
Prostate	hsa-miR-34a	3.417250971	0.002
Prostate	hsa-miR-222	3.142279583	0.005
Prostate	hsa-miR-363	2.974831726	0.0083
Prostate	hsa-let-7c	2.82191799	0.0131
Prostate	hsa-miR-20b	2.417758645	0.0372
Prostate	hsa-miR-205	2.329159193	0.0456
Cervix	hsa-miR-99a	3.777158133	0.0005
Cervix	hsa-miR-100	3.569083632	0.0011
Cervix	hsa-miR-196b	2.999281123	0.0078
Cervix	hsa-miR-101	2.93119979	0.0095
Cervix	hsa-miR-29a	2.857504554	0.0118
Cervix	hsa-miR-195	2.700768754	0.0182
Stomach	hsa-miR-188	6.154640139	0
Stomach	hsa-miR-212	6.06839171	0
Stomach	hsa-miR-211	4.427095082	0
Stomach	hsa-miR-200a*	4.135190679	0.0001
Stomach	hsa-miR-197	4.097113277	0.0001
Stomach	hsa-miR-200a	3.829125066	0.0004

Stomach	hsa-miR-200b	3.805559972	0.0004
Stomach	hsa-miR-200c	3.448764079	0.0017
Stomach	hsa-miR-148a	3.408534292	0.002
Stomach	hsa-miR-346	3.218849413	0.0039
Stomach	hsa-miR-429	2.724576953	0.0171
Stomach	hsa-miR-29c	2.710416384	0.0178
Stomach	hsa-miR-31	2.654642095	0.0206
Stomach	hsa-miR-375	2.484620458	0.0317