Functional Characterization of an Exonic Small Non-coding RNA TIFm71

WANG, Dakui

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in

Biochemistry (Medicine)

The Chinese University of Hong Kong

September 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/Assessment Committee

Professor WAYE Miu-Yee, Mary (Chair)

Professor CHEUNG Wing-Tai (Thesis Supervisor)

Professor CHAN Ting-Fung (Committee Member)

Dr. LAM Yun-Wah (External Examiner)

Abstract

It has been widely acknowledged that non-coding RNAs (ncRNAs) are implicated in almost all known signaling pathways and regulate diverse cellular functions. The number of discovered ncRNA genes is increasing rapidly. However, their biological functions were not yet fully characterized.

In our previous study, Tumor-Induced Factor (TIF) and its transcript isoform SY3 were identified in xenograft tumors induced by oncogene *mas*-stably expressing CHO-K1 cells, coding a novel CXC chemokine protein. SY3 mRNA shared identical sequence with TIF mRNA except lacking of a 71 nt string in the 3'- untranslated region (3'-UTR), and this 71 nt sequence was referred as TIFm71. Genomic analysis indicated that TIF gene was composed of 4 exons and 3 introns, and TIFm71 was located in the fourth exon. It is of interest to note that TIFm71 shared similarity with human short interspersed nuclear elements Alu-J, and had a canonical precursor miRNA-like stem-loop secondary structure predicted with the RNA folding program "RNAfold".

Here we further showed that TIFm71 was released from the 3'-UTR of an exonic TIF transcript in transiently transfected HEK293 cells using polyacrylamide gel Northern blot analysis. TIFm71 was found to be expressed in normal tissues at a low level but greatly elevated in solid tumors. Ribonuclease III enzyme Dicer is a key protein in catalyzing miRNA maturation. In an *in vitro* processing assay, it was shown that TIFm71 could serve as a substrate of Dicer, which is an Ribonuclease III enzyme

playing a crucial role in catalyzing miRNA maturation. Nevertheless, RNAs of miRNA size (~21 nt) was barely detected in TIFm71 transiently transfected HEK293 cells.

An epithelial-to-mesenchymal transition (EMT) was observed in TIFm71-stably over-expressing CHO cells, evidenced by a decreased expression of cell adherent marker E-Cadherin, an increased expression of Snail1, and morphologic change from cobblestone-like appearance of epithelial cells to a spindle-, fibroblast-like morphology. High level of TIFm71 also increased cellular migratory rate and invasive ability of CHO-K1 cells. In addition, we observed a reduction of CHO-K1 cells in G2/M phase induced by TIFm71 over-expression. Potential binding proteins of TIFm71 RNA, including Ran and MPP4, were identified by RNA-pull down assay followed with LC-mass spectrometric analysis.

Taken together, these results suggest TIFm71 is released from the TIF transcript. Upon its release, TIFm71 serves as a scaffold to assemble a protein complex, which mediates EMT and cell cycle deregulation.

摘 要

非編碼RNA（ncRNA）的功能幾乎涉及所有已知的信號通路並對多種細胞功能具有調控作用。新發現的ncRNA基因數量正在迅速增加。但是，許多ncRNA的功能並不十分清楚。

在我們的前期研究中，腫瘤誘導因子TIF及其同源基因SY3在原癌基因*mas*穩定過表達的中國倉鼠卵巢細胞株（CHO）引發的移植瘤中被發現。它們可以編碼壹個CXC型的趨化因子蛋白。與TIF的序列相比，SY3除了在3’端非編碼區（3’-UTR）缺失壹個71 nt的序列以外，其他部分則與TIF完全相同。我們將這段71 nt的缺失序列命名為TIFm71。染色體組分析顯示，TIF基因由4個外顯子和3個內含子組成，TIFm71位於TIF基因的第四個外顯子。有趣的是，序列對比發現TIFm71與短散在核重復序列Alu-J具有相似性。不僅如此，通過RNA二級機構分析軟件RNAfold的預測，TIFm71可形成頸-環狀二級結構，此結構與miRNA前體（pre-miRNA）的二級結構極為類似。

應用聚丙烯酰胺凝膠Northern印跡分析，我們發現在瞬時轉染的HEK293細胞中，TIFm71可從TIF轉錄本中釋放出來。TIFm71在正常狀態下表達水平很低，但在實體腫瘤中卻大大提升。核糖核酸酶III Dicer是催化miRNA成熟的關鍵蛋白。盡管體外剪切實驗證明TIFm71可作為Dicer蛋白的底物，但Northern blot的結果顯示，瞬時轉染的HEK293細胞中TIFm71幾乎沒有生成miRNA。

TIFm71穩定過表達的CHO-K1具有上皮細胞間質轉型（epithelial-mesenchymal transitions, EMT）的傾向，表現在細胞粘附標記蛋白E-Cadherin表達水平的降低、Snail1表達水平的提高，與細胞形態從上皮狀向紡錘狀或成纖維狀的改變。高水平的TIFm71表達亦增加了細胞的遷移速率和侵染性。除此之外，TIFm71導致CHO-K1在細胞周期中G2/M期的比例減少。我們應用RNA沈澱和液相層析串聯質譜分析法鑒定了可與TIFm71 RNA結合的蛋白，包括Ran和MPP4。

以上結果合起來可以看出，TIFm71可從TIF轉錄本中釋放出來。釋放之後，TIFm71作為分子腳手架與蛋白結合而形成RNA-蛋白復合體，因此介導了EMT和細胞周期調節異常。

Acknowledgements

I would like to express my deep appreciation and respect for my supervisor Prof. Cheung Wing-Tai. It has been an honor to be his Ph.D student. I appreciate all his contributions of ideas, encouragements, continued guidance and patience during my research years in his lab. His strong scientific knowledge, remarkable memory together with his large-minded and respectful to others has been a source of inspiration for me and will continue to be so in the future.

My sincere thanks to my committee, Prof. WAYE Miu-Yee, Mary, Prof. CHAN Ting-Fung and Prof. LAM Yun-Wah (CityU), for their suggestions and guidance which has led me towards the completion of this degree. I also want to thank Prof. LAU Chi-Kong, Terrence for his suggestions on electrophoretic mobility shift assay, Prof. TO Ka-Fai for allowing me to do experiments in his lab in the first year of graduate study.

I would like to acknowledge group member LI Zhou-Fang (Iris), who earned her Ph.D degree here two years ago and continued her research study in City University of Hong Kong (CityU). Her great work built a solid foundation for my project, and her valuable suggestions and kind assistance contributed a lot to the accomplishments of my project on time. Other past and present graduate students that I have had the pleasure to work with or alongside of are graduate students JIANG Li-Li, SUN Jing-Xin (Jessie), ZHAO Qi, QI Wei and ZHANG Li. My warmest appreciation to group members such as technician Elaine CHAN, research assistant Shirley CHAN and

postdoc Stone ZHU, with whom I not only exchanged scientific ideas but also shared many happy and memorable moments in my life. Additionally, I would like to thank Mable YEUNG for her great help, encouragements and friendship, which have made the last two years an enjoyable life.

I would also like to acknowledge the people in different labs that shared their equipments and reagents with me, including FENG Lu, XU Jian, LIU Xing, ZOU Chang, CHU Zhi-Qin, LI Jian-zhen, XIE Chuan-Ming and NG Kwok-Shing, Partrick. Many thanks to Raymond LUNG and Joanna TONG for their technical assistance in handling isotope materials. In regard to the proteomics analysis, I am very thankful to LIANG Yi-Min (CityU), DONG Hong-Juan (CityU), Rajkumar (CityU) and SHEN Yun (HKU), who taught me how to prepare sample, helped me in M/S analysis and the following data analyzing.

Finally, I wish to express my greatest gratitude to my parents. Without their unending love and support from childhood to now, I never would have made it through this process or any of the tough times in my life.

WANG, Dakui

The Chinese University of Hong Kong

September 2011

List of Tables

List of Figures

Table of Contents

List of Abbreviations

ATP	Adenosine triphosphate
APS	Ammonium persulfate
BCA	Bicinchoninic acid
BLAST	Basic Local Aligment Search Tool
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CHO-K1	Chinese hamster ovary cells K1
DAPI	4', 6-diamidino-2-phenylindole, dihydrochloride
DEPC	Diethyl pyrocarbonate
DIG	Digoxigenin
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide-5-phosphate
DTT	Dothiothreitol
EB	Ethidium bromide
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FluorDD	Fluorescent differential display
g	Gram

GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GPCR	G protein-coupled receptor
h	Hours
HEK293	Human embryonic kidney cell
IMDM	Iscove's modified DMEM medium
kb	Kilobase
LB	Luric Broth
lincRNA	Large intervening non-coding RNA
M229	TIFm71 with flanking sequence, 229 bp
mg	Microgram
$MgCl_2$	Magnesium chloride
min	Minute
ml	Millilitre
mM	Millimolar
MOPS	Morpholinopropane-sulfonic acid
MTX	Methotrexate
ng	Nanogram
ncRNA	Non-coding RNA
OD	Optical density
Oligo (dT)	Oligopolydeoxythymidine
ORF	Open reading frame
PBS	Phosphate buffered saline

PCR	Polymerase chain reaction
PI	Propidium iodide
PMSF	Phenylmethylsulfonyl fluoride
pre-miRNA	Precursor micro-RNA
pri-miRNA	Primary micro-RNA
P/S	Penicillin and streptomycin
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel
S229	Scrambled TIFm71 with flanking sequence, 229 bp
SSC	Sodium chloride sodium citrate
RT-PCT	Reverse transcription- polymerase chain reaction
SDS	Sodium dodecyl sulfate
TAE	Tris acetate EDTA buffer
TBE	Tris-borate buffer
TEMED	Tetramethylethylenediamine
TIF	Tumor-induced factor
TIFm71	ncRNA in the 3'-UTR of TIF
Tris-HCl	Tris (hydroxymethy) amoinomethane/hydrochloric acid
µl	Microliter(s)
µM	Micromolar
UTR	Untranslated region
V	Voltage

Chapter 1

General Introduction

According to the latest data from genome analyses, humans have approximately 20,000 protein-coding genes as in other vertebrates (Wright and Bruford, 2011). This means that highly complex organisms like human have about the same number of protein-coding genes as other simpler life forms such as the roundworm, *Caenorhabditis elegans* (Wright and Bruford, 2011). It was reported that the percentage of genome coding for proteins decreases linearly with a function of biological evolution, with ~90% in prokaryotes, being decreased to ~68% in yeast, ~25% in nematodes, ~17% in insects, ~9% in pufferfish, ~2% in chicken and 1% in mammals (Taft *et al.*, 2010). The ENCODE (ENCyclopedia of DNA Elements) project showed that at least 90% of the human genome analyzed are transcribed in different cells, but a large portion of eukaryote transcripts does not code protein (Birney *et al.*, 2007). Since protein-coding genes comprise only ~1% of the human genome, it strongly indicates that non-coding RNAs (ncRNAs) which were previously regarded as transcriptional noise, could act as molecular markers in the evolution of complex organisms (Costa, 2008). However, the number of ncRNA types and their biological functions were not yet fully characterized.

In the present study, a novel ncRNA, TIFm71, was found being released from the 3'-untranslated region (3'-UTR) of an exonic TIF (Tumor-Induced Factor) mRNA, which could also encode a CXC chemokine protein. The expression of TIFm71 was inducible, relatively low in normal cell culture, but dramatically elevated in solid tumor. TIFm71 was predicted to fold into a stem-loop structure by a web-based RNA folding program RNAfold, and it could act as a substrate of Dicer, indicating that TIFm71 maybe a precusor microRNA (pre-miRNA). Moreover, functional assays showed that ectopic expression of TIFm71 promoted the non-invasive, epithelial like

ovary cells to undergo mesenchymal transition (EMT), and also caused a deregulation of cell cycle distribution.

1.1 Non-conding RNAs

Previously, the majority of the genomes of eukaryotes were thought to be either 'junk DNA' or 'transcriptional noise' with no functional purpose (Wright and Bruford, 2011). Indeed, in the years after the genome sequencing project, a series of studies in human (Birney *et al.*, 2007) and mouse (Carninci *et al.*, 2005) have shown that at least 63%, and perhaps more than 90%, of each base position in the genome is transcribed at least on one strand to produce large numbers of ncRNAs, whose incidence increases with developmental complexity (Carninci *et al.*, 2005; Mattick, 2009; Szymanski *et al.*, 2003).

1.1.1 Functional classes of ncRNA

The above-mentioned ncRNAs are transcribed from the exons and introns of non-protein-coding genes as well as from the intergenic regions of protein-coding genes by RNA polymerase II and III (Zhou *et al.*, 2010). These non-coding transcripts can be roughly divided into two groups, housekeeping ncRNAs and regulatory ncRNAs. Housekeeping ncRNAs are critical components of many cellular machines, including transfer RNAs (tRNA), ribosomal RNAs (rRNAs), spliceosomal RNAs and so on (Wilusz *et al.*, 2009). The regulatory ncRNAs including microRNAs (miRNAs; ~21 nt), small nucleolar RNAs (snoRNAs; 60-300 nt), small interfering RNAs (siRNAs; 21-25 nt), Piwi-associated RNAs (piRNAs; 26-31 nt), enhancer RNAs (eRNAs) and

short interspersed nuclear elements (SINEs) transcribed RNAs, which are important in regulating diverse cellular functions (Prasanth and Spector, 2007; Ren, 2010). tRNAs and rRNAs are involved in mRNA translation, snoRNAs are involved in the modification of rRNAs and snRNAs (small nuclear RNAs) are involved in RNA splicing. Besides, a number of additional diverse genomic and cellular functions of ncRNAs have been revealed as summarized in Table 1.1 (Zhou *et al.*, 2010). There are certainly many other ncRNAs varieties, both short and long, that are waiting to be identified (Ponjavic and Ponting, 2007).

Table 1.1 Functional classes of ncRNAs

Name	Length (nt)	Main function
microRNA	19-25	Translational repression or mRNA cleavage
siRNA	19-29	mRNA cleavage
piRNA	25-31	Maintaining transposon silencing
rasiRNA	24-29	Direct silencing of retrotransposons and repetitive sequences
rasi/piRNA	24-31	Heterochromatin formation and transcriptional silencing
lsiRNAs (long siRNA)	30-40	Destabilizing mRNA through decapping and 5'-to-3' degradation
tasiRNA	20-25	Endogenous mRNA cleavage
nat-siRNA	20-25	Cleavage of constitutively expressed transcripts
Heterochromatic siRNA	24	Direct DNA and histone methylation
scnRNA (small-scan RNA)	27-30	Histone methylation and DNA elimination and genome re-arrangement
gRNA (guide RNA)	35-78	Directing the insertion or deletion of uridine residues into mRNA
ASRNA	Uncertain	Transcriptional and translational repression
lRNA (long RNA)	>200	Chromatin modification and transcriptional and post-transcriptional regulation
eRNA (enhancer RNA)	Uncertain	Regulation Gene expression regulation during neuronal activity

Adopted and modified from Zhou *et al, Biol Cell* (2010), **102**, 645-55.

1.1.2 microRNA

Recently, microRNAs (miRNAs) emerged as a new class of noncoding genes involved in regulating cell proliferation, differentiation, and viability (Bartel, 2004; Stefani and Slack, 2008). miRNAs are short (19-24 nucleotides), endogenous, single-stranded RNA molecules that regulate gene expression (Zamore and Haley, 2005).

It was suggested that an ncRNA could be defined as miRNA if the following combination of expression and biogenesis criteria can be fulfilled (Ambros *et al.*, 2003; Berezikov *et al.*, 2006; Choi *et al.*, 2008): firstly, mature miRNA should originate from a precursor with a characteristic hairpin secondary structure that does not contain large internal loops or bulges (Gruber *et al.*, 2008); secondly, mature miRNA should be processed by Dicer, as determined by an increase in the accumulation of precursor in Dicer-deficient mutants; and thirdly, mature miRNA should be expressed as a distinct transcript of ~22 nt, detectable by RNA (Northern) blot analysis or other experimental means.

A canonical miRNA biogenesis pathway starts with the transcription of primary miRNA (pri-miRNA) by RNA Polymerase II or III (Borchert *et al.*, 2006; Lee *et al.*, 2004). In the nucleus, pri-miRNAs are processed into 50 to 80 bases hairpin-like pre-miRNA by microprocessor complex of nucleases and associated factors, including the RNase III Drosha and its partner DiGeorge syndrome critical region gene (DGCR8)/Pasha (Denli *et al.*, 2004; Lee *et al.*, 2003). Pre-miRNA molecules, in association with the Ran-GTP dependent factor exportin- 5, are actively transported

into the cytoplasm where a second ribonuclease III enzyme Dicer processes them into duplexes that contain the 19 to 24 nucleotides mature miRNA and an oligonucleotides with similar size from the other arm of the hairpin (Hutvagner *et al.*, 2001; Lund *et al.*, 2004; Yi *et al.*, 2003). The miRNA is incorporated into a ribonucleoprotein complex called RNA-induced silencing complex (RISC) and guides the RISC to the target mRNA (Bartel, 2004).

Most pri-miRNAs are transcribed by RNA polymerase II, as part of non-coding genes or from introns of protein-coding genes (Saj and Lai, 2011). Analysis of the genomic localization of known human miRNAs revealed that the majorities are in intergenic regions, and sometimes in clusters of several miRNAs, and therefore must depend on their own promoters (Baskerville and Bartel, 2005; Lagos-Quintana *et al.*, 2003; Shivdasani, 2006). In addition, ~25% of human miRNA genes are located within introns of known protein coding genes (Cai *et al.*, 2004).

Most animal miRNAs imperfectly base-pair with sequences in the 3`-UTRs of target mRNAs, and inhibit protein synthesis by either repressing translation or promoting mRNA deadenylation and decay (Krol *et al.*, 2010). By forming a RISC complex, miRNA could repress target-gene expression post-transcriptionally at the level of translation. Consistent with translational control, miRNAs using this mechanism to reduce the protein levels of their target genes, but the mRNA levels of these genes are barely affected (Esquela-Kerscher and Slack, 2006). Also, miRNAs could bind with perfect or nearly perfect complementary to protein-coding mRNA sequences to induce the RNA-mediated interference (RNAi) pathway. Briefly, mRNA transcripts

are cleaved by ribonucleases in the miRNA-associated RISC, which results in the degradation of target mRNAs (Tang *et al.*, 2003; Yekta *et al.*, 2004). Bioinformatics data indicated that single miRNA could bind to hundreds of mRNA targets, and these targets could theoretically regulate every biological process (Ambros, 2003; Zamore and Haley, 2005).

Many studies highlighted that miRNAs functions in various physiological processes and their deregulation in cancers (Ryan *et al.*, 2010). Aberrant miRNA expression patterns were found in most, if not all, human malignancies, including the most common ones, such as lung, breast, prostate and gastrointestinal cancers (Visone and Croce, 2009). miRNA genes could function as potential oncogenes and tumor repressors genes in human body. Besides, a strong linkage between alterations in miRNA binding sites and inherited cancer risk was reported recently (Brendle *et al.*, 2008; Christensen *et al.*, 2009; Nelson *et al.*, 2010; Zhang *et al.*, 2011). For instance, let-7 miRNA-binding site polymorphism in the KRAS 3'-UTR is associated with reduced survival in oral cancers (Christensen *et al.*, 2009). Although the pathways that mediate this risk are yet to be elucidated, there is a clear suggestion that cancer risk is mediated by changes in miRNA sequence and maturation (Ryan *et al.*, 2010).

1.1.3 Short interspersed nuclear elements (SINEs)

Recent genome comparisons have revealed the occurrence of numerous conserved non-coding elements (CNEs) in the human genome. Strikingly, many CNEs seem to be derived from ancient transposable elements (TE) and in particular from a class of

retrotransposons known as short interspersed nuclear elements (SINEs), to which human Alu elements and mouse B2 RNA belong (Cordaux and Batzer, 2009). These ancient SINEs-derived sequences are currently evolving under strong negative selection and apparently taken on regulatory functions. While historically viewed as "junk DNA", recent research suggested that in some rare cases, SINEs were incorporated into novel genes, so as to evolve new functionality (Kamal *et al.*, 2006; Nishihara *et al.*, 2006). Furthermore, bioinformatic analysis showed thousands of SINEs were present in constrained non-exonic elements, suggesting that SINEs serve a biological and perhaps an evolutionary role (Lowe *et al.*, 2007; Ponicsan *et al.*, 2010).

Typically, SINEs are 75-500 bp in length and contain internal promoters for RNA polymerase III (Nishihara *et al.*, 2006). It can be transcribed into tRNA, rRNA and other small nuclear RNAs. SINEs do not encode a functional reverse transcriptase protein and rely on other mobile elements for transposition. The most common SINEs in primates are called Alu sequences, which are middle repetitive SINEs dispersed throughout vertebrate genomes. Alu elements are ~280 bp in length, do not contain any coding sequences and can be recognized by the restriction enzyme Alu I. It affects chromatin structure, sites of recombination, RNA processing and stability (Trujillo *et al.*, 2006). Moreover, Alu sequences were reported to be involved in hepatocellular carcinoma (Tang *et al.*, 2005). Although most Alu repeats are transcriptionally silent, these elements are potentially functional class III genes, and can be transcribed by RNA polymerase III. It has been reported that fragmented and/or full-length Alu elements were found in the coding regions of mRNA but their functions are to be clarified (Moolhuijzen *et al.*, 2010).

1.1.4 ncRNAs are important in regulating protein activities

Generally, ncRNAs elicit their biological responses through one of the three basic mechanisms: base-pairing with a target nucleic acid, catalysing biological reactions or binding to and modulating the activity of a protein (Goodrich and Kugel, 2006). For the later two mechanisms, ncRNAs usually fold into unique higher-order structures, which are required for their functioning, much like a protein. Many ncRNAs become integral parts of large complexes that contain proteins or possibly other RNAs, and the components of the complex function together as a unit (Goodrich and Kugel, 2006).

A small double-stranded ncRNA, neuron-restrictive silencer element RNA (NRSE dsRNA), was found to play a critical role in mediating neuronal differentiation (Kuwabara *et al.*, 2004). NRSE dsRNA is ~20 nt long. It could trigger gene expression of neuron-specific genes through interaction with NRSF/REST (neuron-restrictive silencing factor/RE-1-silencing transcription factor) transcriptional machinery, resulting in the transition from neural stem cells to neuronal cells. NRSE dsRNA sequence was found to be complementary to a promoter element, which binds to the repressor protein NRSF/REST in order to silence the transcription of neuron-specific genes in non-neuronal cells. Different from miRNA or siRNA, NRSE dsRNA functions through a direct RNA-protein interaction between itself and the repressor NRSF/REST rather than by base pairing with the promoter element with which it shares sequence homology (Kuwabara *et al.*, 2004).

B2 RNA is ~178 nt long mouse ncRNA transcribed from SINEs, which are retrotransposons abundantly scattered throughout the mouse genome. It directly binds to core RNA Polymerase II (Pol II) and represses transcription from specific genes after heat shock (Allen *et al.*, 2004; Espinoza *et al.*, 2004). During the heat-shock response, the newly transcribed B2 RNA represses the transcription of genes such as actin and hexokinase II (Espinoza *et al.*, 2007). B2 RNA potently inhibits transcription by binding to an RNA docking site on Pol II with a high affinity and specificity. Through this interaction, B2 RNA assembles into preinitiation complexes at the promoter and blocks RNA synthesis (Espinoza *et al.*, 2004). Moreover, ncRNA transcriptional regulators that function by binding directly to an RNA polymerase are not limited to eukaryotes. In bacteria, 6S RNA associates with RNA polymerase in a highly specific and efficient manner, regulating gene expression as cells enter stationary phase (Wassarman and Storz, 2000).

There are two other ncRNAs reported previously that could promote cancer cell metastatic progression, HOTAIR and MALAT-1. HOTAIR is a large intervening non-coding RNAs (lincRNAs). It recruits the PRC2 complex to specific target genes genome-wide, leading to H3K27 trimethylation and epigenetic silencing of metastasis suppressor genes. Within the context of cancer cells, ectopic expression of HOTAIR seems to re-impose the chromatin state, thereby enabling gene expression programs that are conducive to cell motility and matrix invasion (Gupta *et al.*, 2010). MALAT-1, another long ncRNA, was found to play a pivotal role in colorectal cancer metastasis (Ji *et al.*, 2003; Xu *et al.*, 2011), but the underlying mechanism was

unknown. Based on our current understanding of the molecular mechanisms by which ncRNAs functions, ncRNAs commonly act as adaptors that position a target molecule, which is a nucleic acid or proteins to form a ribonucleoprotein unit for enzyme activity (Huttenhofer and Schattner, 2006; Wilusz *et al.*, 2009).

1.2 Chemokine

Chemokines (chemotactic cytokines) are small secretory proteins (8-14 kDa) that direct the movement of circulating leukocytes to sites of inflammation or injury (Charo and Ransohoff, 2006). Chemokine messages are decoded through specific cell-surface receptors. Upon chemokine binding, these receptors unleash cascades of intracellular secondary mediators, turning on cell-specific intrinsic functional programs including directional migration, which is the best-known effect of chemokines (Rot and von Andrian, 2004). To date, approximately 50 human chemokines and 20 receptors were discovered (Allen *et al.*, 2007).

1.2.1 Types of chemokines

Based on the position of two conserved N-terminal cysteine residues of chemokines, they are classified into four groups: CXC, CC, CX3C and C chemokines (Baggiolini *et al.*, 1997; Luster, 1998; Rollins, 1997). The CXC chemokines can be further classified into ELR^+ and ELR^- chemokines based on the presence or absence of the motif "glu-leu-arg (ELR motif)" preceding the CXC sequence (Rossi and Zlotnik, 2000). Although the primary function of chemokines was well recognized as leukocyte attractants, recent evidences indicated that they also play a role in a number

of tumor-related processes, such as growth, angiogenesis and metastasis (Raman *et al.*, 2007).

1.2.2 Roles of chemokines

The main biological function of chemokines is to direct the movement of circulating leukocytes to sites of inflammation or injury, they also connected with cancer and the recruitment of tumor associated macrophages (Mantovani *et al.*, 2010). They can be released by many different cell types and serve to guide cells involving in innate immunity, as well as the lymphocytes in the adaptive immune system. Besides the original studies on their roles in inflammation, it is now clear that they subserve a much broader function ranging from cellular senescence to metastasis (McDonald and Kubes, 2010). Chemokines and their receptors are now known to play a crucial part in directing the movement of mononuclear cells throughout the body, engendering the adaptive immune response and contributing to the pathogenesis of a variety of diseases (Charo and Ransohoff, 2006). For example, in the struggle for optimal host defense against infection by viruses, chemokines exhibit a distinct capacity to regulate both death of the infected host cell and proper immune cell activation (Grayson and Holtzman, 2006).

1.3 Splicing of RNA transcript

Splicing of RNA transcript is a process by which the exons of the RNA produced by a primary gene transcript (pre-mRNA) are reconnected in multiple ways during RNA splicing to produce more than one mRNA isoforms (Black, 2003). Genome-wide

studies based on RNA deep-sequencing revealed that 92-94% of human transcripts are alternatively spliced in a highly tissue-specific fashion, ~86% with a minor isoform frequency of 15% or more (Wang *et al.*, 2008). Alternative splicing plays critical roles in differentiation, development, disease and has been proposed as a primary driver of the evolution of phenotypic complexity in mammals (Luco *et al.*, 2011; Wang *et al.*, 2008). It is a major source for protein diversity in higher eukaryotes, with half of these altering the reading frame (Clark and Thanaraj, 2002) and one third apparently leading to nonsense mediated decay (NMD) of the RNA product (Lewis *et al.*, 2003). Also, alternative splicing of untranslated regions can have important regulatory consequences, including NMD, even though the open reading frame is unchanged (Matlin *et al.*, 2005).

1.3.1 General splicing mechanism

The 5' splice site (5'ss), 3' splice site (3'ss), and branch point sequence (BPS), participate in the splicing reaction and present in every intron, thus are known as the core splicing signals (Wang and Burge, 2008). Each intron has GU at its 5' end and a branch site near the 3' end. The nucleotide at the branch point is always an A. In humans, the branch consensus around this sequence is yUnAy (Gao *et al.*, 2008), which is followed by a series of pyrimidines or polypyrimidine tract, then by AG at 3' end (Matlin *et al.*, 2005). The removal of introns is catalysed by the spliceosome, an assembly of 5 small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U4, U5 and U6) that are associated with a large number of additional proteins (Jurica and Moore, 2003). Eight common types of alternative transcript events derived on the basis of available human cDNA and expressed sequence tag data (Wang *et al.*, 2008)

are shown in Fig. 1.1. The regulation of alternative splicing involves both *cis* and *trans* components, which are composed of sequences in the pre-mRNA and cellular factors (RNA or protein), respectively (Matlin *et al.*, 2005). Its specificity is commonly thought to be achieved by the combinatorial recruitment of basal splicing factors to regulatory RNA motifs on the pre-mRNA (Johnson *et al.*, 2003).

1.3.2 *Cis* ans *Trans* regulatory elements for RNA splicing

Splicing can be regulated by *trans*-acting proteins (repressors and activators), corresponding *cis*-acting regulatory sites (silencers and enhancers) and other RNA features such as RNA secondary structures. These elements form a "splicing code" that governs how splicing will occur under different cellular conditions (Wang and Burge, 2008). The majority of splicing repressors are heterogeneous nuclear ribonucleoproteins (hnRNPs) such as hnRNPA1 and polypyrimidine tract binding protein (PTB). By contrast, most of the activator proteins that bind to intron splicing enhancers (ISEs) and exon splicing enhancers (ESEs) are members of the SR protein family, which contain RNA recognition motifs and arginine/serine-rich (RS) domains (Matlin *et al.*, 2005; Wang and Burge, 2008). The availability of splicing factors depends primarily on their expression levels, nuclear localization and post-translational modifications modulating their activity (Grosso *et al.*, 2008; Luco and Misteli, 2011). *Cis*-acting influences the secondary structure of the RNA, which can either sequester sequence elements or bring them into closer apposition (Buratti and Baralle, 2004), and the relative spacing of sequence elements (Matlin *et al.*, 2005).

Alternative transcript events	Total events ($\times 10^3$)	Number detected ($\times 10^3$)	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon	37	35	10,436	6,822	65	72
Retained intron	1	1	167	96	57	71
Alternative 5′ splice site (A5SS)	15	15	2,168	1,386	64	72
Alternative 3′ splice site (A3SS)	17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE)	4	4	167	95	57	66
Alternative first exon (AFE)	14	13	10,281	5,311	52	63
Alternative last exon (ALE)	9	8	5,246	2,491	47	52
Tandem 3′ UTRs	7	7	5,136	3,801	74	80
Total	105	100	37,782	22,657	60	68

Constitutive exon or region — Body read — Junction read — pA Polyadenylation site

Alternative exon or extension

Figure 1.1 Pervasive tissue-specific regulation of alternative mRNA isoforms. Adopted from Wang *et al, Nature* (2008), **456**, 470-6.

1.3.3 Excised introns could possibly act as ncRNAs

Introns have long been regarded as evolutionary debris with intronic RNA assumed to be simply degraded after splicing excision. Also, the increasing number of non-protein coding transcripts being detected in mammalian cells has been suggested, at least by some, to be largely 'transcriptional noise' (Dennis, 2002; Mattick and Makunin, 2005). However, a significant proportion of ncRNAs appears to be stable in eukaryotic cells. One example is from a study on the evaluation of the half-life of mouse $IVS1_{C\beta1}$, which is a typical vertebrate intron, has a consensus 5' SS (CAG/GTAAGT) and an uninterrupted 35 nt polypyrimidine tract upstream of the 3' SS. The half-life of $IVS1_{C\beta1}$ is 6.0 ± 1.4 min, which is comparable to some of the short-lived mammalian mRNAs including c-myc and c-fos mRNAs (Clement *et al.*, 1999). Introns of *Pem* gene show even longer half-life, ranging from 9-30 min. Interestingly, spliced *Pem* introns are exported from the nucleus to the cytoplasm (Clement *et al.*, 2001). Therefore, excised introns could possibly be functional as ncRNAs.

1.4 Epithelial-to-mesenchymal transition

Epithelial and mesenchymal cells differ in various functional and phenotypic characteristics. Epithelial cells form layers of cells that are closely adjoined by specialized membrane structures. They also have apical-basolateral polarization, which manifests itself through the localized distribution of adhesion molecules. On the other hand, mesenchymal cells contact neighbouring mesenchymal cells focally, not typically associated with a basal lamina. In culture, epithelial cells grow as clusters of cells, maintaining complete cell-cell adhesion with their neighbours,

whereas mesenchymal cells have a spindle-shaped, fibroblast-like morphology, and usually tend to be highly mobile (Thiery and Sleeman, 2006).

An epithelial-to-mesenchymal transition (EMT) is a biologic process that allows a polarized epithelial cell, which normally interacts with basement membrane via its basal surface, to undergo multiple biochemical changes that enable it to assume a mesenchymal cell phenotype, which includes enhanced migratory capacity, invasiveness, elevated resistance to apoptosis, and greatly increased the production of ECM components (Kalluri and Weinberg, 2009). EMT program play key roles in normal tissues such as embryonic development. It is also involved in pathogenesis of cancer and other human diseases (Baum *et al.*, 2008; Hugo *et al.*, 2007; Yang and Weinberg, 2008). Aberrant activations of EMT were usually found in tumors, where they can generate multiple, distinct cellular subpopulations contributing to intratumoral heterogeneity or promote the malignant and stem cell characteristics of cancer cells (Kang and Massague, 2004; Polyak and Weinberg, 2009). An essential difference between the embryonic and tumorigenic processes is that the tumorigenic processes involve genetically abnormal cells that progressively lose their responsiveness to normal growth-regulatory signals and possess the ability to evolve (Polyak and Weinberg, 2009).

1.4.1 Mechanisms of EMT

Several mechanisms are involved in the initiation and execution of EMT in development, and the molecular mechanisms that regulate EMT are considerably

overlapping with those that control cell motility invasion, adhesion, survival and differentiation (Thiery and Sleeman, 2006). A fundamental event of EMT is the loss of cell adhesion through the repression of epithelial marker E-Cadherin (Inui *et al.*, 2010). Repressors of E-Cadherin can be classified into two groups depending on their effects on the *E-Cadherin* promoter. Several transcription factors, either zinc finger proteins or basic helix-loop-helix factors, bind to and repress the activity of the *E-Cadherin* promoter. These factors include Snail/Slug family, ZEB, E47 and KLF8. The other group of factors such as FoxC2, E2.2, Goosecoid and Twist could repress *E-Cadherin* transcription indirectly (Thiery *et al.*, 2009; Wu and Zhou, 2010). A lot of other genes are also involved in the EMT regulation, and the molecular networks that regulate EMT are summized in Fig. 1.2.

The central role of Snail in the regulation of EMT has been underscored by recent studies. The Snail family consists of Snail1 (called SNAI1 in human), Snail2 (Slug) and Snail3 (Smuc), which are induced by TGFβ superfamily members and share an evolutionary conserved role in mesoderm formation in vertebrates (Nieto, 2002). Snail plays a fundamental role in EMT by suppressing E-Cadherin expression in mammalian cells, inducing the expression of genes associated with a mesenchymal and invasive phenotype including MMP9 and fibronectin. Snail also down-regulates the expression of other epithelial molecules such as Muc1, Claudins and Occludins (Wu and Zhou, 2010).

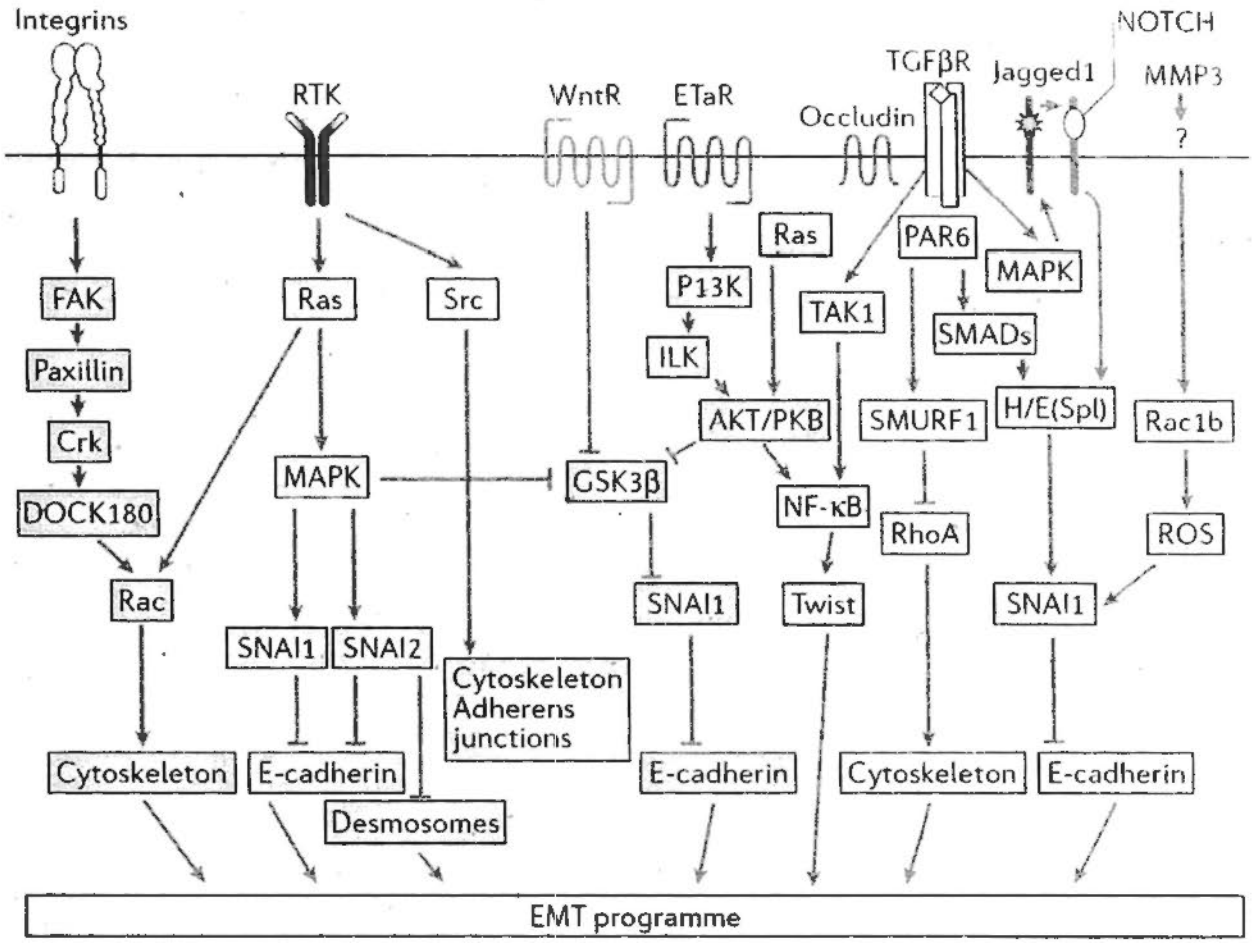


Figure 1.2 Overview of the molecular networks that regulate EMT. A selection of the signalling pathways that are activated by regulators of EMT and a limited representation of their crosstalk is illustrated. Adopted from Thiery *et al, Nat Rev Mol Cell Biol* (2006), **7**, 131-42.

1.4.2 EMT and cancer

Cells that have undergone an EMT program are known to develop an increased resistance to apoptosis, and shares many characteristics with the tumor progression (Schickel *et al.*, 2010). Indeed, EMT is associated with cancer progression and metastasis. The activation of an EMT program is proposed as the critical mechanism for the acquisition of malignant phenotypes by epithelial cancer cells (Thiery, 2002). Progression from normal epithelium to invasive carcinoma goes through several stages. The first stage is the invasive carcinoma stage, which involves epithelial cells losing their polarity and detaching from the basement membrane. The second step involves EMT and an angiogenic switch, facilitating tumor growth in the malignant phase. The third stage is the transition from the malignant phase to metastatic cancer, which again involves EMTs, enabling cancer cells to enter the circulation and exit the blood stream at a remote site, forming micro- and macro-metastases (Kalluri and Weinberg, 2009).

1.5 Regulation of cell cycle

The cell cycle is a series of events that occur in dividing cells between the completion of one mitotic division and the completion of the next division. This cycle includes accurate duplication of the genome during the DNA synthesis phase (S phase), segregation of complete sets of chromosomes to each of the daughter cells (M phase). The somatic cell cycle also contains "Gap" phases, known as G1, which connects the completion of M phase to initiation of S phase in the next cycle. During G1 phase, various enzymes and proteins, which are required in S phase, were synthesized. Dependent on environmental and developmental signals, cells in G1 may temporarily

or permanently leave the cell cycle and enter a quiescent or arrested phase known as G0. Another "Gap" phase is G2, which separates the S phase and M phase. Proteins required for mitosis were synthesized in G2 phase (Ferrell *et al.*, 2011; Maiato, 2010; Nelson *et al.*, 2002).

The process of cell cycle is highly ordered and regulated. Checkpoints exist to delay progression into the next cell cycle phase only when the previous step is fully completed (Nyberg *et al.*, 2002). Regulation of the cell cycle includes the detection, repair of genetic damage and the prevention of uncontrolled cell division. There are a number of proteins involved in the regulation of the cell cycle. Two classes of molecules, cyclin-dependent protein kinases (CDKs) and cyclins are mainly responsible for cell cycle regulation (Satyanarayana and Kaldis, 2009).

In mammalian cells, cyclins bind to cyclin dependent kinases to form complexes that are involved in regulating different cell cycle transitions. For instance, cyclin-DCDK4/6 complex are for G1 progression, cyclin-E-CDK2 are for the G1-S transition, cyclin-A-CDK2 are for S phase progression and cyclin A/B-CDC2 are for entry into M-phase (Stamatakos *et al.*, 2010).

Cells are driven into mitosis in the control of CDKs, which phosphorylate a number of important protein substrates and dictate the necessary changes in nuclear and cytoskeleton architecture. These changes lead to the condensation of DNA and the formation of individual chromosomes. It also promotes the assembly of a

microtubular mitotic spindle. The correct distribution of chromosomes during mitosis relies in their capacity to interact with the mitotic spindle, which is made possible due to the transient disassembly of the nuclear envelope in vertebrates (Maiato, 2010).

1.6 Aims of the present study

In previous study, Tumor-Induced Factor (TIF) as well as its transcript isoform SY3, was identified in xenograft tumors induced by oncogene *mas*-stably expressing CHO-K1 cells (Lin *et al.*, 2009). Sequence alignment showed that SY3 mRNA shared an identical sequence with TIF mRNA except lacking a 71 nt string in the 3'-UTR, and this 71 nt sequence was referred to as TIFm71. Genomic analysis indicated that the lost of TIFm71 in SY3 could be a result of posttranscriptional splicing. It is of interest to note that TIFm71 shared similarities with human SINE Alu-J, and a canonical pre-miRNA-like stem-loop secondary structure was predicted with a RNA folding program. Therefore, we hypothesized that TIF transcript was more than a protein-coding gene; it possibly also encodes an ncRNA or a pre-miRNA.

In the present study, we examined the expression of TIF mRNA in a variety of normal tissues and tried to find the homologies of TIF in other species including human, mouse and rat. Also, efforts were made in trying to analysis the genomic organization of TIF gene within the hamster genome (Chapter 2). To test our hypothesis that TIFm71 could be released from TIF mRNA, firstly we examined the expression of TIFm71 in normal tissues and xenograft tumors. Secondly, several fragments including M229, S229, TIF, SY3 and TIFm71 were subcloned into

mammalian expression vectors pcDNA3 and pIREshyg3. In addition, different transformants that stably over-expressing either TIFm71 or control fragments were prepared (Chapter 3). With the constructed expression vectors, we proceeded to test our hypothesis i) whether TIFm71 is released from exonic TIF transcript in living cells after transient transfection, and ii) could TIFm71 be further processed into miRNA. Moreover, protein-binding activity of TIFm71 was illustrated and its binding proteins were identified (Chapter 4). In Chapter 5, the potential biological roles of TIFm71 were characterized, including EMT and the deregulation of cell cycle distribution promoted by the over-expression of TIFm71. Finally, effects were made to find other human genes that potentially generate ncRNA from their untranslated regions.

Chapter 2

TIF Expression

2.1 Introduction

Mas is a proto-oncogene which encodes an orphan G protein-coupled receptor (GPCR) (Young *et al.*, 1986). In our previous study, we found that the over-expression of Mas facilitated tumor formation in a tumorigenicity assay (Lin *et al.*, 2009). A serial of *mas* over-expressing stable cell colonies were established, including McOM80 and VcOM80. McOM80 is CHO Dhfr deficient cells clone number 0 that stably over-expressing *mas* in the presence of 80 μM MTX, while VcOM80 is an empty vector transfected control cell of McOM80. In search for the molecular mechanisms that enabled the Mas over-expressing McOM80 cells for anchorage-independent growth and induction of xenograft tumor formation in nude mice, a novel chemokine, Tumor-Induced Factor (TIF, GenBank Accession No.EF471205) was found to be up-regulated in McOM80 when comparing with vector control VcOM80 cells by fluorescent mRNA differential display (FluorDD) (Lin *et al.*, 2009).

Full-length TIF cDNA sequence is 1.2 kb, obtained by 5' RACE cloning from the Mas over-expressing McOM80 cell line. It encodes a 101-amino acid protein with a 28 amino acids signal peptide at its N-terminus. Sequence analysis suggested that TIF is a new member of ELR$^+$ CXC chemokine. It shares very high sequence similarity with another hamster growth-regulated (GRO) gene (GenBank Accession No. J03560), which was also found to be significantly up-regulated in both *mas* over-expressing cells and xenograft tumors (Lin *et al.*, 2009). By amino acids sequence comparison, DCIP1 in mouse and CINC 2 in rat share sequence similarity with TIF, therefore we subscribe TIF into a subgroup of proinflammatory, netrophil-attractiving chemokines comprising murine DCIP-1, MIP-2, KC, rat CINC and human GRO

proteins. Functional studies revealed that TIF acted as a chemoattractant to neutrophil and promoted blood vessel formation in an aortic ring assay. An inhibitory effect on tumorigenesis was found when co-injecting TIF stable expressing cells together with embryonic fibroblasts into nude mice, suggesting that TIF has an anti-tumor effect, possibly by inhibiting fibroblast proliferation (Zhoufang, unpublished data).

In this chapter, we first examined the expression of TIF mRNA in a variety of tissues by both RT-PCR and Northern blot analysis, and then we tried to find the homologies of TIF in other species including human, mouse and rat. Finally, efforts were made in trying to analyze the genomic organization of TIF gene within the hamster genome.

2.2 Materials

2.2.1 Chemicals

DMSO (Cat. D2650) was purchased from Sigma (Steinheim, Germany). Dulbecco's Modified Eagle Medium (DMEM) powder (Cat. 12100-046), F12-Nutrient powder (#21700-075), RPMI-1640 powder (#130091-440), penicillin/streptomycin powder (#15140-122), fetal bovine serum and DNAzol Reagent (#10503-027) were purchased from Invitrogen (Carlsbad, CA, USA). Methotrexate (MTX, #M8407) was purchased from Sigma. HPLC grade isopropanol and chloroform were purchased from Scharlau (Sentmenat, Spain) and VWR International (Poole, UK), respectively. Digoxigenin-11-2'-deoxy-uridine-5'-triphosphate (DIG-11-dUTP, #11093088910), DIG-RNA molecular weight marker I (0.3-6.9 kb), DIG-DNA molecular weight marker III, Nitroblue tetrazolium chloride (NEB), 5-bromo-4-chloro-3-indolyl-phoshate, 4-toluidine salt (X-phosphate, 4-toluidine salt, BCIP), and DIG Easy Hyb (#14549300) hybridization buffer were purchased from Roche (Mannheim, Germany). Nitrocellulose transfer membrane was a product of Whatman (Dassel, Germany). [γ-^{32}P] ATP (#BLU502A250UC) was obtained from PerkinElmer (Boston, USA). All other chemicals and reagents of molecular biology grade were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2.2 Enzymes

SuperScript™ reverse transcriptase (#18064-014) was purchased from Invitrogen. Pfu DNA polymerase was purchased from Promega (Madison, WI, USA). T4-Polynucleotide Kinase (T4-PNK, #M0201S) and restriction enzymes *Eco*RI, *Eco*RV

and *Xba*I were obtained from NEB (Ipswich, UK). Anti-digoxigenin-AP (Fab fragments) was purchased from Roche (Mannheim, Germany).

2.2.3 Animals

Health adult golden hamsters (8-10 weeks, male) were ordered from the Laboratory Animal Service Center of the Chinese University of Hong Kong.

2.3 Methods

2.3.1 Cell culture methods

2.3.1.1 Thawing and freezing cells

The liquid nitrogen frozen cells were quickly placed at 37°C with gentle shaking until the entire frozen buffer was thawed. Cells were gently transferred to a fresh 1.5 ml eppendorf tube, mixed with 1 ml of culture medium, followed by spinning down at 100 xg for 5 min and the supernatant was discarded. Cells were re-suspended in culture medium according to the number of cells and transferred to an appropriate culture dish. Cells were frozen prior to confluence in order to assure the viability of the cell line. The cells were trypsinized, transferred into a 1.5 ml eppendorf tube and centrifuged at 100 xg for 5 min. Medium was discarded and 1 ml of freezing solution per 10^6 cells was added to re-suspend the cells (90% FBS, 10% DMSO). The mixture was transferred to a cryogenic vial and placed at 4°C for 30 min, at -20°C for 2 h and then at -80°C for overnight. Finally cells were transferred to liquid nitrogen for long-term conservation.

2.3.1.2 Cell culture and cell lines

HEK293 cells obtained from the American Type Culture Collection (ATCC, Rockville, MD) were grown in DMEM supplemented with 10% FBS at 37°C in 5% CO_2 humidified incubator. Human HepG2, CaCo, A431, SiHa and rat UMR-106 cells were obtained from ATCC and grown in DMEM supplemented with 10% FBS; human MCF-7 cells from ATCC were grown in DMEM supplemented with 10% FBS and 0.01 mg/ml bovine insulin; human ZR-75-1, LNCaP.FGC and Raji cells from

ATCC were grown in RPMI supplemented with 10% FBS; human Phoenix Ampho cells from Orbigen (San Diego, CA) were grown in DMEM supplemented with 10% FBS. McOM80 cells (CHO Dhfr cells clone number 0 stably over-expressing pFRSV-*Mas*) were grown in IMDM medium supplemented with 10% FBS and 80 μM of MTX.

2.3.2 Determination of TIF gene expression in hamster tissues by RT-PCR

2.3.2.1 Tissue preparation

Fresh whole blood (~0.3 ml) was collected in a clear microtube in the presence of 0.2 volume of ACD anti-congregation solution (50 mM Sodium Citrate, 25 mM Citric Acid and 80 mM Glucose) and 1 volume of pre-warmed 3% gelatin (m/v, dissolved in physiological saline). The whole blood cells were kept still in room temperature for 1 h to allow red blood cells settling down. Supernatant was transferred into a new microtube and centrifuged at 1,000 xg for 10 min to collect cell pellet, which contained all the leukocytes and a small amount of red cells. The pellet was washed with PBS twice, followed by RNA extraction using TRIzol (Invitrogen, USA) reagent. Total RNA of 7 tissues, including kidney, liver, lung, testis, heart, spleen and cerebral cortex, were isolated using TRIzol reagent according to the manufacturer's instructions. Total RNA from McOM80 cells was used as a positive control.

2.3.2.2 Total RNA isolation

Fresh tissues were homogenized with a hand-held tissue homogenizer in the present of 1 ml TRIzol/100 mg tissue until the tissue is completely dissolved in the solution.

Cells grown in 100 mm dish (~$1x10^7$) were trypsinized, rainsed with PBS, and lysed in 1 ml of TRIzol reagent by pipetting for several times. The tissue or cell lysate was incubated in room temperature for 5 min, and 0.2 ml of chloroform was added, followed by vigorous shaking or vortex for 15 sec. After keeping in room temperature for 3 min, the lysate was centrifuged at 12,000 xg for 12 min at 4°C, and two divided aqueous phases were formed. The upper aqueous phase was transferred into a new eppendorf tube and mixed with 0.5 ml of isopropanol. The mixture was kept in room temperature for 10 min for RNA precipitation, followed by spinning down at 12,000 xg for 10 min at 4°C. RNA pellet was washed once with 75% ethanol (v/v, diluted with DEPC treated water) and then centrifuged at 7,500 xg for 5 min at 4°C. The RNA pellet was air-dried and dissolved in 20-50 µl DEPC-treated water.

2.3.2.3 Quantification of DNA and RNA

Given that DNA solution at 50 µg/ml gives an absorbance at 260 nm (OD260) equals to 1, the concentration of DNA sample = OD260 × dilution factor × 50 µg/ml. Given that RNA solution at 40 µg/ml gives an absorbance at 260 nm equals to 1, the concentration of RNA sample = OD260 × dilution factor × 40 µg/ml. The concentration of all RNA samples were adjusted to 2 µg/µl by adding DEPC treated water.

2.3.2.4 Reverse transcription

Total RNA was reversely transcribed using SuperScript™ II Reverse Transcriptase (Invitrogen, USA). A reaction mixture containing 5 µg of total RNA, 1 µl of oligo

(dT), and 1 μl of 10 mM dNTP were mixed to a final volume of 13 μl, heated at 65°C for 5 min. Afterwards, 4 μl of first strand buffer, 2 μl of 0.1 M DTT and 0.3 μl of supertranscript reverse transcriptase were added to the mixture. The mixture was then incubated for 50 min at 42 °C and 15 min at 70°C. The cDNA products obtained were stored at -20°C until use.

2.3.2.5 Polymerase Chain Reaction (PCR)

PCR reaction was run for 30 cycles at denaturing temperature 94°C for 30 sec, an annealing temperature 62°C for 30 sec and an extension temperature 72°C for 30 sec. At the end of the amplification cycles, an extended incubation at 72°C for 7 min was added. The sequence of forward (F) and reverse (R) primers and length of amplicons (in parentheses) obtained were: TIF F: GCT CCT GTG CTC CAG ACT T, R: CGG AAT TCC GGG CCT TAG GCA GGA TCA CTT (306 bp); Gro F: ACA TCC AGA GCT TGA AG GT, R: AAC ACA TCC ACA ATG TTA AAT A (333 bp); Tubulin F: GGA ATG GAT CCC CAA CAA, R: TCC TGG TAC TGC TGA TAC (255 bp).

2.3.3 Northern blot

Northern blot was applied to examine the homology of TIF mRNA in cancer cell lines of other species rather than hamster, including human, rat and mouse. Total RNA was isolated from cultured cancer cell lines, separated in formaldehyde agarose gel, transferred onto a positive charged Nylon membrane and hybridized with a DIG-labeled full-length TIF DNA probe.

2.3.3.1 Preparation of RNA samples

RNA samples were adjusted to 2 µg/µl and 10 µl of each RNA (20 µg) sample was mixed with 2 µl of 10x MOPS (200 mM 4-morpholinopropane- sulfonic acid, 50 mM sodium acetate, 10 mM EDTA, pH 7.0), 5.7 µl of 37% formaldehyde, 0.3 µl of ethidium bromide solution (10 mg/ml) and 2 µl of 10x loading dye (0.21% bromophenol blue, 0.21% xylene cyanol FF, 50% glycerol, 0.2 M EDTA, pH 8.0) to give a final volume of 20 µl. The RNA samples mixture were denatured in boiling water for 10 min and chilled on ice immediately.

2.3.3.2 Preparation of formaldehyde agarose gel and electrophoresis

To prepare 1% formaldehyde agarose gel, 1.8 ml of 37% formaldehyde, 10 ml of 10x MOPS (200 mM 4-morpholinopropane- sulfonic acid, 50 mM sodium acetate, 10 mM EDTA, pH 7.0) and 1 g of agarose were added to 100 ml of autoclaved water. The agarose was dissolved by heating in a microwave oven and poured into a gel-cast inside a fume hood cabinet. The gel was kept still for at least 1 h before RNA sample loading. RNA samples, as well as 5 µl of DIG-labeled RNA Molecular Weight Marker I (20 ng/µl, #11526529910, Roche) were loaded into the wells of the 1% formaldehyde agarose gel. The electrophoresis was set at a constant voltage of 10 V :m and allowed to run for ~150 min in 1x MOPS running buffer until the bromophenol blue dye front migrated almost to the bottom of the gel.

2.3.3.3 Capillary transfer

After electrophoresis, the gel was exposed under UV light to assess the integrity of RNA and photos were taken. Then the gel was washed with an excess volume of 10x SSC buffer for 20 min with gentle shaking. During the washing, positively charged nylon membrane was cut to the same dimensions as the gel (~12cm x 14cm) and equilibrated in 10x SSC at least for 10 min. The upward capillary transfer was set up within a plastic box containing 10x SSC transfer buffer and a stack of gel blotting paper in the bottom. On top of the gel blotting paper, the following items were set up in order: three pieces of Whatman 3 MM paper, 1% formaldehyde agarose gel with RNA (upside down), pre-equilibrated positively charged nylon membrane, another three pieces of Whatman 3 MM paper and paper towels (~15 cm in height). All air-bubbles were carefully smoothed out with a glass pipet. Edges of gel were covered with Parafilm to prevent the transfer buffer being absorbed directly into the paper towels. Finally a glass plate was placed on top of the paper stack and a weight of ~1 kg was put on the glass plate to allow capillary transfer. The transfer was allowed to proceed overnight (~16 h). Afterwards, the membrane was first rinsed with 5x SSC for 5 min with gentle shaking, air-dried, then UV cross-linked by exposing to a total energy of 0.12 J/cm^2 for 3 min and baked for 2 h at 80°C.

2.3.3.4 Preparation of DIG-labeled probe

The DIG-labeled probe was prepared by incorporating DIG-dUTP (#11093088910, Roche) into the nascent TIF sequence during PCR amplification. The PCR was carried out with 100 ng of DNA template, 5 µl of 10x PCR buffer, 3 µl of 25 mM $MgCl_2$, 5 µl of PCR DIG-labeling mix (DIG-dUTP: dTTP = 1:19), 1 µl each of 10 µM forward primer and 10 µM reverse primer, 1 µl of Taq DNA polymerase (5 U/µl)

in a final volume of 50 µl. Meanwhile, a parallel reaction, in which 1 µl of normal dNTP was added instead of 5 µl of PCR DIG-labeling mixture. was set up as a control. Finally, both the labeled and not labeled PCR products were analyzed in a 1.2% agarose gel in the presence of 0.5 µg/ml ethidium bromides, quantified by UV-spectrometryand and stored at -20°C until use. The full-length TIF probe was a 1147 bp fragment that was amplified by PCR using the oligonucleotide pair 5'-GCT CCT GTG CTC CAG ACT T-3' and 5'-CCT TTA ATC CCA GCA CTC AG-3'. The TIFm71 probe was a 71 bp fragment that was amplified by PCR using oligonucleotide pair 5'-TTT CTC TGT GTA GCT TT-3' and 5'-CTC TGT GAG TTT GAG GC -3'.

2.3.3.5 Pre-hybridization and hybridization

The baked membrane was rinsed once with 2x SSC and transferred to a baked hybridization tube. Pre-hybridization was carried out by adding of 7 ml of Dig Easy hyb buffer (#11603558001, Roche) with gentle rotation at 42°C. After 2 h of pre-hybridization, 50 µl of DIG-labeled DNA probe was boiled for 10 min and chilled on ice immediately for 5 min. The boiled DNA probe was then added to 10 ml of fresh hybridization buffer to give a final concentration of 10 ng/ml. The pre-hybridization buffer was discarded and replaced with the fresh hybridization buffer containing the DIG-labeled probe. The blot was then incubated in a hybridization machine at 42°C for 16 h with gentle rotation.

2.3.3.6 Post-hybridization washing, blocking and detection

After the hybridization was over, the membrane was first washed with 2x SSC/0.1% SDS for 10 min at room temperature twice, then washed with 0.5x SSC/0.1% SDS for 10 min at 60°C twice. After the washings, the membrane was rinsed once with washing buffer (0.1 M maleic acid, 0.15 M NaCl, Tween 20, pH 7.5), then incubated with 1x blocking buffer (1% w/v Blocking Reagent in maleic acid buffer, #11096176001, Roche) with gentle shaking for 1 h. After blocking, anti-digoxingenin-AP (Fag fragments) was added to a final concentration of 1:10,000 and incubated for another 1 h at room temperature with gentle shaking. Unbound anti-Dig antibody was removed by washing the membrane with washing buffer (0.1 M maleic acid, 0.15 M NaCl, Tween 20, pH 7.5) for 10 min with gentle shaking, twice. The membrane was then equilibrated with detection buffer (0.1 M NaCl, 0.1 M Tris-HCl pH 9.5) for 5 min. After that, the membrane was transferred to a clean baked glass tray and developed by adding the NBT/BCIP substrate solution (0.1 M NaCl, 0.1 M Tris-HCl pH 9.5, 0.05 M $MgCl_2$, 0.5 mg/ml NBT, 0.19 mg/ml BCIP). The glass tray was covered with a tin foil and the membrane was allowed to develop overnight until the signals were detected.

2.3.3.7 Stripping and de-probing

To strip developed color, the membrane was first incubated with N'-N' dimethylformamide in a hybridization bag at 65°C until all the color was removed. The membrane was then rinsed with an excess amount of water for 10 min at room temperature with gentle shaking, followed by adding 0.1% boiling SDS solution and incubated with shaking at room temperature for 20 min to de-probe. After de-probing,

the membrane was rinsed once with 2x SSC, air-dried and stored at 4°C until next hybridization as described above.

2.3.4 Southern blot analysis

Localization of particular sequences within genomic DNA is usually accomplished by the Southern blot analysis. Genomic DNA is digested with one or more restriction enzymes, and the resulting fragments are separated by electrophoresis through an agarose gel. The DNA is then denatured in situ and transferred from the gel to a nitrocellulose filter or nylon membrane. The relative positions of the DNA fragments are preserved during their transfer to the filter. DNA attached to the filter is hybridized to DIG-labeled or isotope-labeled probes.

2.3.4.1 Preparation of genomic DNA from cultured cells

Wild type CHO-K1 cells were grown to confluence in 100 mm culture dishes with F12 medium supplemented with 10% FBS. The cells were trypsinized, washed with PBS once and lysed by adding 1 ml of DNAzol Reagent (#10503-027, Invitrogen), with repetitive pipetting using a micropipette. DNA was precipitated from the lysate by addition of 0.5 ml of 100% ethanol, mixed by inverting for several times and kept in room temperature for 1-3 min. DNA quickly became visible as a cloudy precipitate. The sample was then centrifuged at 4,000 xg for 2 min to pellet the DNA, followed by washing with 75% ethanol twice. At each wash, the DNA pellet was suspended by inverting the tube for 3-6 times. Finally, the DNA pellet was air-dried, resuspended in

8 mM NaOH. The concentration of DNA solution was determined by the absorbance at 260 nm and 280 nm.

2.3.4.2 Preparation of probe

The DIG-labeled probe preparation protocol was described in section 2.3.3.4. The full-length genomic TIF probe (including introns) was a 1870 bp fragment amplified by PCR using the oligonucleotide pair 5'-GCC ATG GCC CCA GTC AC-3' and 5'-CCT TTA ATC CCA GCA CTC AG-3'. The open reading frame (ORF) region of *mas* probe was a 921 bp fragment PCR-amplified using the oligonucleotide pair 5'-CCG GAA TTC GAT GGA CCA ATC AAA TAT GAC-3' and 5'-TTG CAT CTC GTC TTT GAA AGC TCT GGT-3'.

Isotope-labeled probe was prepared by 5'-labeling of [γ-^{32}P] ATP with T4-Polynucleotide Kinase (T4-PNK, #M0201S, NEB). TIF sequence was generated by PCR as described in section 2.3.3.4, purified by adding equal volume of isopropanol, kept on ice for 20 min, and then precipitated the DNA pellet by centrifuge at 1,2000 rpm for 10 min under 4°C. The DNA pellet was washed once with 75% ethanol, air-dried, and then dissolved in distilled water. Probe concentrations were quantified by UV-spectrometry. The isotope-labeling mixture was set up by adding 200 ng of TIF sequence, 1.5 µl of T4-PNK (10,000 U/ml), 5 µl of 10x T4-PNK buffer and 2 µl of [γ-^{32}P] ATP (10 µCi/µl). The total volume was made up to 50 µl by the addition of distilled water. The mixture was incubated in 37°C for 1 h and purified by passing

through CHROMA SPIN +TE-30 Column (Clontech). The probe was stored at -20°C until use.

2.3.4.3 Enzymatic restriction and gel electrophoresis

Genomic DNA (20 μg) extracted from CHO-K1 cells was digested with *Eco*RI, *Eco*RV or *Xba*I restriction enzymes. The reaction mixture contained *Eco*RI (2 μl, 20,000 U/ml) in 1x NEB *Eco*RI buffer (100 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 0.025 % Triton X-100, pH 7.5); *Eco*RV (2 μl, 20,000 U/ml) in NEBuffer 3 (50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9) and 1 mg/ml BSA; or *Xba*I (2 μl, 20,000 U/ml) in 1x NEBuffer 4 (20 mM Tris-acetate, 50 mM potassium acetate, 10 mM Magnesium Acetate, 1 mM Dithiothreitol, pH 7.9) and 1 mg/ml BSA were mixed. The final volume was made up to 50 μl by adding sterile water. The digestion systems were kept in 37°C water bath overnight. The next day, digested samples, as well as equal amount of not digested DNA, 5 μl of DIG-labeled DNA Molecular Weight Marker III (#10528552001, Roche), were loaded into a 0.8% agarose gel. The electrophoresis was set at 120 V, and usually it took 150 min to finish the running.

2.3.4.4 DNA transferring to positive charged Nylon membrane

Following electrophoresis, the DNA samples in the gel were visualized under UV illuminant. The gel was soaked in 0.125 M HCl and agitated gently for ~10 min, until the bromophenol blue band changed its color into light yellow. The gel was then rinsed with sterile distilled water once and placed into sufficient denature buffer (0.5

M NaOH, 1.5 M NaCl) with gentle agitation for 30 min. Then gel was rinsed with sterile water and placed in sufficient neutralization buffer (0.5 Tris-HCl, 1.5 M NaCl, pH 7.5) with gentle agitation for 30 min and rinsed with sterile water. After neutralization, the gel was incubated with 10x SSC buffer for 10 min with gentle shaking. Positively charged nylon membrane was cut to the same dimensions as the gel and equilibrated in 10x SSC for at least 10 min. The upward capillary transfer was set up by placing a stack of gel blotting paper in a plastic box containing 10x SSC transfer buffer, three pieces of Whatman 3 MM paper, gel (upside down), pre-equilibrated positively charged nylon membrane, three pieces of Whatman 3 MM paper and paper towels (~15 cm in height). All air bubbles were carefully smoothed out with a glass pipet. Edges of gel were covered with Parafilm to prevent the transfer buffer being absorbed directly into the paper towels. Finally, a glass plate was placed on top of the paper stack and a weight of ~1 kg was put on top of the glass plate. The transfer was allowed to proceed overnight. The next day, the membrane was first rinsed with 5x SSC for 5 min with gentle shaking, air-dried, and then UV cross-linked by exposing to a total energy of 0.12 J/cm^2 for 3 min and baked for 2 h at 80°C.

2.3.4.5 Pre-hybridization and hybridization with DIG-labeled probe

Pre-hybridization and hybridization were performed as described in section 2.3.3.5.

2.3.4.6 Pre-hybridization and hybridization with [γ-^{32}P] ATP-labeled probe

The baked membrane was rinsed once with 2x SSC and transferred to a baked hybridization tube. Pre-hybridization was carried out by the addition of 7 ml of

Rapid-hyb buffer (RPN163, Amersham) with gentle rotation at 42°C. After 2 h of pre-hybridization, ~50 µl of purified [γ-^{32}P] ATP-labeled probe was heated for 10 min at 95°C and chilled on ice immediately for 5 min. The probe was then added directly into the hybridization buffer in the hybridization tube, and additional 3 ml of Rapid-hyb buffer was added to make the total hybridization volume to 10 ml. The blot was then incubated in a hybridization machine at 42°C for 16 h with gentle rotation. After the hybridization was over, the hybridization buffer was discarded into a ^{32}P liquid waste container. The blot was firstly washed with 2x SSC/0.1% SDS for 10 min at room temperature twice, then washed with 0.5x SSC/0.1% SDS for 10 min twice at 60°C. After washing, the blot was wrapped with saran wrap and exposed overnight to X-ray film at -70°C.

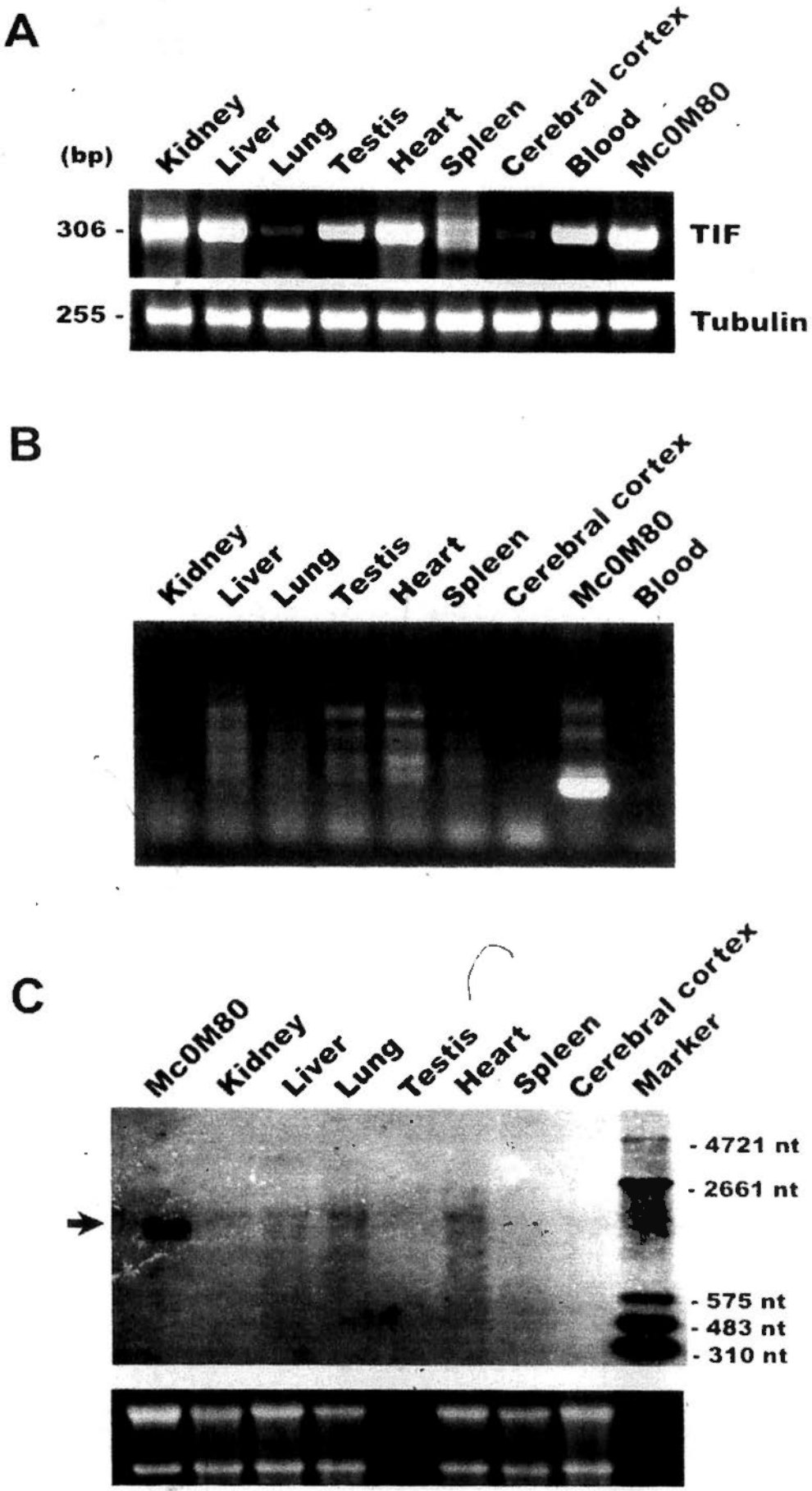
2.4 Results

2.4.1 Tissue distribution of TIF mRNA

Chemokines are critical mediators of cell migration during routine immune surveillance, inflammation, and development (Allen *et al.*, 2007). A novel member of CXC chemokine family, TIF, was identified in hamster. Expression pattern was examined in various tissues by RT-PCR (Fig. 2A). RT-PCR results showed that TIF mRNA was ubiquitously expressed, and high expression was detected in kidney, liver, heart, testis and blood cells, suggesting an important function of TIF in these tissues. Mc0M80 is an oncogenic stable cell line, which expresses a relatively high level of TIF mRNA (Lin *et al.*, 2009), was used as a positive control (Fig. 2.1A). *Gro* gene, a homology of TIF in hamster, was not detected using a pair of primers specific to *Gro* gene, indicating the specificity of RT-PCR (Fig. 2.1B).

To further confirm the RT-PCR result, total RNA isolated from normal adult hamster tissues was subjected to Northern blot analysis to examine the TIF mRNA expression pattern. DIG-labeled full-length TIF DNA (~1.2 kb) was used as probe. Northern blot analysis revealed that the weak expression of ~1.2 kb, which was the expected size of TIF mRNA, were detected in tissues such as liver, lung and heart (Fig. 2.1C). However, no signal was detected in tissues like kidney and testis, probably because of the expressing level in these tissues was under the detectable capacity of DIG-labeled probe.

THIS IS A BLANK PAGE


A
(bp)
Kidney
Liver
Lung
Testis
Heart
Spleen
Cerebral cortex
Blood
McOM80
306 -
TIF
255 -
Tubulin
B
Kidney
Liver
Lung
Testis
Heart
Spleen
Cerebral cortex
McOM80
Blood
C
McOM80
Kidney
Liver
Lung
Testis
Heart
Spleen
Cerebral cortex
Marker
- 4721 nt
- 2661 nt
- 575 nt
- 483 nt
- 310 nt

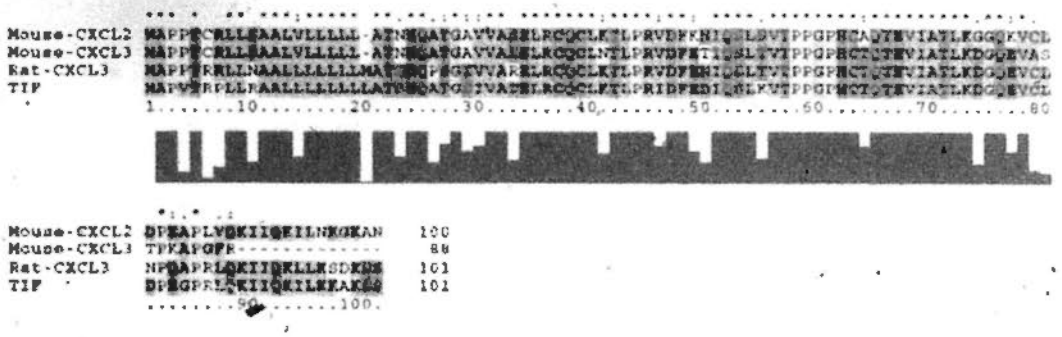
2.4.2 Homologue gene of TIF was not detected in human or rat cell lines

Blast analysis of TIF mRNA sequence in NCBI database indicated that the mouse gene CXCL3 shared the highest similarity with TIF, with identities equal 78%. When we use the blastp algorithm for the protein blast, several proteins were found to be the homologues of TIF, including mouse CXCL3 (mDCIP1), CXCL2 (MIP2) and rat CXCL3 (CINC2), with identities equal ~90% (Fig. 2.2A). Phylogenetic analysis of CXC chemokines also showed that TIF was closest to rat CXCL3 (Fig. 2.2B), suggesting TIF may be the hamster homologue of rat CXCL3 (Lin *et al.*, 2009).

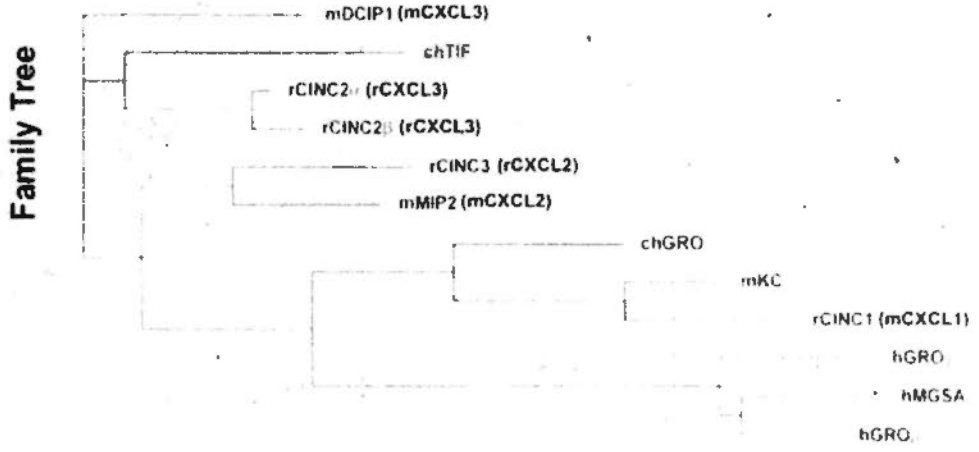
To examine the expression of the TIF-like transcript in cancers, a panel of cancer cell lines was analyzed by Northern blot using a DIG-labeled full-length TIF probe (Fig 2.2C). A high expression level of the ~1.2 kb TIF fragment was detected in xenograft tumor induced by the inoculation of Mc0M80 cells. Although the expression of TIF was only detected in *mas*-transfected but not in empty vector-transfected cells, expression of TIF was also detected in xenograft tumor derived from empty-vector transfected Vc0M80 cells. However, no TIF-like transcript was detected in human mammary adenocarcinoma MCF-7, human mammary ductal carcinoma ZR-75-1, human hepatocarcinoma HepG2, human epidermoid carcinoma A431, human colorectal adenocarcinoma CaCo, human cervix squamous carcinoma SiHa, human prostate carcinoma LNCap-FGC, human lymphoma Raji and rat osteosarcoma UMR-106 cells. Neither TIF-like transcript was detected in the transformed human embryonic kidney Pheonix Ampho cells. These results implied that the expression of TIF was detected only in tumor tissues and/or a specific subset of tumor cells. Possibility cannot be excluded that there is no human homologue of TIF.

THIS IS A BLANK PAGE

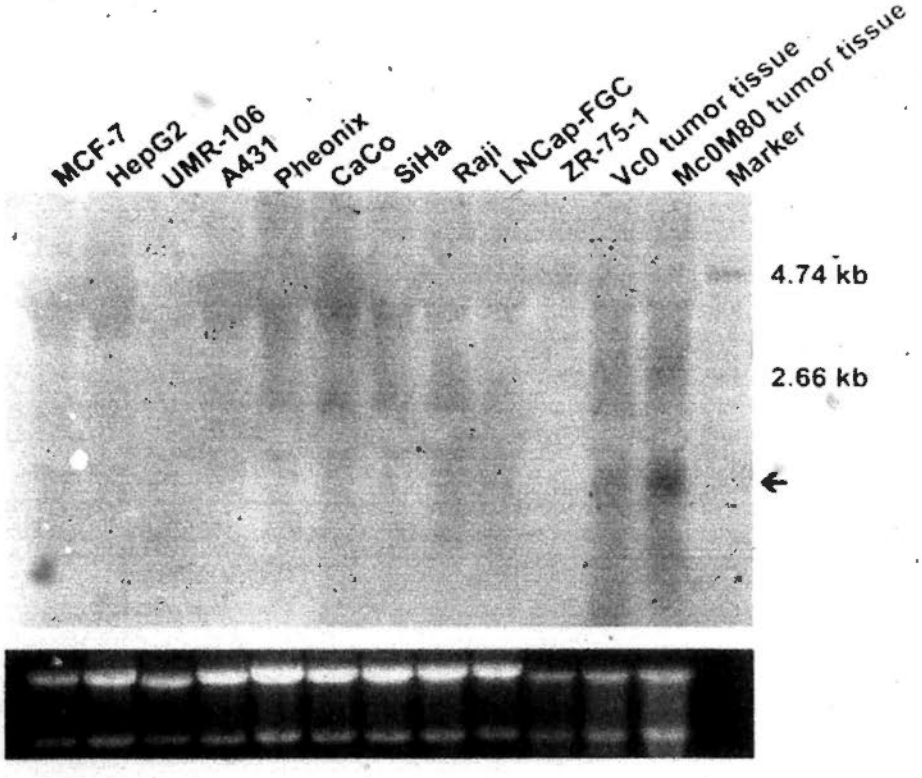
A


Mouse-CXCL2
Mouse-CXCL3
Rat-CXCL3
TIF

B


Family Tree
mDCIP1 (mCXCL3)
chTIF
rCINC2α (rCXCL3)
rCINC2β (rCXCL3)
rCINC3 (rCXCL2)
mMIP2 (mCXCL2)
chGRO
mKC
rCINC1 (mCXCL1)
hGRO
hMGSA
bGRO

C


MCF-7
HepG2
UMR-106
A431
Pheonix
CaCo
SiHa
Raji
LNCap-FGC
ZR-75-1
Vc0 tumor tissue
Mc0M80 tumor tissue
Marker
4.74 kb
2.66 kb

2.4.3 Genomic TIF examined by Southern blot

TIF transcript containing introns was amplified from the genome DNA of Chinese hamster ovary CHO-K1 cells. One single band that spanned ~2 kb was sequenced. As shown in Fig 2.3, sequencing results showed that TIF transcript consisted of four exons ranging between 84 and 842 bp in size, together with three introns. All exon-intron junctions conformed to the GT/AG rule established for eukaryotic genes (Shi *et al.*, 1999). The open reading frame (ORF) region of TIF spanned through exon 1 to exon 4 (Zhoufang, unpublished data).

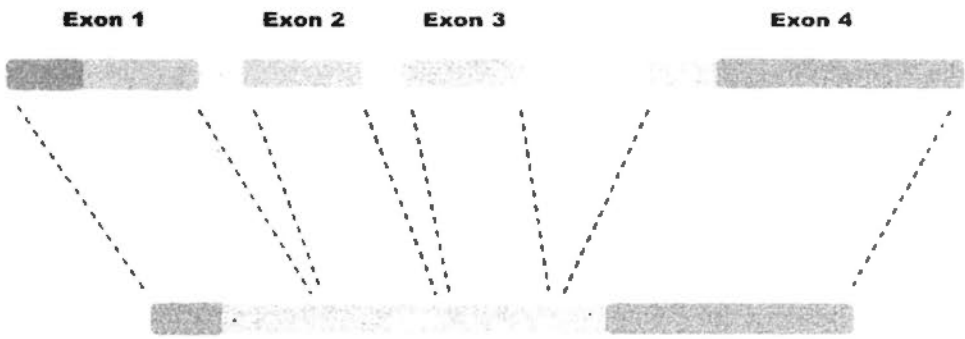
As for the whole genome map of hamster is not available yet, we tried to examine the genomic organization of TIF gene by Southern blot analysis, expecting to identify the promoter region and its neighbor genes. Southern blot analyses were performed with both DIG-labeled and [γ-^{32}P] ATP labeled TIF probe. Genomic DNA (20 μg) isolated from CHO-K1 cells was digested with restriction enzymes *Eco*RI, *Eco*RV or *Xba*I. DNA fragments were separated with a 0.8% agarose gel and transferred to positive charged Nylon membrane. About 0.1 ng of genomic TIF DNA, which was amplified by PCR, was loaded as a positive control. However, we failed to get concrete bands hybridized by DIG-labeled probe but a smear-like binding range from ~500 bp to ~21,000 bp (Fig. 2.4A).

To rule out the possibility that DIG-labeled was not sensitive enough, we tried [γ-^{32}P] ATP labeled probe (Fig. 2.4B). The result was more or less the same as that shown in Fig. 2.4A. Reducing the loading amount of genomic DNA to 10 μg or 5 μg did not

help getting any discrete banding either (last two lanes of Fig. 2.4B). We also tried using different parts of TIF gene as the hybridization probes, but still, the results were similar to those previous ones.

In order to test whether it was caused by technique problems of Southern blot, a parallel experiment was performed to detected *mas* gene in Mc0M80 cells using DIG-labeled ORF of *mas* probe. As shown in Fig 2.4C, we successfully detected the DNA fragment containing *mas* (indicated by arrow). These results implying that the background in the blots detected with TIF probe was probably due to TIF gene itself.

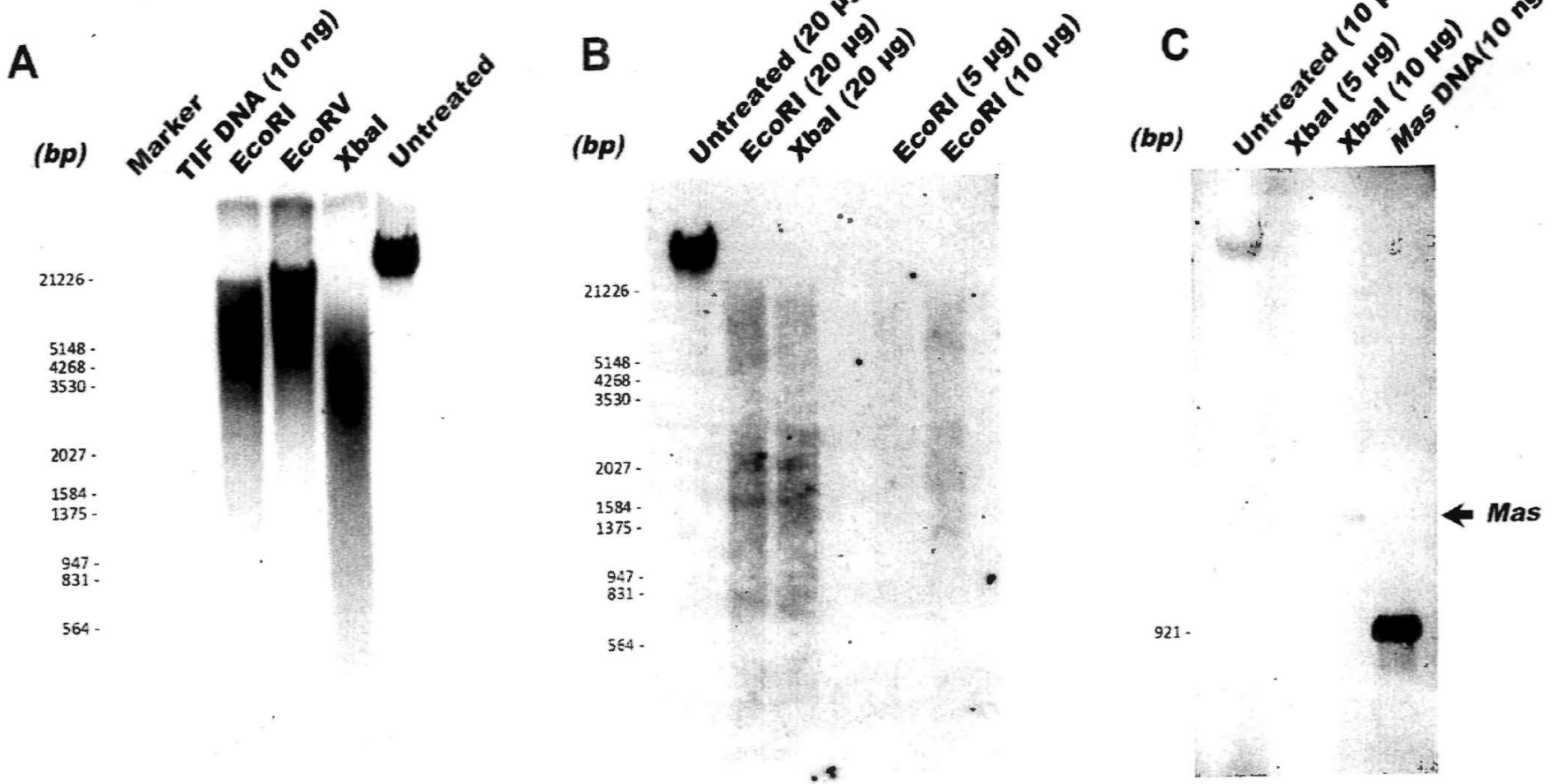
A



B

Figure 2.3 Organization and sequence of TIF gene. (A) Schematic diagram showing the genomic organization of TIF gene. It consisted of four exons and three introns. The ORF sequence was shown in grey and the untranslated regions were indicated by dark red. (B) Sequence of TIF gene. Four exons were highlighted with yellow and start/stop codons ("ATG " and "TGA") were indicated with red color.

THIS IS A BLANK PAGE


A
(bp)
Marker
TIF DNA (10 ng)
EcoRI
EcoRV
XbaI
Untreated
21226 -
5148 -
4268 -
3530 -
2027 -
1584 -
1375 -
947 -
831 -
564 -
B
(bp)
Untreated (20 µg)
EcoRI (20 µg)
XbaI (20 µg)
EcoRI (5 µg)
EcoRI (10 µg)
21226 -
5148 -
4268 -
3530 -
2027 -
1584 -
1375 -
947 -
831 -
564 -
C
(bp)
Untreated (10 µg)
XbaI (5 µg)
XbaI (10 µg)
Mas DNA(10 ng)
921 -
Mas

2.5 Discussion

After a proper stimulation, either by proinflammatory cytokines or even with whole bacteria, a wide variety of cell types secret chemokines, which exert their effects on one or more target cell populations (Visser *et al.*, 1998). It was reported that several other cell types, including monocytes (Olsnes *et al.*, 2009) and macrophages (Goodman *et al.*, 1998) can produce chemokines without stimulation. Tumor cells can also secrete chemokines that act as autocrine growth factors or enhance metastasis (Kershaw *et al.*, 2002). However, there is no report on the chemokines expression in normal tissues.

By both RT-PCR and Northern blot analysis, we found that TIF is a broadly expressed gene. According to the RT-PCR results, the expression of TIF mRNA was detected in a wide variety of normal tissues, including kidney, liver, testis, heart, blood cells and so on. The blood cells expressed TIF mRNA in a relatively high level compared to other tissues. Some tissues such as heart and liver containing plenty of blood vessels inside, therefore it was possible residual blood cells would affect the RT-PCR result in these tissues. From the result of Northern blot analysis, TIF expression level was actually quite low when compared to that in Mc0M80 cells, and even no expression was detected in several of the tissues examined. Blood cells were not included in the Northern blot analysis in the examination of TIF expression because of the difficulty in getting adequate total RNA. Our result indicated that the roles of chemokines might not restrict to inflammatory or tumor, it probably has a fundamental function under physiological conditions.

Blast analysis of TIF cDNA sequence and amino acid sequence indicated that mouse CXCL3, CXCL2 and rat CXCL3 might be the homologues of TIF. However, no TIF-like transcript was detected in a variety of tumor cells derived from human and rat, probably because of the differences between TIF and TIF homologues were too large.

DNA sequences upstream of transcriptional start points usually encompass most of the regulatory elements that control gene expression (Suzuki *et al.*, 2001). In an attempt to examine the genomic organization of TIF by Southern blot, we failed to detect discrete bands with either DIG-labeled or [γ-^{32}P] ATP-labeled probe. Indeed, we obtained a smear-like bindings range from ~500 bp to ~21,000 bp. We performed a parallel experiment to detect proto-oncogene *mas*, and the result turned out to be successful. Therefore we believed that the background in the blots detected with TIF probe was probably because of the gene itself. A possibility is that TIF contains some conserved motifs, which are highly repeated in the genome.

Chapter 3

Construction of TIFm71 Expression Vectors and Stable Cell Lines

3.1 Introduction

Growing evidence suggests that non-coding RNAs (ncRNA), including microRNA (miRNA), are implicated in almost all known cellular processes and regulate diverse cellular functions (Mattick, 2009). Studies on the biogenesis of small RNAs in animals revealed that approximately 50% of mammalian miRNAs locates in the introns of protein-coding genes or long ncRNAs transcripts, whereas the remainder are independent transcription units with specific core promoter elements and polyadenylation signals (Kim and Nam, 2006). Among the intragenic miRNAs, 40% are in the introns of protein-coding genes, whereas ~10% are in the introns of long ncRNA transcripts (Di Leva and Croce, 2010; Kim *et al.*, 2009). However, there is no experimental evidence on miRNAs that are derived from the exonic region of protein coding genes.

Besides full-length transcript, an isoform of TIF named SY3 was also identified to be up-regulated in Mc0M80 cells. SY3 gene shares identical sequence with TIF except lacking a 71 nt string in the 3' untranslated region (3'-UTR), and this 71 nt string was referred as TIFm71. It was of interest to note that TIFm71 was predicted to fold into a pre-miRNA-like stem-loop structure by RNA folding programs (Gruber *et al.*, 2008). Moreover, sequence homologs were found using BLAST against miRNA database. Therefore, we hypothesized that TIF transcript was more than a protein-coding gene, and it may also encode a small ncRNA or a pre-miRNA TIFm71.

To test our hypothesis, we first examined the expression of TIFm71 in normal tissues and xenograft tumors. Moreover, several fragments including M229, S229, TIF, SY3 and TIFm71 were subcloned into pcDNA3 and pIREshyg3 expression vectors. Both pcDNA3 and pIREshyg3 carrying CMV promoter but with different antibiotics resistance genes, which provided two selection methods in stable cell line construction using either G418 (pcDNA3) or Hygromycin B (pIREShyg3). M229 was a fragment including TIFm71 as well as flanking sequences in both 5' and 3', while S229 contained the same flanking sequence with M229 but a scrambled TIFm71 sequence to avoid the formation of a stem-loop secondary structure. Also, CHO-K1 cells that over-express M229 and S229 were constructed.

3.2 Materials

3.2.1 Chemicals

Iscove's modified DMEM medium, F12-Nutrient Mixture (#21700-075), fetal bovine serum (FBS, #10270-106), penicillin/streptomycin (#15140-122), trypsin, TRIzol reagent, Lipofactamine 2000 (#11668-019), all DNA primers, dNTP, 1kb plus DNA ladder and agarose (electrophoresis grade) were purchased from Invitrogen (Carlsbad, CA, USA). ULTRAhyb®-Oligo Hybridization Buffer (#AM8663) was purchased from Ambion (CA, USA). G418 were from Merck Biosciences (Whitehouse Station, NJ, USA). Hygromycin B (#843555), Nitroblue tetrazolium chloride (NBT), 5-bromo-4-chloro-3-indolyly-phosphate-4-toludine salt (BCIP), Digoxigenin-11-2'-deoxy-uridine-5'-triphosphate (DIG-dUTP, #11093088910), DIG Easy Hyb (#14549300) hybridization buffer and DNAse-free RNAse (#11 579 681 001) were purchased from Roche (Roche Diagnostics Corporation, IN, USA). Nitrocellulose transfer membrane was from Whatman (Dassel, Germany). All other chemicals and reagents of molecular biology grade were purchased from Sigma-Aldrich (St. Louis, MO, USA).

3.2.2 Enzymes

All restriction enzymes including *Bam*HI, *Xba*I and *Bst*XI were purchased from New England BioLabs (Ipswich, UK). T4 DNA ligase and SuperScriptTM were purchased from Invitrogen Corporation (Carlsbad, CA, USA). Anti-digoxigenin-AP (Fab fragments) was purchased from Roche (Mannheim, Germany).

3.2.3 Vectors

pcDNA3 cloning and expression vector was kindly provided by Prof. TT. Kwok. pIREShyg3 (#631621) cloning and expression vector was purchased from Clontech laboratories, Inc (CA, USA).

3.2.4 Kits and Instruments

Gel extraction kit was from GE Healthcare (UK). GeneAmp PCR system 9700 was a product of Applied Biosystem. DU-640 spectrophotometer was manufactured with Beckman Coulter Inc. (#4326045, USA). QIAquick gel extraction kit, Qiagen plasmid midi kits were from Qiagen (Hilden, Germany).

3.2.5 Animals

Health adult golden hamsters (8-10 weeks, male) and nude mice (20-25 g, male) were from the Laboratory Animal Service Center of the Chinese University of Hong Kong. Tumorigenesis assay on nude mice was performed in animal house.

3.3 Methods

3.3.1 Northern blot analysis

Total RNA was isolated from cells, xenograft tumors or fresh tissues using TRIzol reagent as described in section 2.3.2.2. To detect the expression of TIF and TIFm71 in normal tissues, 20 μg of total RNA from each sample was separated on 1% denaturing agarose gel, blotted to membrane, and hybridized to DIG labeled probe overnight at 42°C in DIG Easy Hyb hybridization buffer (Roche). Full-length TIF probe, a 1147 bp fragment, was amplified by PCR using the oligonucleotides pair 5'-GCT CCT GTG CTC CAG ACT T-3' and 5'-CCT TTA ATC CCA GCA CTC AG-3'. The TIFm71 probe, a 71 bp fragment, was amplified by PCR using oligonucleotides pair 5'-TTT CTC TGT GTA GCT TT-3' and 5'-CTC TGT GAG TTT GAG GC-3'. The open reading frame (ORF) of TIF probe was 306 bp, amplified by PCR using the oligonucleotides pair 5'-ATG GCC CCA GTC ACC AGG-3' and 5'-TCA GCT GGA CTT GGC CTT-3'. To detect the generation of TIFm71 or TIF-miR, 20 μg of each total RNA sample was resolved on a 12% urea-polyacrylamide gel and transferred to Nytran SPC membrane (Whatman) using a semidry electroblotting apparatus. Membranes were crosslinked and hybridized overnight at 30°C in ULTRAhyb-Oligo hybridization buffer (Ambion) with ^{32}P-labeled oligonucleotides complementary to potential TIF-miR (5'-TCT GTG AGT TTG AGG CCA GCC TGG TCT CCA-3'). U6 RNA was used as control for total RNA loading, and the sequence was 5'- GCA GGG GCC ATG CTA ATC TTC TCT GTA TCG -3'. Membranes were exposed to Kodak Phosphor Screen SD230 for quantification.

3.3.2 Construction of pcDNA3-M229/S229

M229 and S229 fragments tailed with *Bam*HI and *Xba*I were amplified by PCR and were constructed into the multiple cloning sites of pcNDA3 (Fig. 3.1).

3.3.2.1 Amplification of M229 and S229

The M229 fragment was PCR-amplified from genomic DNA of CHO-K1 cell. PCR reaction was run for 35 cycles with denaturing temperature at 94°C for 30 sec, an annealing temperature at 60°C for 30 sec and an extension temperature at 72°C for 30 sec. At the end of the amplification cycles, an extended incubation at 72°C for 7 min was added. The forward (F) and reverse (R) primers and length of amplicons (in parentheses) obtained were *Bam*HI-F1528: 5'-CGG GAT CCC GTG TCT TAG AAC TGA GAA GTT-3', *Xba*I-R1522: 5'-GCT CTA GAG CTT CCA AAT TTT TTG TAA AAA TAC CA-3' (229 bp). The S229 fragment was constructed based on previously constructed pcDNA3-R71 by successive addition of 3'-flanking sequence using reverse oligos in 3-round PCR amplification (Fig. 3.5B). In the first round of PCR amplification, the fragment that contains scrambled TIFm71 was amplified from pcDNA3-R71 with forward primer *Bam*HI-F1528 (5'-CGG GAT CCC GTG TCT TAG AAC TGA GAA GTT-3') and reverse primer R1523 (5'-ATA AAT ACC GGC GTT GGT GGT GCA CAC CTT TAA T-3'). In the second round amplification, 1 µl of 20x diluted PCR product from first round PCR was added as the template in the second round PCR mixture that was then amplified with *Bam*HI-F1528 and R1526 (5'-AAT TTT TTG TAA AAA TAC CAA TAA ATA CCG GCG TTG G-3'). Finally, S229 DNA was amplified from the 20x diluted second round PCR product by forward primer *Bam*HI-F1528 and reverse primer *Xba*I-R1522. The PCR program in the above-mentioned 3-step elongation was similar to that of the M229 amplification.

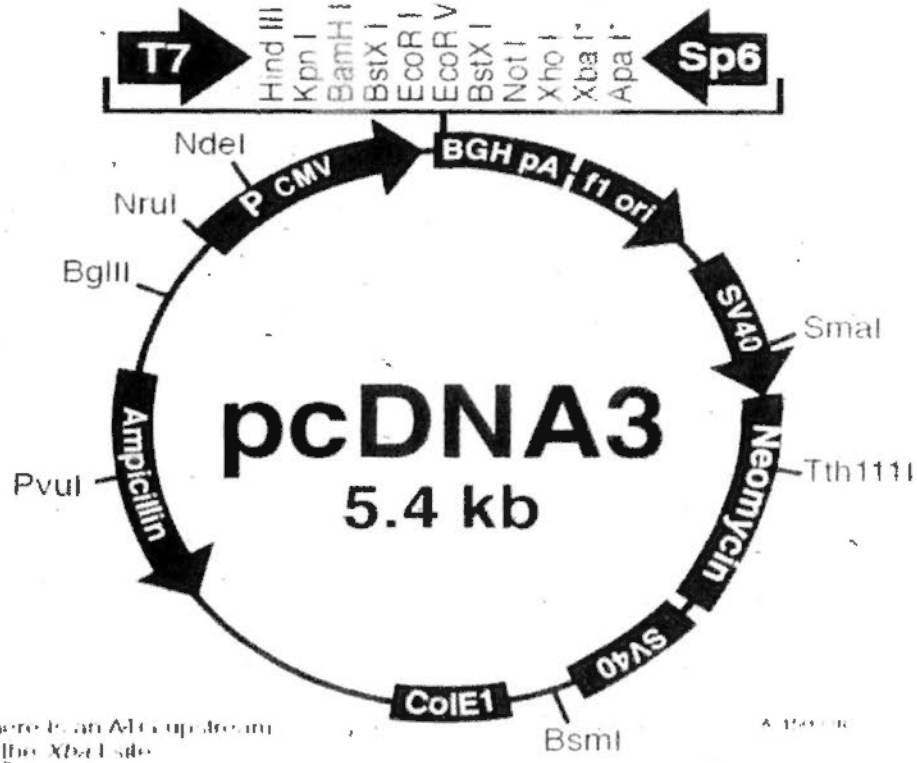


Figure 3.1 Vector map of pcDNA3. M229 and S229 fragments were subcloned into pcDNA3 vector between the *Bam*HI and *Xba*I restriction enzyme cutting sites as indicated by the arrows. pcDNA3 expression vector contains both ampicillin resistant gene and neomycin resistant gene, therefore once transformed into bacteria or mammalian cells, it will enable the transformant to resist ampicillin and hygromycin.

3.3.2.2 DNA fragment purification

PCR products of M229 and S229 were mixed with 10x DNA loading dye (0.21% bromophenol blue, 0.21% xylene cyanol FF, 50% glycerol, 0.2 M EDTA, pH 8.0) and loaded into the wells of 1.5% agarose gel (w/v, agarose in 1x TAE buffer) containing 0.5 μg/ml ethidium bromide. Electrophoresis was carried out at 120 V constant voltage for ~20 min. Afterwards, DNA bands were cut with a clean blade under UV light visualization. Gel Extraction kit (GE Healthcare) was used to purify the DNA fragment from agarose gel. To dissolve the gel, 3 gel volumes of Capture buffer 2 (100 mg of agarose gel is approximately 100 μl) was added to the gel slices and the mixture was heated at 50°C for 10 min. When the gel was completely dissolved, 1 gel volume of isopropanol was added to increase the yield of DNA. The mixture was then loaded to the assembled GFX Microspin column and collection tube, followed by spinning down at 16,000 xg for 30 s. The column was washed with 500 μl of Washing Buffer PE (4 volumes of 100% ethanol was added before use) once. To elute DNA, 50 μl of water was added to the bottom center of the column, incubated at room temperature for 2 min and centrifuged at 16,000 xg for 1 min.

3.3.2.3 Quantification of DNA

DNA quantification was performed as described in section 2.3.2.3.

3.3.2.4 Restriction enzyme digestion

Purified M229 and S229 fragments (~50 ng) were digested in a system containing 3 μl of 10x NEBuffer 4 (20 mM Tris-acetate, 50 mM potassium acetate,

10 mM Magnesium Acetate, 1 mM Dithiothreitol, pH 7.9), 3 µl of 10x BSA (1µg/µl), 2 µl of *Bam*HI (20U/µl) and 2 µl of *Xba*I (20U/µl) in a final volume of 30 µl. The digestion was carried out at 37°C overnight and purified by agarose gel electrophoresis and gel extraction. Vector plasmid pcDNA3 (1 µg) was digested using the same condition.

3.3.2.5 Ligation of M229/S229 fragment with pcDNA3

The *Bam*HI/*Xba*I digested M229/S229 fragments and pcDNA3 vector were ligated at 1:3 (mol/mol) ratio. The ligation system consisted of 4 µl of 5x ligation buffer (250 mM Tris-HCl, 50 mM $MgCl_2$, 5 mM ATP, 50 mM Dithiothreitol, pH 7.5), 1 µl of T4 ligase (1 U/µl) and 1:3 mixed DNA fragments. The ligation was carried out at 16°C for overnight.

3.3.2.6 Preparation of DH5α competent cells

DH5α *Escherichia coli* were thawed from -80°C stock and streaked on a LB agar plate, followed by incubating at 37°C for ~14 h. Single colonies were picked and cultured in 2 ml of LB medium at 37 °C overnight with shaking at 250 rpm. The next day, 1 ml of bacteria was seeded into 100 ml of LB medium and continued to culture until OD_{600} reached 0.4-0.6. Then the DH5α cells were harvested by centrifugation at 3,000 g for 10 min at 4°C, resuspended in 25 ml of 0.1 M ice-cold $CaCl_2$ and incubated on ice for 15 min. After that, the cells were pelleted by centrifugation at 3,000 xg for 10 min and resuspended in 5 ml of 0.1 M ice-cold $CaCl_2$ containing 10%

glycerol. Finally, cells were aliquoted (50 µl/tube) and stored at -80°C to avoid cycles of freeze-thawing.

3.3.2.7 Transformation

DH5α competent cells were slowly thawed on ice, and 10 µl of ligation products were added followed with gentle mix. After incubated on ice for 30 min, the competent cells were heat-shocked in 42°C water bath for 90 s followed with incubation on ice for 2 min. Following addition of 1 ml of LB medium, the bacteria were incubated at 37°C with shaking at 250 rpm for 1 h. Afterwards, cells were collected by centrifugation at 4,000 rpm for 5 min and resuspended in 100 µl of LB medium. The bacteria culture was spread onto a 100 mm LB agar plate containing 100 µg/ml of ampicillin and cultured at 37°C overnight. The majority of the colonies resulting from this transformation contained only recombinant plasmid, confirmed by restricted digestion and subsequent sequencing.

3.3.2.8 Mini-prep and midi-prep of plasmid

QIAprep Spin Miniprep Kit (Hilden, Germany) was used in mini-prep of recombinant plasmids following the manufacture's instruction. A single colony from transformed bacteria containing recombinant plasmid was picked and cultured in 5 ml of LB medium containing 100 µg/ml ampicillin with vigorous shaking at 37°C overnight. The next day, ~2 ml of the culture bacteria were collected by centrifugation at 2,000 xg for 5 min, resuspended in 250 µl of buffer P1 supplemented with 100 mg/ml RNase A, and lysed by adding 250 µl of buffer P2. After kept at room temperature for

2 min, 350 µl of buffer P3 was added followed by gentle inverting the tube up and down for 6 times. After that, the mixture was centrifuged at 14,000 rpm for 10 min at 4°C to pellet undissolvable debris, and the supernatant was transferred into a spin column followed by centrifugation at 13,000 rpm for 1 min. The column was washed once with 0.75 ml of buffer PE and 50 µl of nano pure water was added onto the filter inside the column. After incubation at room temperature for 2 min, the vector DNA was rescued by centrifugation at 13,000 rpm for 1 min.

For the plasmid DNA midi-prep, 100 µl of positive bacterial suspension were innoculated into 100 ml of LB medium containing 100 µg/ml ampicillin and grown with vigorous shaking for 12-16 h. The bacterial cells were harvested by centrifugation at 4,000 xg for 15 min, followed with resuspension in 5 ml of solution I (50 mM glucose, 10 mM EDTA, 25 mM Tris pH 8.0) supplemented with 100 mg/ml RNase A. Bacteria were lysed by adding 10 ml of freshly prepared solution II (0.2 M NaOH, 1% SDS). The contents were mixed thoroughly by gently inverting the tube 4-6 times and incubated for 5 min at room temperature. After that, 6 ml of ice-cold solution III (3M potassium acetate, 11.5% glacial acetic acid) was added, mixed gently by inverting several times and incubated on ice for 10 min until a flocculent white precipitate formed. Afterwards, the mixture was centrifuged for 15 min twice at 20,000 xg to pellet the cellular debris and the genomic DNA. DNA in the cleared supernatants was precipitated by addition of 12 ml of isopropanol, followed by incubation at room temperature for 10 min. Plasmid DNA was recovered by centrifugation at 2,000 xg at room temperature for 15 min, air-dried, and dissolved in 3 ml of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Afterwards, equal volume of autoclaved 5 M LiCl was added, precipitated RNA was removed by centrifugation at

10,000 xg for 10 min at 4°C. Supernatant was transferred into a fresh conical tube, and equal volume of ice cold isopropanol was added. After thorough mixing, the tube was kept still at -20°C for 10 min. Plasmid DNA was collected by spinning at 2,000 xg for 10 min at 4°C. To get rid of residual RNA, the pellet was dissolved in 0.5 ml TE buffer containing 20 ug/ml DNAse-free RNAse and incubated at 37°C for 30 min. The DNA was then extracted with 0.5 ml of phenol for 3 times. At each time of extraction, 0.5 ml of phenol was added, vortexed for 20 s, and centrifuged for at 15,000 xg for 5 min. The upper phase containing plasmid DNA was transferred into a new tube, followed with the second round of extraction. The plasmid DNA was then precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol, kept at room temperature for 10 min and centrifuged at 12,000 xg for 5 min at room temperature or 4°C. The supernatant was carefully decanted and the pellet was washed with 1 ml of 70% ethanol, centrifuged at 75,00 xg for 5 min. The plasmid DNA pellet was air dried and then dissolved in appropriate volume of autoclaved nano water. Finally the plasmid DNA was quantified by DU-640 spectrophotometer (Beckman Coulter, USA).

3.3.2.9 DNA sequencing

The purified recombinant pcDNA3-M229 and pcDNA3-S229 were sequenced by Beijing Genomics Institute, Hong Kong.

3.3.3 Construction of pIREShyg3-M229/S229

Both M229 and S229 sequences were amplified from pcDNA3-M229 and pcDNA3-S229 by the same pair of primer *Bam*HI-F1528 (5'-CGG GAT CCC GTG TCT TAG AAC TGA GAA GTT-3') and *Bst*XI-R1527 (5'-ATC CCA GCA CAC TGG TTC CAA ATT TTT TGT AAA AA-3'). The fragments were constructed into the multiple cloning sites of pIREShyg3 vector between *Bam*HI and *Bst*XI cutting sites using the same methods described in section 3.3.1 (Fig. 3.2).

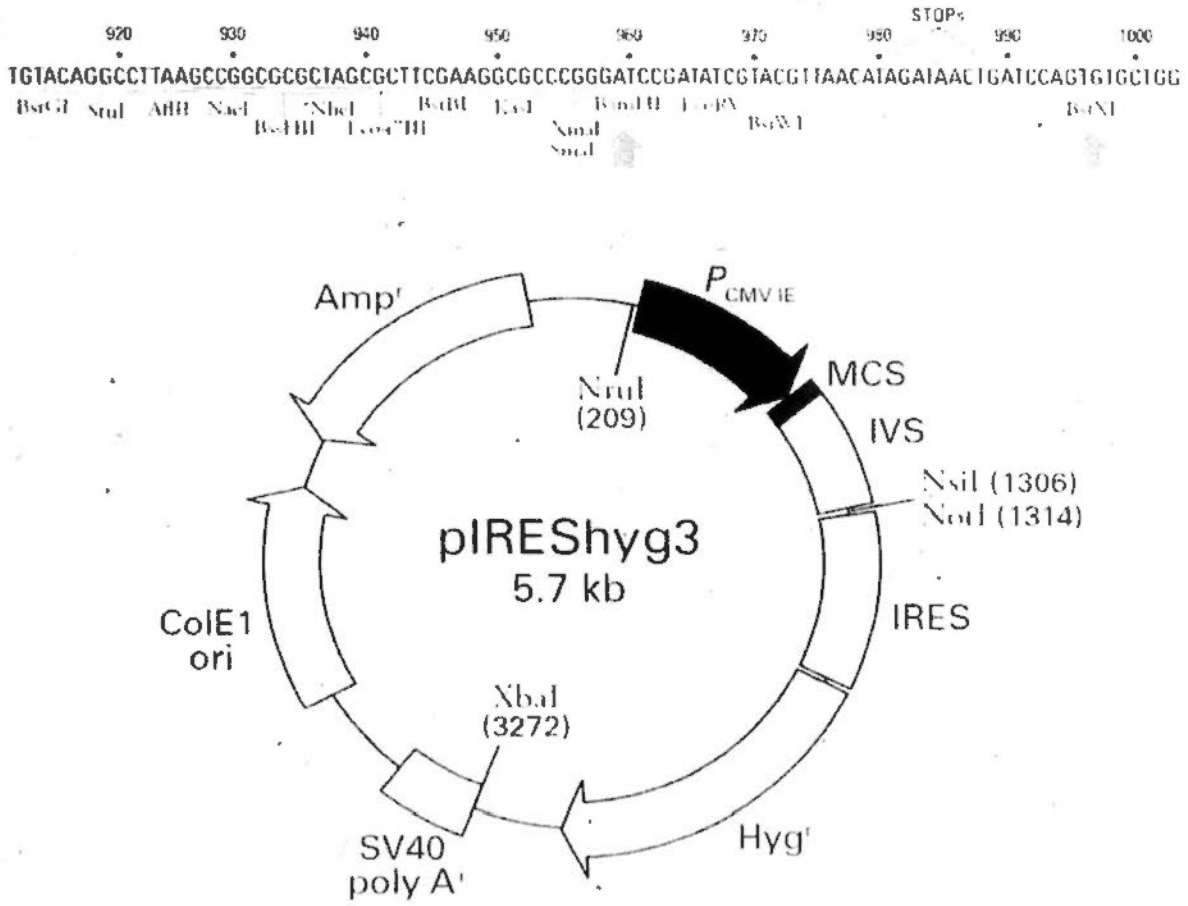


Figure 3.2 Multiple cloning sites and vector map of pIREShyg3. M229 and S229 fragments were subcloned into pcDNA3 vector between the *Bam*HI and *Bst*XI restriction enzyme cutting sites as indicated by arrows. pIREShyg3 expression vector contains both ampicillin resistant gene and hygromycin resistant gene. Therefore, once transformed into bacteria or mammalian cells, it will enable the transformant to resist ampicillin and hygromycin.

3.3.4 Stable expression of TIFm71 in CHO-K1 cells

To establish cell clones stably expressing TIFm71, CHO-K1 cells were transfected with pcDNA3-M229 and pIREShyg3-M229 constructs using lipid transfectant Lipofactamine 2000 (Cat. 11668-019), and then selected against 1 mg/ml G418 or 500 μg/ml hygromycin B. Meanwhile, negative control constructs including pcDNA3-S229, pIREShyg3-S229, pcDNA3/pIREShyg3 empty vectors were applied in the transfection.

3.3.4.1 Linearization and purification of vectors

Prior to transfection, pcDNA3-M229, pcDNA3-S229 and pcDNA3 vector (30 μg) were linearized with *Bsm*I enzyme in a reaction system containing 1x NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 1 mM Dithiothreitol, pH 7.9), 1mg/ml BSA and 100 units of *Bsm*I in a final volume of 200 μl. The reaction mixture was incubated at 65°C overnight. pIREShyg3-M229, pIREShyg3-S229, and pIREShyg3 vector (30 μg) were linearized with *Xba*I in a reaction mixture containing 1x NEBuffer 4, 1mg/ml BSA and 100 units of *Xba*I in a final volume of 200 μl. The reaction mixture was incubated at 37°C overnight. Linearized DNA was precipitated by adding 2.5 volumes of ethanol, 1/10 volume of 3 M sodium acetate and incubated at -20°C for 15 min, followed by centrifugation at 12,000 xg for 5 min to obtain the DNA pellet. The DNA pellet was then washed once with 70% ethanol and dissolved in sterile water.

3.3.4.2 Transfection by Lipofectamine 2000

CHO-K1 (~10^5) were seeded in 35 mm culture plate with 2 ml of Ham's F-12 complete medium containing 10% (v/v) FBS and kept at 37°C in 5% CO_2 humidified incubator until 90-95% confluence. Linearized plasmid DNA (2 µg) and lipofectamine 2000 (4 µl) was diluted in 250 µl serum free Ham's F-12 medium and incubated at room temperature for 5 min. The diluted plasmid DNA and diluted lipofectamine 2000 were combined with gentle mixing and incubated for 20 min at room temperature. The mixture was added to the CHO-K1 cells containing 1 ml of Ham's F-12 complete medium supplemented with 10% (v/v) FBS. Cells were incubated at 37°C in a 5% CO_2 humidified incubator for 24 h.

3.3.4.3 Selection of stably transfected cells

One day after the transfection, cells were selected against 1 mg/ml G418 (transfected with pcDNA3 vectors) or 500 µg/ml hygromycin B (transfected with pIREShyg3 vectors) in Ham's F-12 complete medium. The selection medium was refreshed every 3 days to remove untransfected dead cells killed by the antibiotics. After one week, cells were trypsinized and seeded in a 96-well plate at a concentration of one cell per milliliter of selection medium. The plate was kept at 37°C in a 5% CO_2 humidified incubator for 10 days to allow the formation of cell colonies. Afterwards, cell colonies were trypsinized and subcultured in 35 mm dish, followed by the testing of transgene expression by RT-PCR.

3.3.4.4 Verification of stable cells by RT-PCR

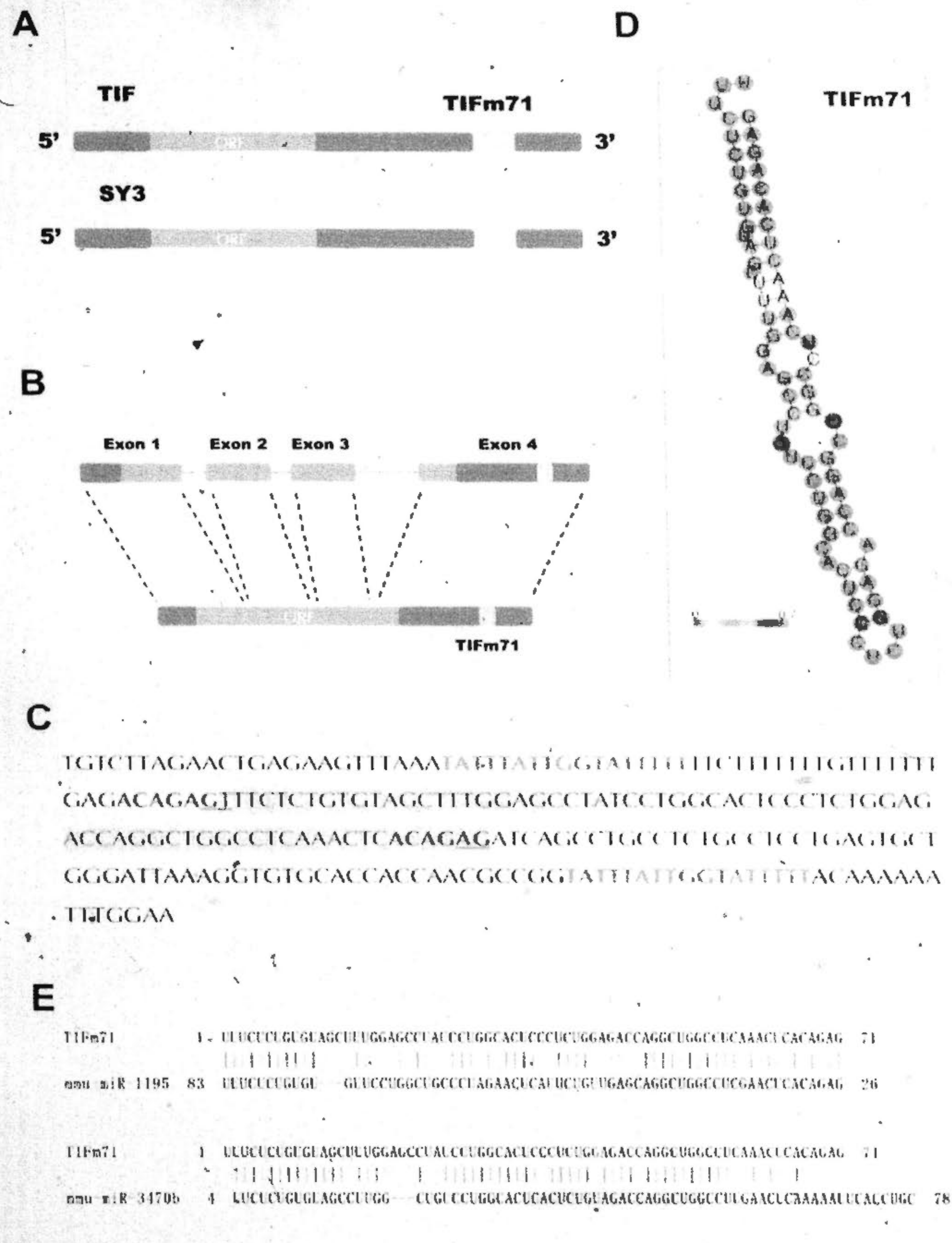
A similar procedure was described in section 2.3.2.4 and 2.3.2.5. Briefly, total RNA of stable cells (2 μg), dNTP (10 mM, 1 μl), oligo dT (0.5 μg/μl, 1 μl) and RNase-free H_2O were added into a PCR tube to make a final volumn of 13 μl. The mixture was incubated at 65 °C for 5 min, and was immediately cooled on ice to denature the RNA. Then 4 μl of first strand buffer (5x) and 2 μl of 0.1M DTT were added into the denature RNA sample and mixed thoroughly. The mixture was pre-warmed at 42 °C for 2 min and SuperScript™ II Reverse Transcriptase (10 units/μl, 1 μl) was added to the mixture. The reaction system was firstly incubated at 42 °C for 50 min, and then at 70 °C for 15 min. The product was stored at -20°C until use.

PCR procedure was as described in section 2.3.2.5. Briefly, the reaction was run for 30 cycles at denaturing temperature 94°C for 30 sec, an annealing temperature 62°C for 30 sec and an extension temperature 72°C for 30 sec. At the end of the amplification cycles, an extended incubation at 72°C for 7 min was added. The forward (F) and reverse (R) primers and length of amplicons (in parentheses) obtained were: M229/S229 F: 5'-TGT CTT AGA ACT GAG AAG TT-3', R: 5'-TTC CAA ATT TTT TGT AAA AA-3' (229 bp); GAPDH F: 5'-CAA GGC TGA GAA TGG AAA GC, R: GGG GTT ATT GGA CAG GGA CT-3' (916 bp).

3.4 Results

3.4.1 Preliminary evidence indicating TIFm71 is an ncRNA

In our previous study, a couple of genes including the novel chemokine TIF, were found up-regulated in *mas*-oncogenic Mc0M80 cells compared to vector control Vc0M80 cells by FluorDD (Lin *et al.*, 2009). SY3, an isoform of TIF, was also identified to be up-regulated in oncogene *mas* over-expressing Mc0M80 cells. Sequence alignment revealed that SY3 mRNA shared identical sequence with TIF mRNA except lacking a 71 nt string in the 3'- untranslated region (3' UTR), and this 71 nt sequence was referred as TIFm71 (Fig. 3.3A). TIFm71 located in the fourth exon, and GT/AG was observed at both ends of TIFm71, indicating that TIFm71 might be a potential intron (Fig. 3.3B-C). In addition, two pairs of short repetitive RNA motifs, R-SBE-like motifs and 17 nt AT-rich elements were found at both 5' and 3' end of TIFm71 (Fig. 3.3C). R-SBE$_s$ sequence is critical for the TGFβ/BMP-dependent pri-miRNA to pre-miRNA processing (Davis *et al.*, 2010); while the function of AT-rich sequence was yet to be clarified. It was of interest to note that TIFm71 could fold into a canonical pre-miRNA-like stem-loop secondary structure (Fig. 3.3D) predicted with RNA folding program RNAfold (Gruber *et al.*, 2008). Pre-miRNAs such as pre-miR-1195 and pre-miR-3470b were found to share high sequence similarities with TIFm71 (Fig. 3.3D-E). Moreover, the reverse complement of TIFm71 shared 56.3% similarities with a human short interspersed nuclear element (SINE) Alu-J in the consensus region (Fig. 3.3F). These observations indicated that TIFm71 might be released from TIF transcript and act as an ncRNA or even generate miRNA.


A
TIF
TIFm71
5'
3'
SY3
5'
3'
B
Exon 1
Exon 2
Exon 3
Exon 4
TIFm71
C
GAGACAGAGTTTCTCTGTGTAGCTTTGGAGCCTATCCTGGCACTCCCTCTGGAG
ACCAGGCTGGCCTCAAACTCACAGAGATCAGCCTGCCTCTGCCTCCTGAGTGCT
D
TIFm71
E
TIFm71
mmu-miR-1195
TIFm71
mmu-miR-3470b

F

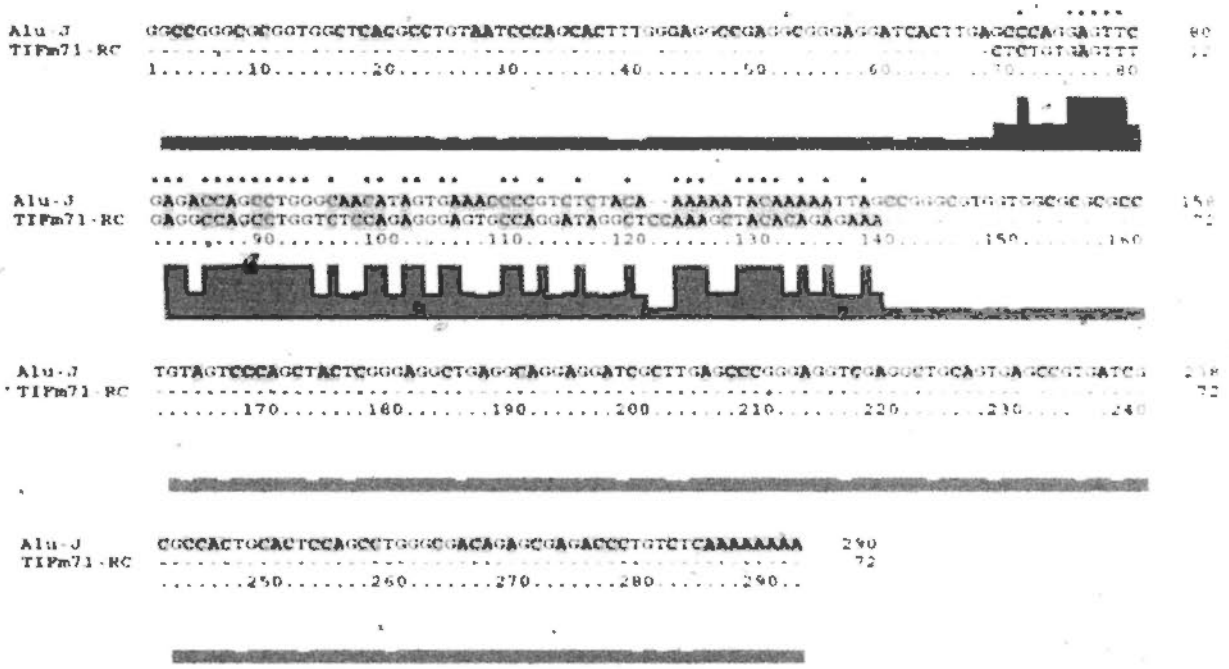


Figure 3.3 Preliminary evidences implied that TIF transcript was more than a protein-coding gene, but also encoded an ncRNA or a pre-miRNA. (A) Schematic diagram showing the genomic organization of TIF gene. (B) Structure of TIF and SY3 mRNA. (C) TIFm71 with flanking sequence. TIFm71 region was labeled with yellow, R-SBE-like sequences were labeled with blue, 17 bp AT-rich sequences were labeled with red and GT/AG were underlined. (D) Stem-loop secondary structure of TIFm71 predicted by a RNA folding program RNAfold. Different colors of nucleotides indicated values of positional entropy. (E) Sequences alignment of TIFm71 with pre-miRNAs. Highlighted parts represented mature miRNA regions. (F) Sequences alignment between reverse complement of TIFm71 and human Alu-J by ClustalX 2.0 software. "RC" was short for "reverse complement".

3.4.2 Expression of TIFm71 in tissues and xenograft tumors

In our previous study, the expression level of TIF mRNA detected in tumor tissues was higher than that in normal tissues examined by Northern blot analysis. Meanwhile, additional mRNA bands at small RNA region (< 300 nt) were detected in tumor tissues (Lin *et al.*, 2009). Consistent with TIF mRNA expression, the expression level of this small RNA was much higher in solid tumors than that of the normal tissues, and it could only be recognized by the probe containing 3'-UTR but not the ORF region of TIF by Northen blot analysis (Fig. 3.4A), implying that this small size RNA might be TIFm71. To test this possibility, Mc0M80 cells were inoculated subcutaneously into nude mice to induce xenograft tumor formation as described previously (Lin *et al.*, 2009). Total RNA of both normal tissue and xenograft tumor were extracted followed by polyacrylamide gel Northern blot analysis using radioactive labeled TIFm71 probe. The result confirmed the expression of TIFm71 in normal tissues, and a great up-regulation in xenograft tumor. However, smaller RNAs that less than 71 nt was detected in solid tumors (Fig. 3.4B). The mechanism of the formation of TIFm71 related small RNAs was unknown.

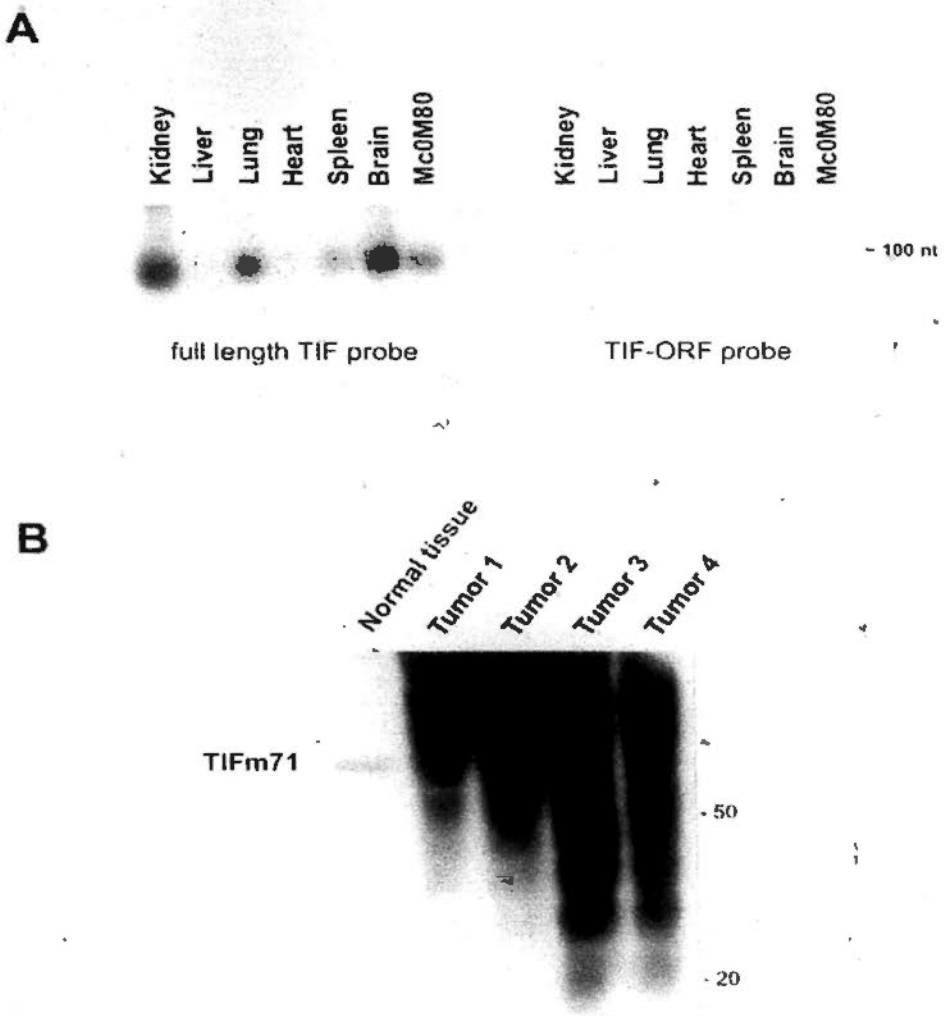


Figure 3.4 Expression of TIFm71 in normal tissues and tumors. (A) Expression of TIFm71 in hamster normal tissues were examined by Northern blot analysis with DIG-labeled full-length TIF or open reading frame region of TIF DNA probe as described in the Methods section. (B) TIFm71 in xenograft tumors examined by polyacrylamide gel Northern blot analysis. Mc0M80 cells were subcutaneously inoculated into nude mice to induce xenograft tumors. Total RNA of these tumors was extracted and analyzed with Northern blot. RNA isolated from healthy hamster brain tissue was loaded as control.

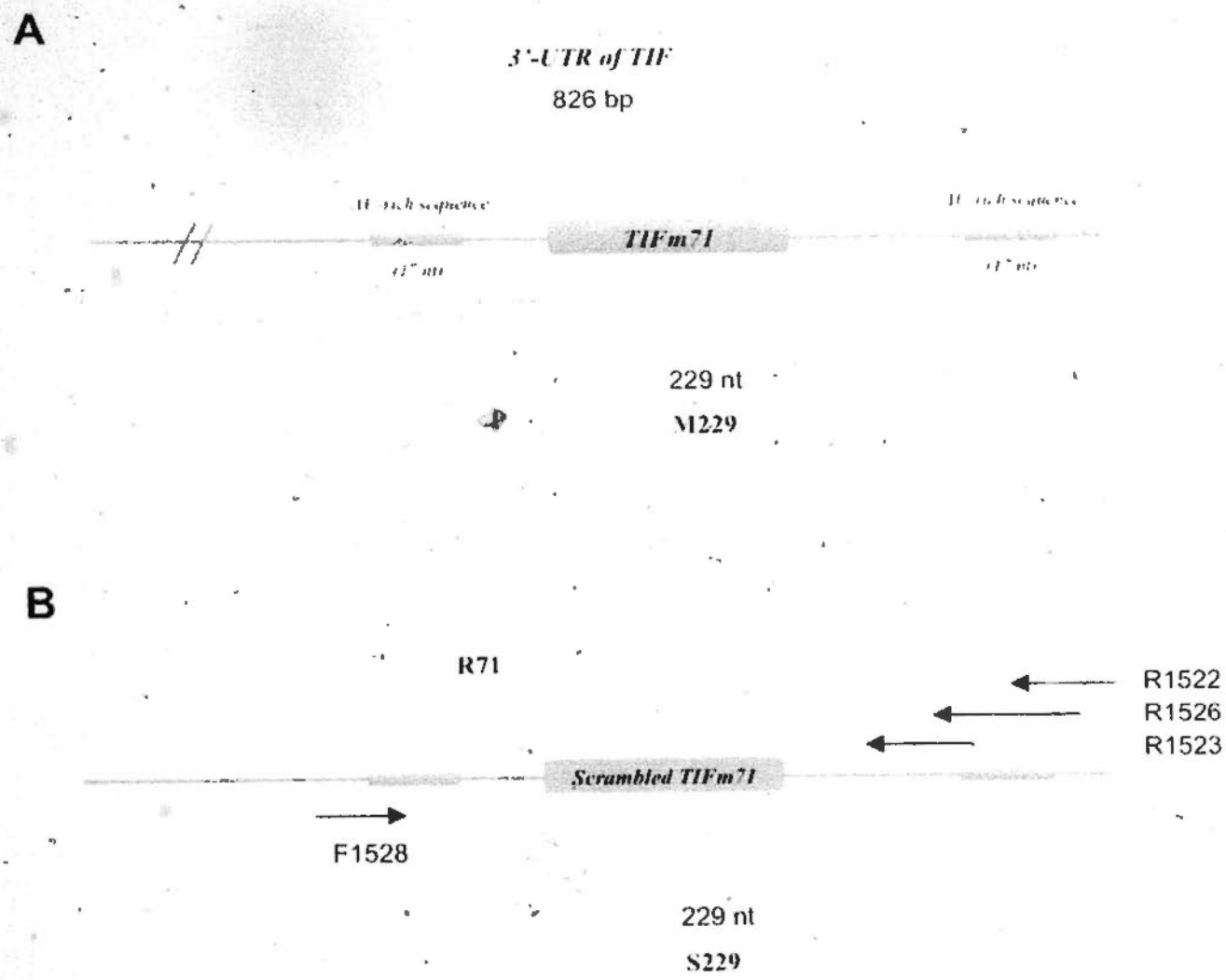
3.4.3 Cloning of TIFm71 and control cDNA

In order to investigate whether TIFm71 was a pre-miRNA or an ncRNA released from TIF mRNA, several fragments were subcloned downstream of CMV promoter of pcDNA3 and pIREShyg3 vectors as shown in Fig. 3.1 and Fig.3.2. In pri-miRNA to pre-miRNA processing, flanking sequence of stem-loop structure may facilitate the recognition and contain the binding sites for Drosha or DGCR8 (Ohler *et al.*, 2004; Zeng and Cullen, 2005). Therefore, a 229 bp fragment M229 containing TIFm71 together with 5' and 3' flanking regions was PCR-amplified from genomic DNA of CHO-K1 cell and subcloned into pcDNA3 and pIREShyg3 vectors (Fig. 3.5A and C). The flanking sequence in M229 contained two 17 bp AT-rich sequence (5'-TAT TTA TTG GTA TTT TT-3') at both 5' and 3' of TIFm71, which maybe functional in TIFm71 processing (Fig. 3.5A).

Scrambled-M229 (S229), which contained the same flanking sequence with M229 but a scrambled TIFm71 sequence to avoid the formation of stem-loop secondary structure, was synthesized as a negative control (Fig. 3.5B). S229 fragment was amplified from previously constructed pcDNA3-R71 (Zhoufang, unpublished data) by 3-step elongation as described in section 3.3.2.1. The PCR product of the third round elongation was the S229 fragment, which was analyzed by agarose gel electrophrosis as shown in Fig. 3.5D.

Besides, TIFm71, SY3 (TIF that lacking TIFm71 in 3'-UTR) and full-length TIF sequences were subcloned into both pcDNA3 and pIREShyg3 vectors using the same protocol as described above (Fig. 3.6).

THIS IS A BLANK PAGE


A
3'-UTR of TIF
826 bp
TIFm71
229 nt
M229
B
R71
R1522
R1526
R1523
Scrambled TIFm71
F1528
229 nt
S229
C
(bp)
1 kb+ DNA ladder
M229
16,50 -
500 -
300 -
200 -
100 -
D
(bp)
1 kb+ DNA ladder
S229
16,50 -
500 -
300 -
200 -
100 -

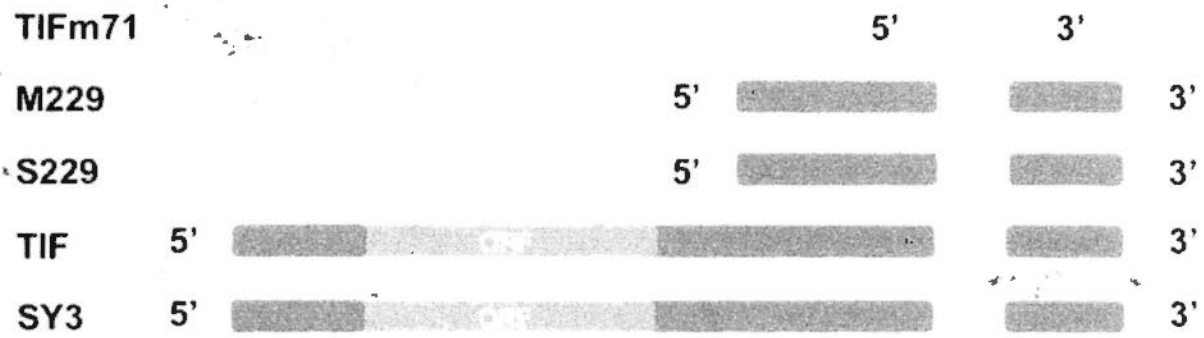


Figure 3.6 Expression constructs. Five fragments, including TIFm71, TIFm71 with flanking sequence (M229), scrambled TIFm71 with flanking sequence (S229), full-length TIF and SY3 were PCR-amplified and subcloned into pcDNA3 and pIREShyg3 vector. The ORF sequence was marked by gray color, while the 5' and 3' UTR were indicated by dark red. TIFm71 was indicated by yellow and the scrambled TIFm71 was marked with mosaic.

3.4.4 Development of stable cell lines

To establish cell lines stably over-expressing M229 and S229 RNA, expression vectors of these two fragments, as well as the empty vectors, were linearized and transfected into CHO-K1 cells with Lipofectamine 2000 reagent following the manufacture's instructions. Antibiotics resistant cell colonies were selected as described in section 3.3.4.3. Total RNA of these cells was extracted and reverse transcribed to cDNA, and the expression of M229 or S229 were detected by PCR with specific primers targeting the flanking sequence. Most of hygromycin B selected cells (transfected with pIREShyg3) were positive cells which stably expressing M229 or S229 (Fig. 3.7A). Clone 2 was the only negative cell clone we got (Fig. 3.7A, lane 3), and the weak band was probably amplified from endogenously expressed TIF transcript. Representative RT-PCR verification result on G418 selected cells (transfected with pcDNA3) was shown in Fig. 3.7B.

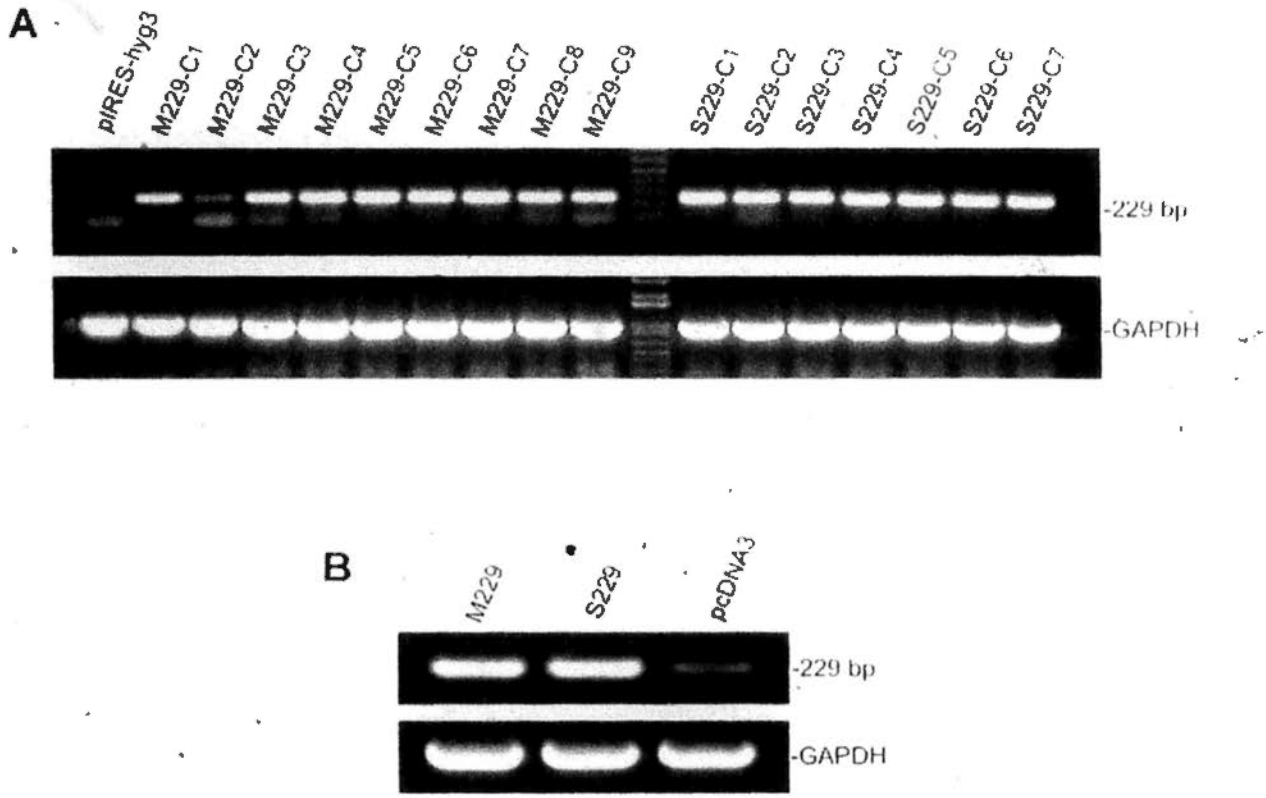


Figure 3.7 Verification of cell clones stably expressing of M229 and S229 by RT-PCR. (A) CHO-K1 cells were transfected with pIREShyg3 empty vector, pIREShyg3-M229 or pIREShyg3-S229. Cell colonies that survived 500 μg/ml hygromycin B selection were passed to 35 mm dish and cultured until confluence. Total RNA was extracted and reverse transcribed to cDNA. The expression M229 or S229 were detected by PCR with specific primers targeting the flanking sequence. (B) CHO-K1 cells were transfected with pcDNA3 empty vector, pcDNA3-M229 or pcDNA3-S229. Cell colonies were selected with 1 mg/ml G418 and verified with RT-PCR using the same protocol as described in (A). GAPDH was amplified as a control for equal loading.

3.4.5 Morphology of stable cell lines

Cell morphology of selected stable cells was analyzed by microscopy. The wild type CHO-K1cells and the S229 stably expressing cells displayed a polygonal, cobblestone-like appearance of epithelial cells. These cells tend to grow in cluster. It is of interest to note a striking change in cellular morphology of cells over-expressing TIFm71. Stable clones over-expressing M229 showed a fibroblast-, spindle-like morphology with extensive cellular scattering (Fig. 3.8). These were hallmarks of epithelial-to-mesenchymal transitions (EMT), which is a pivotal cellular program in regulating cellular plasticity in normal adult tissues and tumors, where they can generate multiple, distinct cellular subpopulations contributing to intratumoral heterogeneity (Kang and Massague, 2004; Polyak and Weinberg, 2009). Further analysis would be needed to test whether the over-expression of TIFm71 could promote EMT.

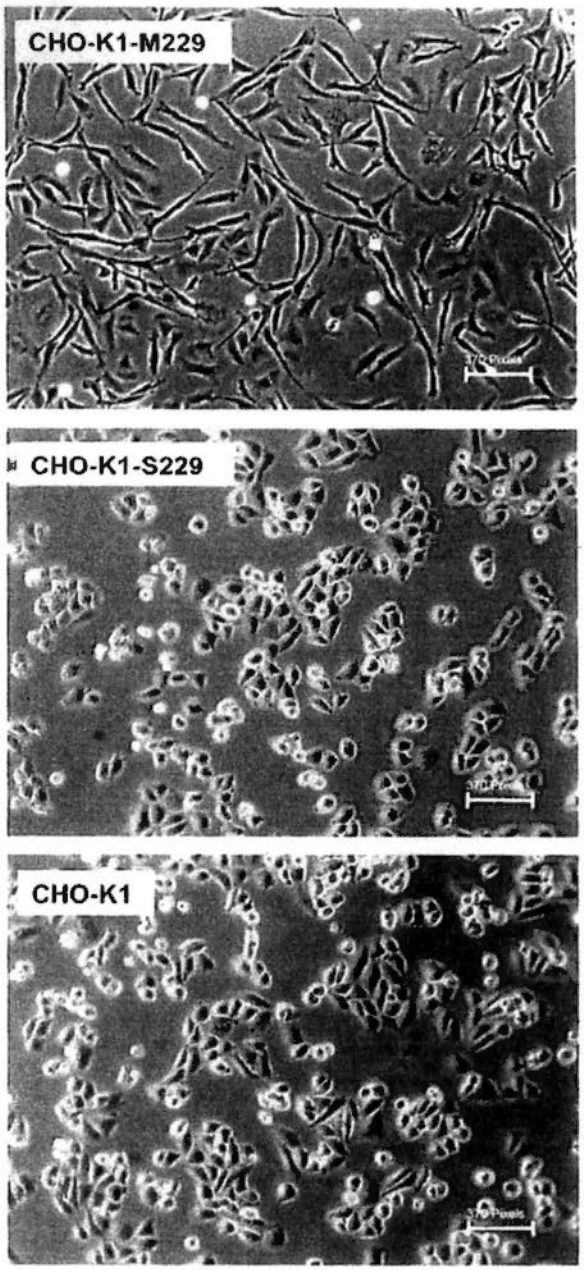


Figure 3.8 Morphology of CHO-K1 cells stably expressing M229/S229 and wild type CHO-K1 cells. Equal number of each type of cells (10^5) were seeded in 35 mm dish and cultured for 24 h. Images were captured using a 20x objective with ZEISS Axiovert 135 microscope, which is coupled to a Leica camera (DFC490). Image capture and processing were controlled via Leica Application Suite (LAS). Scale bar was 50 μM.

3.5 Discussion

Sequence analysis and Northern blot suggested that a new exonic ncRNA TIFm71 was potentially derived from the 3'-UTR of TIF transcript. Although expression level was low, endogenous TIFm71 was detectable in normal tissues by Northern blot. However, under solid tumor microenvironment, which usually contains poorly vascularized regions characterized by severe hypoxia (oxygen deprivation), low pH, and nutrient starvation (Keith and Simon, 2007), the expression of TIFm71 was greatly up-regulated. It was probably due to some inducible factor that can bind to the flanking sequence of TIFm71 and facilitate TIFm71 processing. However, the mechanism remains to be clarified.

A canonical miRNA biogenesis pathway is a two-step process. In the first step, the primary miRNA (pri-miRNA) is transcribed by RNA Polymerase II or III (Borchert *et al.*, 2006; Lee *et al.*, 2004) and processed into 50 to 80 bases hairpin-like precursor miRNA (pre-miRNA) by microprocessor complex of nucleases and associated factors, including the RNase III Drosha and its partner DGCR8/Pasha (Denli *et al.*, 2004; Lee *et al.*, 2003). This step happens within the nucleus. In the second step, pre-miRNA molecules, in association with the Ran-GTP dependent factor exportin- 5, are actively transported into the cytoplasm, where a second ribonuclease III enzyme Dicer processes them further into duplexes that contain the 19 to 24 nucleotides mature miRNA and an oligonucleotides of similar size from the other arm of the hairpin (Hutvagner *et al.*, 2001; Lund *et al.*, 2004; Yi *et al.*, 2003). In an *in vitro* Drosha processing assay with RNA substrates of various sizes and structures, it showed that Drosha function also requires the attachment of a substantial single-

stranded RNA flanking sequence to an extensive, 60-80 nt pri-miRNA stem-loop structure (Zeng and Cullen, 2005). Therefore, in order to study the processing of TIFm71, a fragment termed M229, which contained TIFm71 with single-stranded RNA flanking sequences, was subcloned into expression vectors.

In the flanking sequence of TIFm71, we found two pairs of short repetitive RNA motifs, R-SBE-like motifs and 17 nt AU-rich elements, whose functions were yet to be clarified. Smad binding element (SBE) is a consensus sequence (5'-CAGAC-3') found in the promoter region of transforming growth factor β (TGFβ) target genes, such as germline Igα constant region, plasminogen activator inhibitor type 1 (PAI-1) and TGFβ1 (Dennler *et al.*, 1998; Massague *et al.*, 2005). A recent study reported a set of miRNAs (T/B-miRNAs), which was regulated posttranscriptionally by TGFβ and BMP signaling, contains a conserved sequence similar to SBE (R-SBE). Smad protein could directly associate with R-SBE and facilitate the maturation of T/B-miRNAs (Davis *et al.*, 2010). In our present study, the R-SBE-like (5'- ACAGAG -3') sequence was found within the 3' end and immediately in front of the 5' end of TIFm71. We believed that the R-SBE-like motif would play a role in the TIFm71 processing, probably by binding to some yet to be identified factors.

AU-rich elements (ARE) are important *cis*-acting short sequences in the 3'-UTR that mediate the recognition of an array of RNA-binding proteins, affecting mRNA stability and translation (Khabar, 2010). AREs from different mRNAs vary dramatically but can be divided broadly into three distinct classes based on their sequence features. Class I AREs contain multiple independent repeats of the AUUUA

pentamer coupled with a nearby U-rich region or U stretch; class II has at least two overlapping copies or closely juxtaposed of the nonamer UUAUUUA(U/A)(U/A) motifs in a U-rich region; while class III has a AU or U-rich region but lacks a core AUUUA element (Chen and Shyu, 1995; Novotny *et al.*, 2005). As shown in Fig. 3.3 C and Fig. 3.5A, a pair of type II AREs (5'- UAUUUAUUGGUAUUUUU -3') were found in -27 nt and +58 nt of TIFm71. It probably acts as a *cis*-element in the regulation of TIF mRNA stability or translational efficiency (Khabar, 2010; Piecyk *et al.*, 2000). However, it may also have other biological functions such as facilitating the process of TIFm71. Therefore, this pair of ARE was included in M229 sequence during the construction of TIFm71 expression vectors.

We observed a striking change in cellular morphology by the over-expression of TIFm71, whereby the cobblestone-like appearance of epithelial cells switched to a fibroblast-, spindle-like morphology with extensive cellular scattering, which are hallmarks of EMT. Therefore, we hypothesized that the over-expression of TIFm71 could promote EMT.

In summary, a novel ncRNA TIFm71 was found expressed in normal tissues at a low level and greatly elevated in solid tumors. Preliminary evidences implied that TIFm71 was possibly derived from an exonic region, the 3'-UTR of TIF. In the next chapter, we would like to prove whether TIFm71 was released from the 3'-UTR of TIF and to test if TIFm71 could generate miRNA.

Chapter 4

Processing of TIFm71

4.1 Introduction

In our present study, a novel ncRNA TIFm71 was found expressed in normal tissues at a low level and greatly elevated in solid tumors. With the constructed expression vectors as described in Chapter 3, we proceeded to test our hypothesis i) whether TIFm71 is released from exonic TIF transcript in living cells after transient transfection, and ii) could TIFm71 be further processed into miRNA.

As discussed in Chapter 1, miRNAs could be defined as ncRNAs that fulfill the following combination of expression and biogenesis criteria (Ambros *et al.*, 2003; Berezikov *et al.*, 2006; Choi *et al.*, 2008): firstly, mature miRNA should originate from a precursor with a characteristic hairpin secondary structure that does not contain large internal loops or bulges (Gruber *et al.*, 2008); secondly, mature miRNA should be processed by ribonuclease III enzyme Dicer, which is one of the most critical proteins in the processing of mature and functional miRNAs (Bartel, 2004); and thirdly, mature miRNA should be expressed as a distinct transcript of ~22 nt that is detectable by RNA (Northern) blot analysis or other experimental means. Our present study revealed that TIFm71 was predicted to fold into a canonical pre-miRNA-like stem-loop secondary structure predicted with a RNA folding program. In the following study, we would like to find out whether TIFm71 could serve as substrate of Dicer and if any miRNA would be released from TIFm71 in transiently transfected HEK293 cells detected by polyacrylamide gel Northern blot analysis.

Generally, ncRNAs elicit their biological responses through interacting with RNA-binding proteins (RBPs) within ribonucleoprotein particles (RNPs) (Charon *et al.*, 2010). Therefore, we are of interest to identify the binding proteins of TIFm71, which will facilitate our understanding of the biological functions of TIFm71 as well as the mechanisms on TIFm71 processing.

The application of mass spectrometry (M/S) to complex proteomic samples was benefited from high-throughput fashion and unbiased data processing. It greatly facilitated the analysis of proteins and peptides, including the identification of RNA binding proteins (Helsens *et al.*, 2011). We performed RNA pull-down assay followed with liquid chromatography/mass spectrometry/mass spectrometry (LC/MS/MS) proteomics analysis to identify TIFm71 RNA binding proteins. After separation by liquid chromatography, the peptides are subjected to M/S analysis using a mass spectrometer, which is consisted of an ion source, a mass analyzer and a detector. Electrospray Ionization (ESI) and Matrix-Assisted Laser Desorption/Ionization (MALDI) are the two most commonly employed ion sources for proteins and peptides (Pitt, 2009). LC-MS is able to determine the mass, charge and retention time of the peptides contained in the sample under examination, while an MS/MS spectrum provides actual amino acid composition for selected signals in the LC-MS map. The peptides selected for MS/MS recording are called precursor ions. The MS/MS spectrum is a mass spectrum containing *m/z* and intensity measurements of fragments of the precursor ion (Herrin *et al.*, 2005).

4.2 Materials

4.2.1 Chemicals

GeneRuler™ Ultra Low Range DNA Ladder (#SM 1211) and PageRuler™ Plus Prestained Protein Ladder (#SM 1811) were purchased from Fermentas (Maryland, USA). Precision Plus Protein™ Standards (#161-0374) was purchased from Bio-Rad (Hercules, CA, USA). Lipofactamine 2000 (#11668-019), TRIzol reagent, SYBR® Gold nucleic acid gel stain (#S11494) were purchased from Invitrogen (Carlsbad, CA, USA). CHROMA SPIN TE-30 column (#PR16808) was a product of Clontech (CA, USA). PERFORMA® DTR Gel Filtration Cartridges (#98780) was purchased from EdgeBio (Maryland, USA). ULTRAhyb®-Oligo Hybridization Buffer (#AM8663) and Biotin-11-UTP (#AM8450) were purchased from Ambion (CA, USA). Protease Inhibitor Cocktail Tablets (#04 693 116 001) was purchased from Roche (Roche Diagnostics Corporation, IN, USA). Streptavidin sepharose high performance beads and MicroSpin G-25 columns (#27-5325-01) were purchased from GE Healthcare (UK). Nytran® SPC membrane (#10416296) was ordered from Whatman (Dassel, Germany). Proteomics grade trypsin (#T 6567) was purchased from Sigma-Aldrich (MO, USA). All other chemicals and reagents of molecular biology grade were purchased from Sigma-Aldrich (St. Louis, MO, USA).

4.2.2 Enzymes

Dicer enzyme was purchased form Invitrogen (Carlsbad, CA, USA). T7 RNA polymerase (#9PIP207) RNasin® Ribonuclease Inhibitors (#N2511) and Pfu DNA

polymerase were purchased from Promega (Madison, WI, USA). T4-polynucleotide kinase (T4-PNK, #M0201S) was obtained from New England Biolabs (Ipswich, UK).

4.2.3 Kits and Instruments

Mighty Small II gel set was purchased from GE Healthcare (UK). MAXIscript Kit was obtained from Ambion (Austin, TX, USA). Galaxy® 48R Incubator was manufactured by New Brunswick Scientific (New Jersey, USA). BCA™ Protein Assay Kit (#23225) was purchased from Thermo SCIENTIFIC Inc (Rockford, IL 61101, USA). KCjunior software and μQuant microplate reader were the products of Bio-Tek Instruments Inc (Winnoski, VT, USA). MicrOTOF-Q II ESI-Qq-TOF mass spectrometer was manufactured by Bruker Daltonics (Bremen, Germany).

4.3 Methods

4.3.1 Polyacrylamide gel Northern blot analysis

4.3.1.1 Cell transfection and preparation of RNA samples

HEK293 cells (2 x 10^6) were seeded in 100 mm culture plate and grown for 24 h at 37°C, 5% CO_2. Plasmids were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. In brief, 4 µg of each plasmid and 10 µl of lipofectamine 2000 were incubated separately in 500 µl of serum free DMEM for 5 min, then mixed at room temperature and incubated for another 15 min. The mixture together with 4 ml of serum free DMEM were added to the HEK293 cells. After 5 h of incubation, medium was changed to DMEM supplemented with 10% FBS. Total RNA of transfected cells was isolated with TRIzol reagent (Invitrogen) 24 h post transfection followed the protocol described in section 2.3.2.2.

4.3.1.2 Preparation of polyacrylamide gel and electrophoresis

To prepare 12% denaturing polyacrylamide gel, 2.25 ml of 40% acrylamide stock (the ratio of acrylamide: bis-acrylamide was 19:1), 3.6 g of urea and 0.75 ml of 10x TBE buffer were mixed and the final volume was made up to 7.5 ml by adding autoclaved water. The urea was dissolved by heating gently in a microwave (~15 seconds) or warming up in a 37°C water bath for several minutes. After that, 37.5 µl of 10% APS and 7.5 µl of TEMTD were added, and the gel solution was loaded into the Mighty Small II (8 x 9 cm) gel set (GE Healthcare). The gel was kept still for at least 1 h before RNA sample loading. Then the polyacrylamide gel was set at a constant voltage of 150 V and pre-ran for 30-60 min with 1x TBE as running buffer. During

pre-running, RNA samples (20 µg) were mixed with 2x Gel Loading Buffer II (95% Formamide, 18 mM EDTA, 0.025% each of SDS, Xylene Cyanol and Bromophenol Blue) and were heated denature at 95°C for 5 min followed by snap cooling on ice. Residual urea in the wells was flushed out with running buffer using a small syringe immediately before sample loading. RNA samples, as well as GeneRuler™ Ultra Low Range DNA Ladder (#SM 1211, Fermentas) were loaded into the wells of 12% denaturing polyacrylamide gel and subjected to electrophoresis at a constant voltage of 150 V for ~2 h.

4.3.1.3 Gel staining with SYBR gold

To confirm the quality of the RNA samples, the gel was stained with 10,000-fold diluted (with 1x TBE buffer) SYBR® Gold nucleic acid gel stain (#S11494, Invitrogen) for 5 min, and photos were taken under UV using MultiImage™ Light Cabinet.

4.3.1.4 Semi-dry transfer

The gel was washed with 0.5x TBE buffer to remove the dye, and the RNA was transferred onto the Nytran® SPC membrane (#10416296, Whatman) using a PANTHER semi-dry electroblotter. Briefly, 3MM paper and Nytran® SPC membrane were pre-wet with 0.5x TBE buffer. On the negative pole of the electroblotter, the following items were set up in order: three pieces of Whatman 3 MM paper, 12% denaturing polyacrylamide gel with RNA (upside down), pre-equilibrated Nytran® SPC membrane, and another three pieces of Whatman 3 MM paper. All air-bubbles

were carefully smoothed out with a glass pipet. The transfer was set up at a constant voltage of 20 V for 1 h.

4.3.1.5 Preparation of probes

Complementary oligonucleotide probe was prepared by 5'-labeling of [γ-^{32}P] ATP with T4-polynucleotide kinase (T4-PNK, #M0201S, NEB). To set up the isotope-labeling mixture, 1 µl of DNA oligonucleotide (10 µM), 1.5 µl of T4-PNK, 3 µl of 10x T4-PNK buffer and 2 µl of [γ-^{32}P] ATP (10 µCi/µl) were added. The total volume was made up to 30 µl by the addition of distilled water. The mixture was incubated in 37°C for 1 h and purified by passing through CHROMA SPIN +TE-30 Column (#PR16808, Clontech). The probe was stored at -20°C before use. To detect TIFm71 and TIF-miR, oligonucleotide complementary to potential TIF-miR (5'-TCT GTG AGT TTG AGG CCA GCC TGG TCT CCA-3') was labeled. The potential mature TIF-miR located in the 3'-stem of TIFm71 stem-loop structure (Fig. 3.3D, left stem), which was predicted by MatureBayes, a probabilistic algorithm for identifying the mature miRNA within novel precursors (Gkirtzou *et al.*, 2010). U6 RNA was examined to control the total RNA loading, and the sequence of the oligonucleotide was 5'- GCA GGG GCC ATG CTA ATC TTC TCT GTA TCG -3'.

4.3.1.6 Hybridization and detection

After semi-dry transfer, the Nytran® SPC membrane carrying RNA was washed once with 1x TBE buffer and air-dried. RNA was fixed by crosslinking for 3 min under UV irradiation (1,200J/100cm^2). Pre-hybridization was carried out by addition of 5 ml of

ULTRAhyb®-Oligo Hybridization Buffer (#AM8663, Ambion) with gentle rotation at 42°C. After 2 h of pre-hybridization, 30 µl of purified [γ-^{32}P] ATP-labeled probe was heated for 10 min at 95°C and chilled on ice immediately for 5 min. The probe was then added directly into the hybridization buffer in the hybridization tube, and 3 ml more ULTRAhyb®-Oligo Hybridization Buffer was added to make the total hybridization volume to 8 ml. The blot was then incubated in a hybridizer at 30°C overnight with gentle rotation. After hybridization, the hybridization buffer was discarded into a ^{32}P liquid waste container. The blot was first washed with 2x SSC/0.1% SDS for 5 min at room temperature twice, then with 0.5x SSC/0.1% SDS for 5 min at 42°C twice. After washing, the blot was wrapped with saran wrap and exposed to FUJI medical X-ray film at -70°C.

4.3.2 Dicer *in vitro* processing assay

RNA transcripts were transcribed directly from the DNA templates containing a T7 RNA polymerase promoter upstream of the sequence following MAXIscript Kit (Ambion, USA) manual instruction. Briefly, 2 µg of DNA template, 2 µl of 10x Transcription Buffer, 1 µl of 10 mM ATP, 1 µl of 10 mM CTP, 1 µl of 10 mM GTP, 1 µl of 10 mM UTP and 2 µl of T7 enzyme mix were added into a microtube. The final volume was made up to 20 µl by the addition of RNase free water. The reaction mixture was thoroughly mixed and incubated at 37°C for 1 h. After that, 1 µl of DNase I was added and incubated at 37°C for 15 min to digest the DNA template. The reaction was terminated by adding 1 µl of 0.5 M EDTA.

Prior to Dicer treatment, RNA fragments were purified using PERFORMA® DTR Gel Filtration Cartridges (#98780, EdgeBio). Briefly, cartridges were centrifuged at 1,000 xg for 4 min to remove the equilibrium buffer in the columns. Then RNA samples were added into the column, followed by spinning down at 1,000 xg for 4 min to collect RNA at the bottom of the tube. Ten picomole of each RNA fragments, including TIF-FL, M229, TIFm71, SY3 and S229 were treated with or without Dicer at 37°C overnight, followed by Northern blot analysis using ^{32}P-labeled oligonucleotide as described in section 4.3.1.

4.3.3 Electrophoretic mobility shift assay

4.3.3.1 Preparation of cell fraction

Cell fraction was prepared as described before (Huang *et al.*, 2006; Wang *et al.*, 2000a). To obtain cytosolic protein, ~1x10^7 HEK293, MCF-7 or CHO-K1 cells (100 mm dish) were trypsinized, rainsed with phosphate buffered saline, and incubated with 200 µl of ice-cold lysis buffer (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5% Nonidet P40, 10% glycerol) supplemented with 0.5 mM DTT and protease inhibitor Cocktail. Lysate was placed on ice for 10 min and centrifuged at 1,200 xg for 5 min at 4°C to remove nuclear proteins. Supernatants were saved, freeze-thawed for 5 times. Insoluble materials were removed by centrifugation at 16,000 xg for 15 min at 4°C. For preparing nuclear fractions, cell lysate was centrifuged at 10,000 xg for 5 min at 4°C to remove cytoplasmic proteins within the supernatant. Nuclear proteins were extracted from the pellet in 200 µl of ice-cold extraction buffer (20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol), in which 0.5 mM DTT and protease inhibitor Cocktail were

added before use. Samples were kept on ice for 2 h with agitation every 15 min. Insoluble materials were removed by centrifugation at 16,000 xg for 15 min at 4°C.

4.3.3.2 Quantification of protein concentration

Protein concentration was measured with the bicinchoninic Acid (BCA) protein assay kit (Pierce) according to the manufacturer's instructions. Serial dilutions of BSA in concentration of 2 mg/ml, 1 mg/ml, 0.5 mg/ml, 0.25 mg/ml, 0.125 mg/ml and 0.0625 mg/ml were applied as protein standards. Protein samples were 4-fold and 8-fold diluted, and 10 µl of each sample as well as serial dilutions of BSA were added to a 96-well plate. Solution A and solution B were mixed in a ratio of 50:1, 80 µl of the mixture were added into each well of the plate and incubated for 10 min at room temperature. The protein concentration was determined by endpoint absorbance at 562 nm with a µQuant microplate reader (Bio-Tek Instruments Inc, USA) and analyzed with KCjunior software (Bio-Tek Instruments Inc). Samples were stored at -70°C as multiple small aliquots to avoid cycles of freeze-thawing.

4.3.3.3 Probe preparation

TIFm71 DNA was PCR-amplified with a forward probe containing a T7 RNA polymerase promoter. Using T7-TIFm71 DNA as template, TIFm71 RNA was generated by *in vitro* transcription. In short, 1 µg of DNA template, 2 µl of 10x Transcription Buffer, 1 µl of 10 mM ATP, 1 µl of 10 mM CTP, 1 µl of 10 mM GTP, 1 µl of 10 mM UTP and 2 µl of T7 RNA polymerase were added into a microtube. Total volume was made up to 20 µl by the addition of RNase free water. The reaction

mixture was thoroughly mixed and incubated at 37°C for 1 h. At the end of the synthesis reaction, 1 µl of RNase-free DNase I was added and kept at 37°C for 15 min to digest the DNA template. RNA was purified using CHROMA SPIN TE-30 columns. Prior to purification, columns were spun at 700 xg for 3 min to remove the equilibration buffer. RNA samples were applied to the center of the gel bed's flat surface, followed by spinning down at 700 xg for 5 min to collect RNA at the bottom of the tube. Concentration of RNA fragments was measured by a spectrophotometer as described in section 2.3.2.3. T4 polynucleotide kinase (T4-PNK) was used to transfer the terminal phosphate group from [γ-^{32}P] ATP to a free 5'-hydroxyl group of RNA fragments. A similar procedure was described in section 4.3.1.4. In each labeling reaction, 4 pmol RNA was labeled. The unincorporated radionucleotides, enzyme and salts were removed by passing the mixture through MicroSpin G-25 columns (#27-5325-01, GE Healthcare).

4.3.3.4 RNA-protein binding reaction

RNA-protein binding mixure was consisted of 1 µl of [γ-^{32}P] ATP-labeled probe, 8 µg or specific amount of protein extracts, 1 µl of RNase inhibitor and 4 µl of 10x binding buffer (100 mM Tris-HCl, 10 mM EDTA, 1 mM DTT, 50% glycerol, pH 7.5). Final volume was made up to 20 µl by adding RNase free water. The mixture was kept on ice for 30 min.

4.3.3.5 Electrophoresis and detection of RNA-protein binding complexes

The RNA-protein binding products were subjected to electrophoresis through a 5% non-denaturing acrylamide gel. To prepare the gel solution, 3 ml of 40% acrylamide stock (the ratio of acrylamide: bis-acrylamide was 79:1), 1 ml of 5x TBE buffer, 100 µl of 10% APS and 20 µl of TEMED were mixed and the final volume was made up to 20 ml by adding RNase free water. The gel was kept still in room temperature for at least 1 h and pre-ran at a constant voltage of 150 V for ~2 h. RNA-protein binding products were loaded onto the gel and ran in the same voltage for another 3-4 h. Finally, the gel was dried and exposed to FUJI medical X-ray film at -70°C.

4.3.4 Biotinylated RNA pull-down assay

4.3.4.1 Biotinylation of RNA

Biotinylated M229, TIFm71 RNA were synthesized by incorporating Biotin-11-UTP during *in vitro* transcription with T7 RNA polymerase. For *in vitro* synthesis of biotinylated M229, TIFm71 and control transcripts, PCR amplicons of target genes under the transcriptional control of T7 RNA polymerase were separated on agarose gels, purified, and used as templates for the synthesis of corresponding biotinylated RNAs using T7 RNA polymerase and biotin-11-UTP. Briefly, 2 µg of DNA template, 4 µl of 5x Transcription Buffer, 1 µl of 10 mM ATP, 1 µl of 10 mM CTP, 1 µl of 10 mM GTP, 0.83 µl of 10 mM UTP, 0.17 µl of 10 mM Biotin-11-UTP (#AM8450, Ambion) and 1 µl of T7 enzyme mix were added into a microtube. The final volume was made up to 20 µl by the addition of RNase free water. The reaction mixture was thoroughly mixed and incubated at 37°C for 1 h. Biotinylated RNAs were treated with 1 µl of RNase-free DNase I at 37°C for 15 min to digest DNA templates and purified

on Chroma Spin TE-30 columns (Clontech). Each biotin-RNA was analyzed by agarose gel electrophoresis and quantified by UV-spectrometry.

4.3.4.2 Preparation of protein extracts

The protocol of subcellular protein extration was developed as described before (Huang *et al.*, 2006; Wang *et al.*, 2000a). To obtain nuclear protein, ~$1x10^7$ HEK293 cells (100 mm dish) were trypsinized, rainsed with PBS, and incubated with 200 µl of ice-cold lysis buffer (10% glycerol, 10 mM KCl, 1.5 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 0.5% Nonidet P40) supplemented with 1 mM PMSF, 0.5 mM DTT and protease inhibitor Cocktail. Cell lysate was kept on ice for 10 min and centrifuged at 10,000 xg for 5 min at 4°C to remove the cytosolic proteins within the supernatant. Nuclear proteins were extracted from the pellet in 200 µl of ice-cold extraction buffer (20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol), in which 0.5 mM DTT, 1 mM PMSF and protease inhibitor Cocktail were added before use. Samples were kept on ice for 2 h with agitation every 15 min. Insoluble materials were removed by centrifugation at 15,000 xg for 15 min at 4°C. To obtain cytosolic protein, cells were incubated with 200 µl of ice-cold lysis buffer (10% glycerol, 10 mM KCl, 1.5 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 0.5% Nonidet P40) supplemented with 1 mM PMSF, 0.5 mM DTT and protease inhibitor Cocktail. The lysate was placed on ice for 10 min and centrifuged at 2,000 xg for 5 min at 4°C to remove nuclear proteins. Supernatants were saved and freeze-thawed 5 times. Insoluble materials were removed by centrifugation at 15,000 xg for 15 min at 4°C. Protein concentration was measured with the bicinchoninic Acid (BCA) protein assay kit

(Pierce) as described in section 4.3.3.2. Samples were stored at -70°C in small aliquots to avoid cycles of freeze-thawing.

4.3.4.3 Biotinylated RNA-streptavidin binding

The biotinylated RNA-streptavidin binding assay was performed using the following reagents per reaction: 10 pmol biotin-RNA representing either M229, TIFm71 or nonspecific RNA, 60 µl of HEK293 nuclear extract or cytosolic protein (10 µg/µl), 65 µl of streptavidin sepharose high performance beads (GE Healthcare), 1 µl of RNasin (20 U/µl), and 5 µg of tRNA (to decrease nonspecific interactions). Streptavidin sepharose beads (40 µl) were washed with 200 µl of cold 1x PBS three times and 50 µl of Buffer D (100 mM KCl, 0.2 mM EDTA, 20 mM HEPES at pH 7.9, 20% [v/v] Glycerol, 1 mM PMSF, 1 mM DTT) once and placed on ice until use. HEK293 cytosolic protein or nuclear extract was pre-incubated with 25 µl of beads with gentle rocking for 15 min at room temperature. The extract was centrifuged at 4,000 xg for 30 sec to remove the beads and the supernatant was transferred to a new tube and placed on ice. Next, 10 pmol of biotinylated RNA oligo (approximately equal to one reaction of 20 µl *in vitro* transcriptional labeling product) and 1 µl of RNasin were added to pre-cleared cytosolic protein or nuclear extract. The mixture was incubated at 30°C for 20 min. A second set of pre-washed streptavidin beads (40 µl) were added to the mixture of RNA-protein and the mixture was incubated on ice with vigorous shaking for an additional 30 min. The reaction mixture was centrifuged for 5 min at 1,000 xg and the supernatant was removed. The beads were washed with 1 ml PBS three times, and 30 µl of 2x Laemmli Buffer (0.004% bromophenol blue, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 125 mM Tris-HCl, pH 6.8) was added. The

bead-bound protein mixtures were boiled for 5 min and loaded onto a 10% denaturing SDS-PAGE gel.

4.3.5 Identification of TIFm71 binding proteins

4.3.5.1 Silver staining of PAGE gel

The protein mixtures were size-fractionated on a 10% SDS-PAGE and stained with silver staining for visualization of protein bands. The gel was first fixed for 30 min to overnight in 40% ethanol and then sensitized with 0.2% sodium thiosulfate, 4.1% sodium acetate and 30% ehtanol solution for 45 min at room temperature. After the sensitization, the gel was washed with autoclaved water for 10 min 3 times, followed by impregnating with 0.1% silver nitrate for 40 min with gentle shaking at 4°C. After impregnation, the gel was rinsed with autoclaved water twice and developed in 2.5% sodium carbonate containing 0.04% formaldehyde solution for less than 10 min in dark. Once the color was developed, the silver staining reaction was stopped by replacing the development solution with 40 mM EDTA and incubated for 15 min with gentle shaking. The gel was then washed with water for 5 min 3 times and digitized with Epson Perfection 2450 scanner at a resolution of 300 dpi.

4.3.5.2 LC-mass spectrometry analysis

Bands of interest were excised and cut into small cubes. Gel pieces were incubated with freshly prepared 100 µl of 25 mM ammonium bicarbonate including 15 mM ferricyanide potassium and 50 mM sodium thiosulphate for 10 min until the yellow color had gone. The gel pieces were then washed with 1 ml of 25 mM ammonium

bicarbonate for 10 min, followed by dehydrating with 600 µl of 25 mM ammonium bicarbonate containing 50% acetonitrile (v/v) for 10 min and dried in Speed-Van evaporator for 5 min. After that, the gel pieces were chemically reduced by swelling in 100 µl of 10 mM DTT/25 mM ammonium bicarbonate for 20 min in dark and alkylated with 100 µl of 55 mM iodoacetamide/25 mM ammonium bicarbonate for another 20 min in dark. The solution was removed and the gel pieces were washed with 1 ml of 25 mM ammonium bicarbonate twice, dehydrated again by 600 µl of 25 mM ammonium bicarbonate containing 50% acetonitrile (v/v) for 10 min, followed by drying in Speed-Van evaporator for 5 min. The dried gel pieces were then rehydrated with 20 µl of proteomics grade trypsin solution (20 ng/µl) and kept on ice for 10 min and 20 µl of 25 mM ammonium bicarbonate was added to cover the gel pieces. This mixture was incubated at 37°C for overnight digestion. By the end of digestion, peptides were extracted by adding 40 µl of 1% formic acid, incubated for 20 min at 37°C, followed with sonication for 10 min. Extracts were collected in a new Lobind tube (Invitrogen). The extraction was repeated once more, and extracts were collected. After two rounds of formic acid extraction, 40 µl of neat acetonitrile was added to the gel pieces, sonicated for 10 min and collected the extracts into the same Lobind tube. The extraction mixture containing the tryptic peptides was spun down and dried in Speed-Van evaporator for 3-4 h. The dried gel pieces were kept in -70°C in case of the failure of the digestion. Peptides were analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as described before (Tong *et al.*, 2010). Briefly, the resulting peptides from each gel slice were separated by HPLC (Dionex, Sunnyvale, CA) on a commercial C18 reverse phase column (inner diameter 75 µm, 5 µm Acclaim pepMap100 medium; Dionex) over an 80-minute gradient (Mobile phase A: 0.1% fluoroacetic acid in 2% ACN in Milli-Q water;

mobile phase B: 0.1% fluoroacetic acid in 98% ACN) and then analyzed by a micrOTOF-Q II ESI-Qq-TOF mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany).

4.4 Results

4.4.1 TIFm71 releasing in transiently transfected cells

As discussed in Chapter 3, in order to investigate whether TIFm71 was a pre-miRNA or an ncRNA that released from TIF mRNA, several fragments were subcloned downstream of CMV promoter of pcDNA3 vector as shown in Fig. 4.1A. In pri-miRNA to pre-miRNA processing, flanking sequence of stem-loop structure may facilitate the recognition and contain the binding sites for Drosha or DGCR8 (Ohler *et al.*, 2004; Zeng and Cullen, 2005). Therefore, a 229 bp fragment M229 containing TIFm71 together with 5' and 3' flanking regions was subcloned. Scrambled-M229 (S229), which contained the same flanking sequence as M229 but a scrambled TIFm71 sequence to avoid the formation of stem-loop secondary structure, was synthesized as a negative control fragment. Besides, TIFm71, SY3 (TIF that lacking TIFm71 in 3'-UTR), and full-length TIF sequences were subcloned.

Following the transient transfection of HEK293 cells with those expression vectors, total RNA was subsequently extracted for polyacrylamide gel Northern blot analysis as described in section 4.3.1. The result revealed that TIF, M299 and TIFm71 constructs produced TIFm71, but SY3, S229 and pcDNA3 empty control did not (Fig. 4.1B). Also, TIFm71 was detected in the oncogene *mas* over-expressing stable cell line McOM80, in which TIF gene was greatly up-regulated (Lin *et al.*, 2009). These results strongly indicated that TIFm71 was processed from TIF transcript.

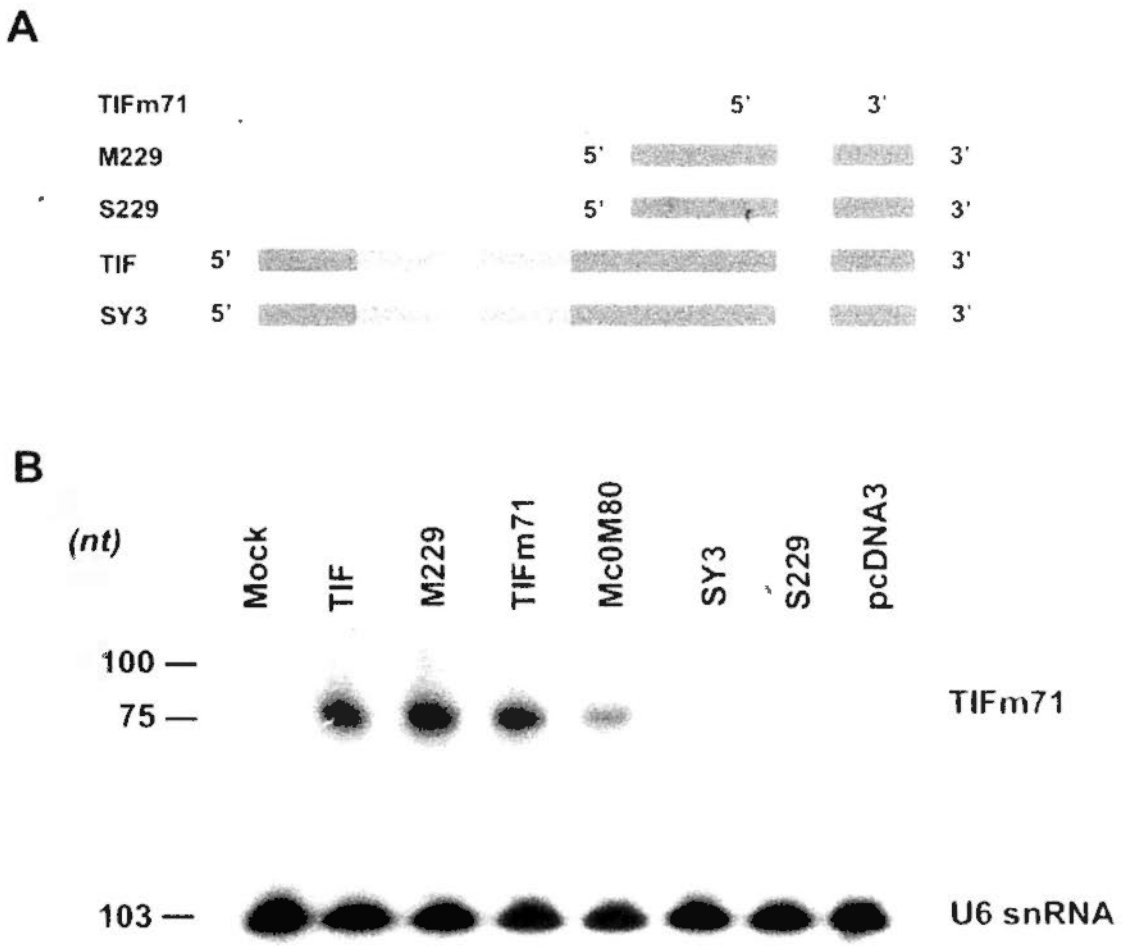


Figure 4.1 TIFm71 was released from TIF transcript. (A) Five fragments, including TIFm71, TIFm71 with flanking sequence (M229), scrambled TIFm71 with flanking sequence (S229), full-length TIF and SY3 were PCR amplified and subcloned into pcDNA3 vector. (B) TIFm71 was processed in transiently transfected HEK293 cells. Sequences shown in (A) were transfected into HEK293 cells. After 24 h, total RNA was extracted and subjected to polyacrylamide gel Northern blot analysis to examine the TIFm71 expression. Mc0M80 cells RNA was loaded as the positive control. U6 RNA was used as control to indicate equal amount of total RNA being load. The data shown was representative of 3 separate experiments with similar results.

In the miRNA region of the blot, only a very weak signal in ~20 nt was recognized in M229 transiently transfected HEK293 cells (Fig. 4.2, left arrow), while wide type full-length TIF transcript or TIFm71 failed to produce any detectable levels of this miRNA species by polyacrylamide gel Northern blot analysis with radio-labeled 3'-stem of TIFm71 probe (Fig. 4.2). Meanwhile, a relatively high level of miRNA (TIF-miR) expression was observed in Mc0M80 cells (Fig. 4.2, right arrow). These results implied that TIFm71 might be a potential pre-miRNA; it did not process into miRNA under normal conditions but acted as an ncRNA.

The same miRNA was also detected in the solid tumor tissues, indicating that TIF-miR maybe an inducible miRNA. To investigate whether hypoxia is an induction of TIF-miR expression, HEK293 cells with transient over-expression of TIFm71 were cultured in a hypoxic condition with 2% oxygen (Galaxy® 48R Incubator) for 16 h. However, Northern blot analysis indicated that hypoxia did not induce TIF-miR formation (data not shown).

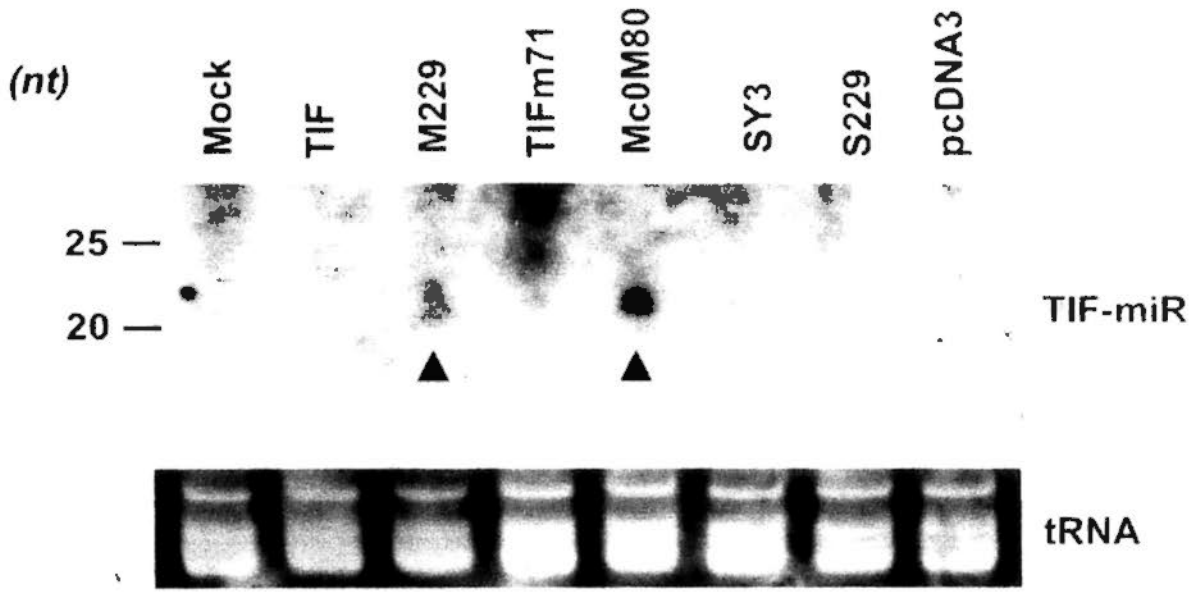


Figure 4.2 Detection of miRNA processed from TIFm71. Expression vectors shown in Fig. 4.1A were transfected into HEK293 cells. Twenty-four hours after transfection, total RNA was extracted 20 μg of the total RNA was subjected to polyacrylamide gel Northern blot analysis to examine the expression of TIF-miR. Mc0M80 cells RNA was loaded as the positive control. Ethidium bromide staining of tRNA was shown to indicate RNA quality.

4.4.2 TIFm71 *in vitro* processing by Dicer

Multiple studies suggested that ribonuclease III enzyme Dicer is one of the most critical proteins in the processing of mature and functional miRNAs (Bartel, 2004). In cytoplasm, Dicer cuts the hairpin loop pre-miRNA to release mature miRNA, which is incorporated into the RISC to impede mRNA translation into protein and induce RNA degradation (Esquela-Kerscher and Slack, 2006). To investigate whether TIFm71 is a potential target of Dicer, an *in vitro* processing assay was performed by incubating TIFm71 and its related RNA fragments with or without Dicer protein, followed by polyacrylamide gel Northern blot analysis using radio-labeled 3'- stem of TIFm71 probe (Fig. 4.3). After the treatment of Dicer, small RNA (~20 nt) was generated in TIFm71 and TIFm71 containing fragments (Fig. 4.3, lane 6-8), suggesting that TIFm71 could serve as a substrate of Dicer. An additional hybridization signal at ~50 nt was observed that might represent a processing intermediate.

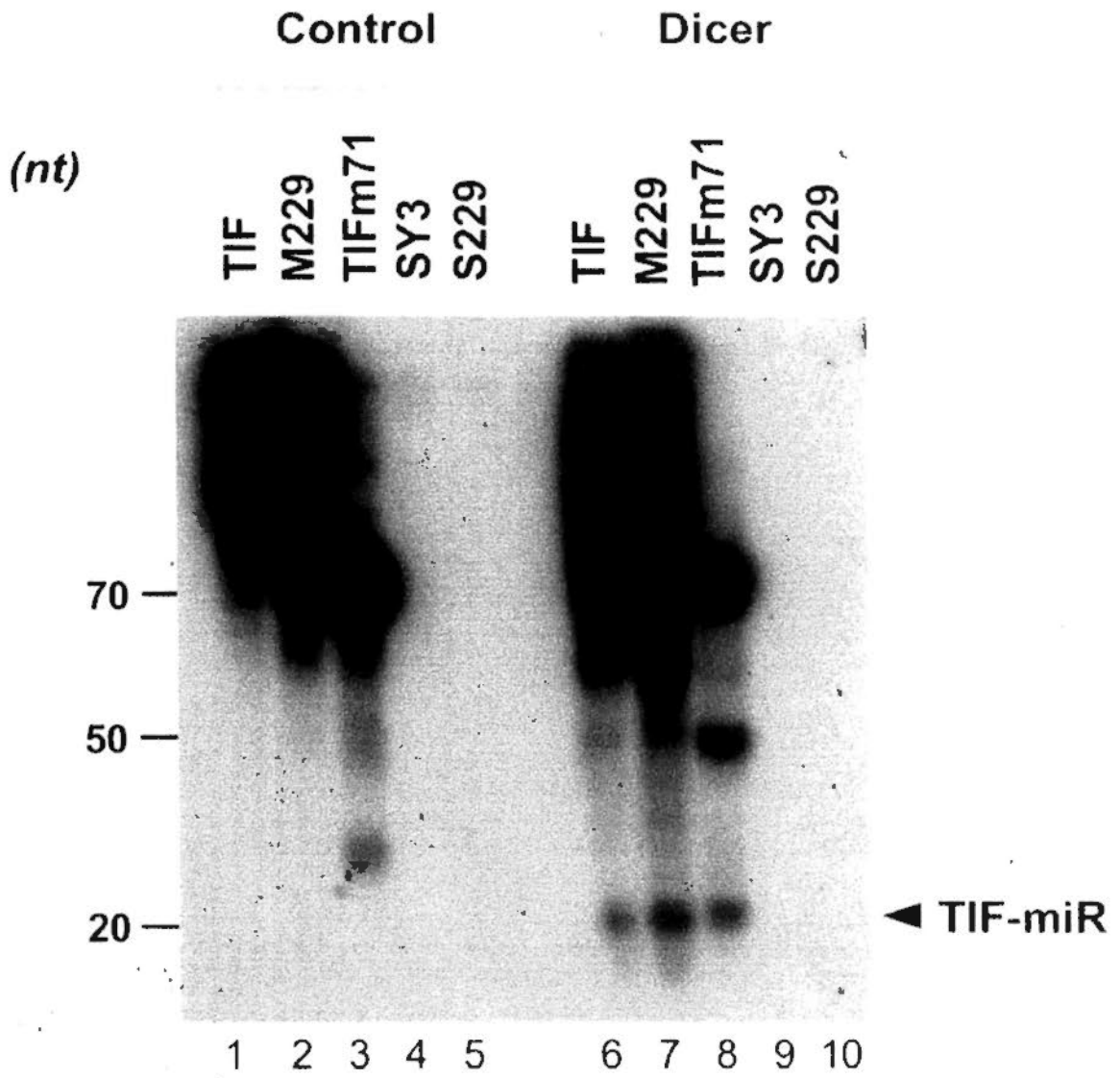


Figure 4.3 Dicer *in vitro* processing assay. By in vitro transcription, five RNA fragments, including full-length TIF, M229, TIFm71, SY3 and S229 were generated, and incubated at 37°C with or without Dicer for 16 h. Small RNA was generated after treating with Dicer (lane 6-8) detected by Northern blot analysis.

4.4.3 TIFm71 RNA electrophoretic mobility shift assay (EMSA)

One of the most important function of ncRNA is catalysing biological reactions through binding to proteins and modulating their activity (Goodrich and Kugel, 2006). ncRNAs usually fold into unique higher-order structures to become integral parts of large complexes that contain proteins or possibly other RNAs, and the components of the complex function together as a unit (Goodrich and Kugel, 2006). In order to identify the protein binding activity of TIFm71 RNA, electrophoretic mobility shift assay (EMSA) was carried out. EMSA is classically used to detect DNA/RNA binding proteins, the tenet of the EMSA is that DNA/RNA with protein bound migrates more slowly than the corresponding free unbound DNA/RNA through a polyacrylamide gel (Holden and Tacon, 2010).

TIFm71 RNA was *in vitro* transcribed by T7 RNA polymerase using TIFm71 DNA template containing a T7 RNA polymerase promoter. T4-PNK was used to transfer the terminal phosphate group from [γ-^{32}P] ATP to a free 5'-hydroxyl group of TIFm71 RNA fragment. The radiolabeled RNA was then incubated with different amounts of HEK293 nuclear protein extract. The RNA-protein binding products were subjected to electrophoresis analysis through a 5% non-denaturing acrylamide gel.

As shown in Fig. 4.4, three major RNA-protein binding complexes were formed, and the binding activity increased with the elevation of protein amount. However, the binding signals smeared in lanes, and very large complex formed on the top of the acrylamide gel. Therefore we could not distinguish the binding proteins from EMSA,

probably because TIFm71 could bind to too many proteins by folding into complicated secondary structure *in vitro*.

Binding activity of TIFm71 RNA in the nuclear fraction of CHO-K1 and MCF-7 cells nuclear extracts were also examined by EMSA. Although background was high, binding complex 2 was detected in all of the three cell extracts as shown in Fig. 4.5. The binding activity of TIFm71 in cytoplasmic fraction of these cells showed a similar pattern (data not shown).

These results suggested TIFm71 could bind to many proteins, and possibly affect the activities of these proteins to affect the morphology of CHO-K1 cells as discussed in Chapter 3.

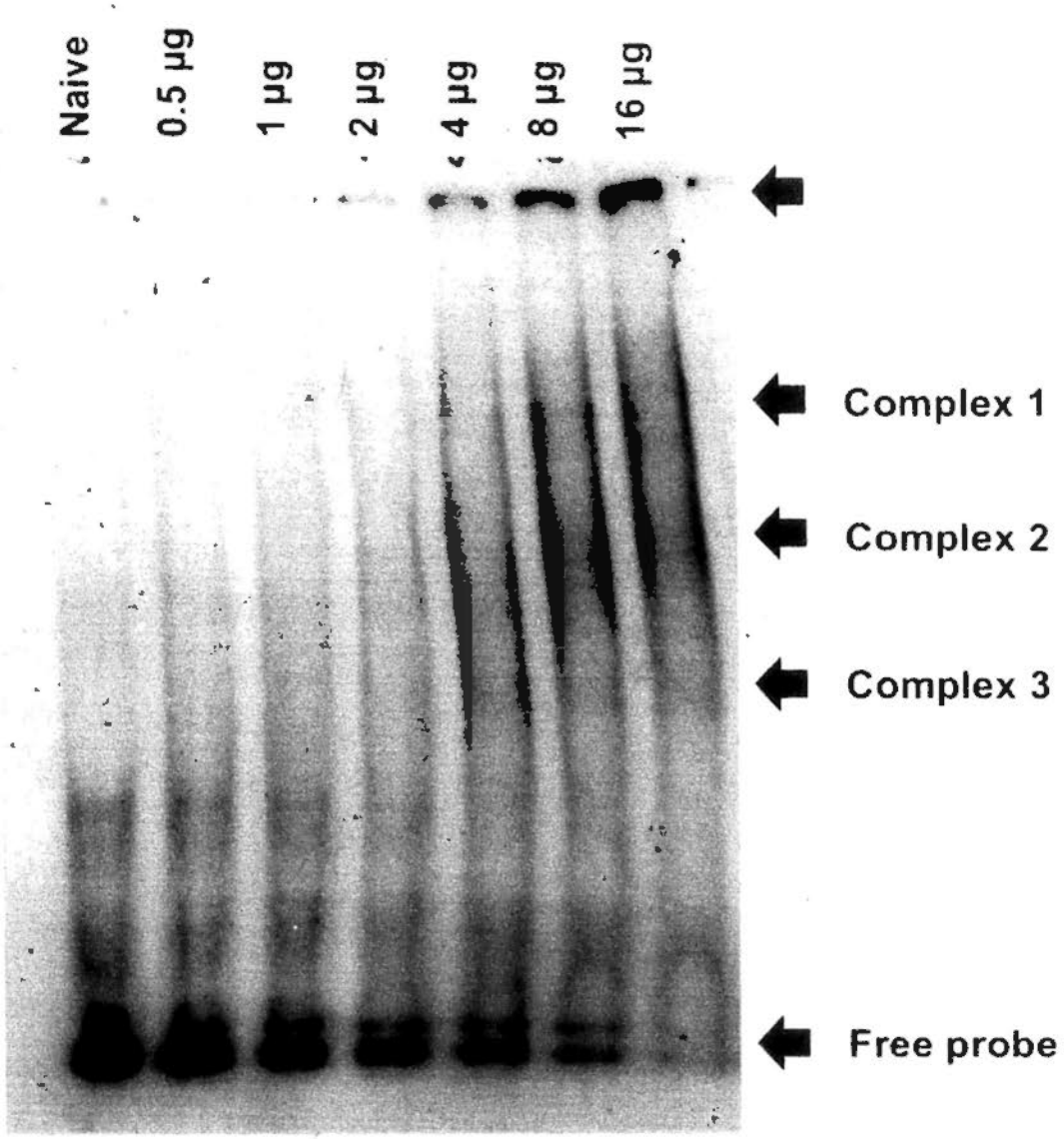


Figure 4.4 Binding activity of TIFm71 RNA in the nuclear fraction extracted from the HEK293 cells analyzed by EMSA. End-labeled TIFm71 RNA was incubated with different amount of HEK293 nuclear protein extract or without protein extract (Naive) as described in Method section. Excess unbound probes and protein-RNA complexes were marked with arrows.

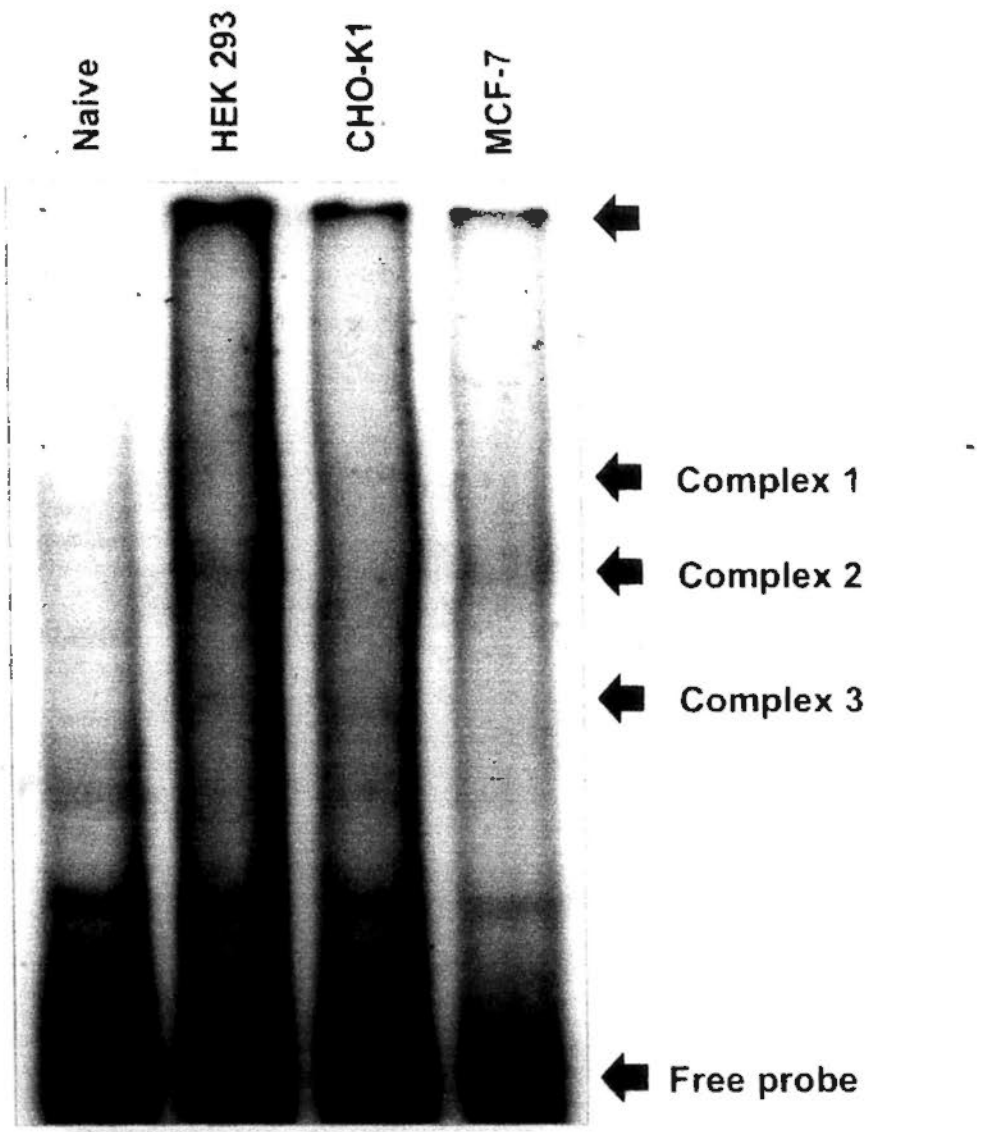


Figure 4.5 Binding activity of TIFm71 RNA in the nuclear fraction extracted from the HEK293, CHO-K1 and MCF-7 cells analyzed by EMSA. End-labeled TIFm71 RNA was incubated with nuclear protein extracts (8 μg) or without protein extract (Naive) as described in Method section. Excess unbound probes and protein-RNA complexes were marked with arrows.

4.4.4 Biotinylated TIFm71 RNA pull-down assay

We hypothesized that TIFm71 might act as a scaffold to recruit EMT related proteins and affect their biological activities. Result of EMSA suggested TIFm71 could bind to many proteins and forming RNA-protein complexes. As for the small amount of complexes detected, it's difficult to identify the binding proteins based on the EMSA results. Then we turned to biotinylated RNA pull-down assay.

As shown in Fig. 4.6, biotinylated M229, TIFm71 RNA were synthesized by incorporating Biotin-11-UTP during *in vitro* transcription with T7 RNA polymerase using M229/TIFm71 DNA template containing a T7 RNA polymerase promoter. The biotin-labeled RNAs were allowed to bind to proteins in the cell extract, followed by incubating with streptavidin agarose resins that bound biotin with high affinity. Proteins specifically bound to RNAs were precipitated with the complex composed of RNA-protein-Streptavidin agarose resins, and then analyzed by SDA-PAGE.

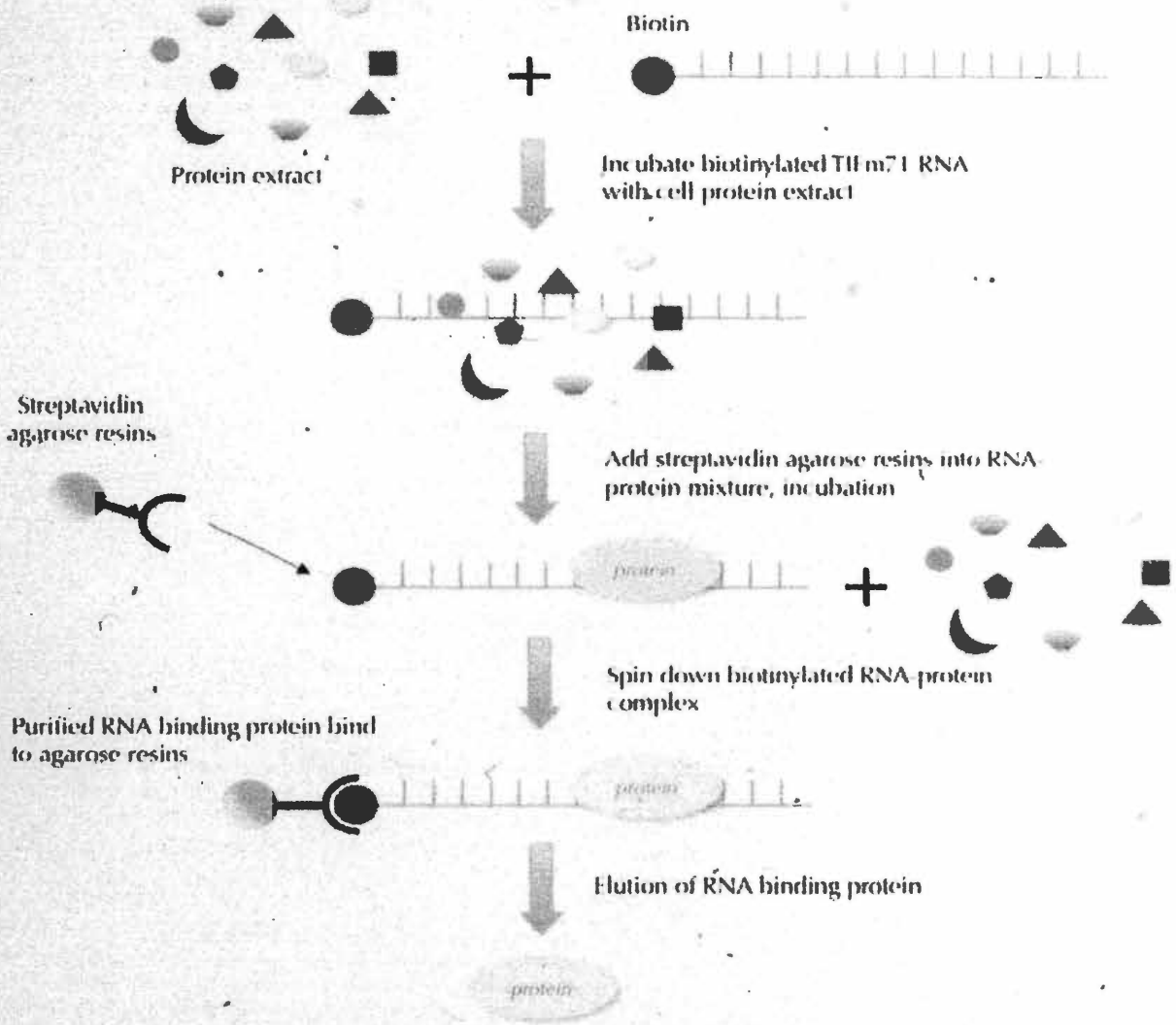


Figure 4.6 Experimental scheme for RNA pull-down assay.

To optimize the biotin labeling, M229 RNA were labeled in several different ratios of normal UTP to Bio-UTP and used in small-scale pull-down assay. It could be noticed that the size of RNA was increased with an increase of Bio-UTP to UTP ratio as analyzed by agarose gel electrophoresis (Fig. 4.7A), indicating that the biotin labeling was successful. These RNAs were incubated with cytosolic extracts of HEK293 cells, and specific proteins were precipitated with immobilized Streptavidin. The captured protein complexes were size-fractionated on a small-scale 10% SDS-PAGE and stained with silver staining (Fig. 4.7B). A specific protein was seen as indicated by a black arrow in Fig. 4.7B, suggesting the ratio of 1:5 for Bio-UTP to normal UTP was the most effective. Therefore this ratio was chosen in the following large scale RNA pull-down analysis.

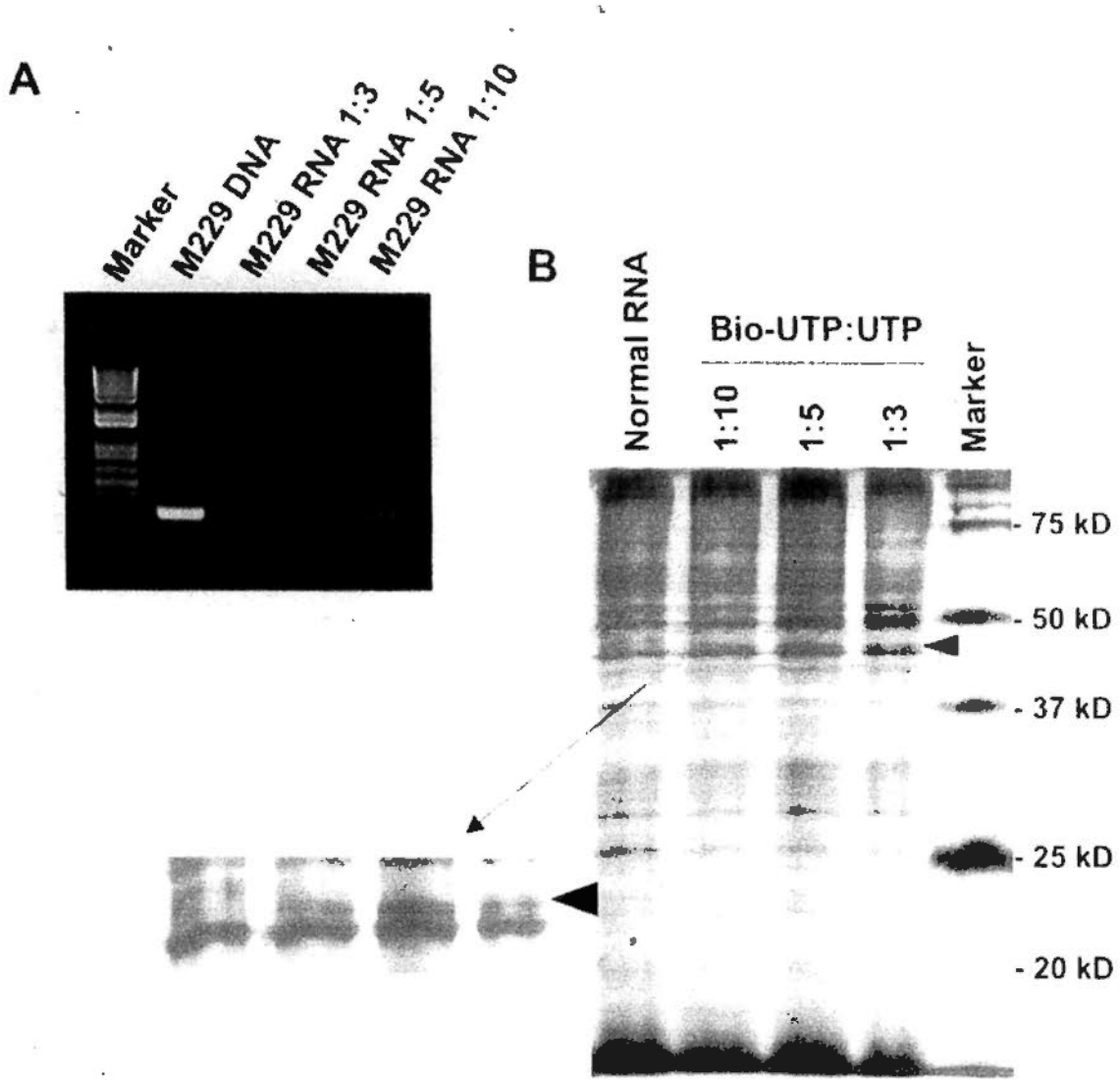


Figure 4.7 Optimization of the biotin labeling condition. (A) Agarose gel electrophoresis of biotin-labeled M229 RNAs in different Bio-UTP to normal UTP ratio. Marker was 1 kb plus DNA ladder and M229 DNA was loaded as a control. (B) RNA pull-down. HEK293 cytosolic protein (600 μg) was mixed with biotin-labeled TIFm71 RNA and incubated at 30°C for 20 min. RNA bound proteins were pull-down by Streptavidin sepharose high performance beads (40 μl). Proteins were analyzed in 10% SDS-PAGE followed with silver staining. Protein marker was Precision Plus Protein™ Standards (#161-0374).

We next performed large-scale experiments of RNA preparation and affinity capture, and the captured proteins were analyzed in large SDS-PAGE gel. Eight protein bands (Band 1-8) were found to specifically bind to TIFm71 RNA compared to that of the scrambled TIFm71 control as indicated by arrows in Fig. 4.8. For the M229 RNA, three more bands, namely bands 9, 10 and 11, were observed besides the bands that were binding to the TIFm71 region (Fig. 4.9), implying that these proteins were binding to the flanking sequence on M229, and also this might be responsible for the inducible expression of TIFm71 in tumor tissues (Fig. 3.4B).

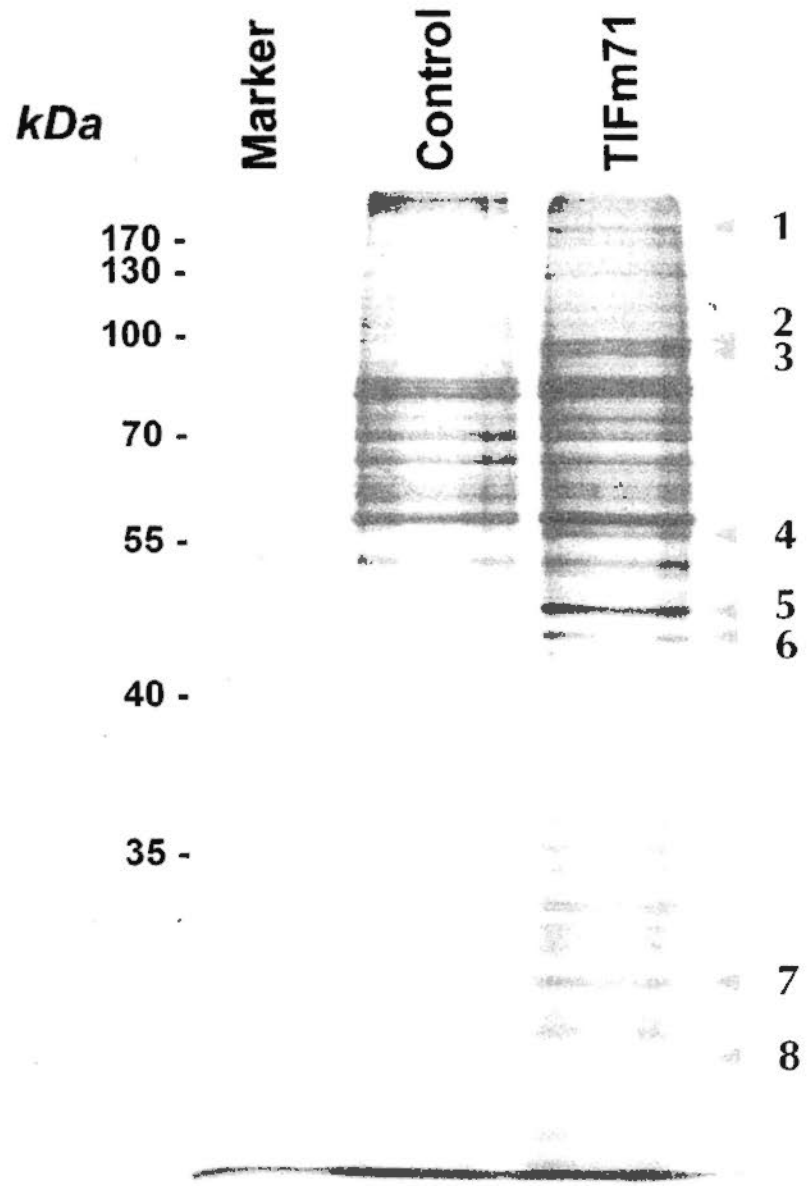


Figure 4.8 Potential TIFm71 RNA binding proteins examined by RNA pull-down and SDS-PAGE. HEK293 cytosolic extracts (600 μg) were applied in the pull-down assay using biotin-labeled TIFm71 RNA as described in the Methods section. The protein mixtures captured were denatured in the loading buffer and size-fractionated by electrophoresis of 10% SDS-PAGE. The gel was stained with silver staining for visualization of protein bands. Protein marker was PageRuler™ Plus Prestained Protein Ladder (#SM 1811). Figure shown was representative of 4 separate experiments with similar results.

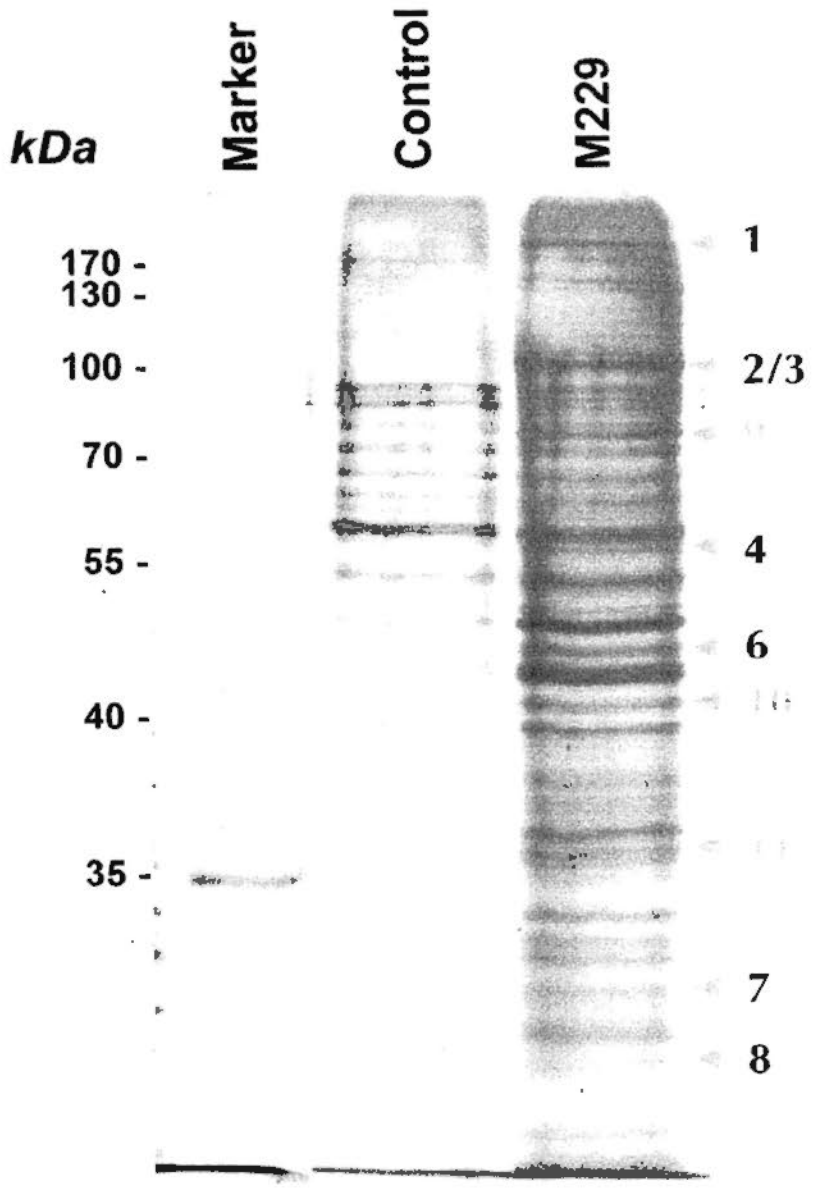


Figure 4.9 Potential M229 RNA binding proteins examined by RNA pull-down assay and SDS-PAGE. HEK293 cytosolic extracts (600 μg) were applied in the pull-down assay using biotin-labeled M229 RNA as described in the Methods section. The protein mixtures captured were size-fractionated by 10% SDS-PAGE and stained with silver staining for visualization of protein bands. Protein marker was PageRuler™ Plus Prestained Protein Ladder (#SM 1811). Figure shown was representative of 3 separate experiments with similar results.

4.4.5 M/S analysis on TIFm71/M229 binding proteins

After RNA pull-down, proteins were separated by one-dimensional gel, and bands of interest were excised, in-gel digested and subjected to LC/MS/MS proteomics analysis. Data were analyzed using the MASCOT 1.9 search engine (Matrix Science) to search against the human (148,148 sequences) MSDB protein database (Lam *et al.*, 2010).

There was a total 15 unique proteins identified (Table 4.1). Usually each band of interest excised from the 1D SDS-PAGE gel contained more than one protein. For example, five proteins from band 7, three proteins from band 1 and two proteins from band 2, 3 and 4, were identified. However, proteins from band 5, 6, 8, 9, 10 and 11 were not detected by LC-MS analysis, probably due to the insufficient protein amount and subsequent lost during extraction.

Among these proteins identified, several have RNA binding activities as indicated by previous studies, including Ras-related GTP-binding nuclear protein (Ran), Interleukin enhancer binding factor 3 (ILF-3), M-phase phosphoprotein 4 (MPP4), RNA helicase A (RHA), heterogeneous nuclear ribonucleoprotein H2 (hnRNP H2), and G-rich sequence factor 1 (GRSF 1).

Ran (band 7) is a small GTP binding protein belonging to the RAS superfamily. It is involved in cell cycle progression through the regulation of mitotic spindle formation (Clarke and Zhang, 2008). Besides, it was proposed that the deregulation of Ran

expression may lead to genomic instability and may have a significant effect in cell transformation or cancer progression (Azuma *et al.*, 2004). Moreover, Ran is essential for the translocation of RNA and proteins through the nuclear pore complex, and also involved in the control of DNA synthesis (Sorokin *et al.*, 2007; Stewart, 2007).

ILF-3 (band 2) and MPP4 (band 3) are isoforms of each other. MPP4 is 611 amino acids in length, which recognizes a phosphoepitope present on a set of 40-50 kDa proteins that become phosphorylated at the G2/M transition (Abaza *et al.*, 2003; Matsumoto-Taniura *et al.*, 1996), probably functions in the cell cycle regulation. While ILF-3, also referred to as nuclear factor 110 (NF110), is a 894 amino acids protein contains 282 more amino acids than MPP4 in the C'-terminal. It is a double-stranded RNA (dsRNA) binding protein which complexes with other proteins, dsRNAs, small ncRNAs and mRNAs to regulate gene expression, transport and stabilize RNAs. ILF-3 protein presents two dsRNA binding domains (dsRBDs), as well as an RGG domain (arginine-glycine-rich domain). It has been demonstrated that ILF-3 regulate several promoter activities, either positively or negatively, depending on the promoter context (Reichman and Mathews, 2003; Reichman *et al.*, 2003).

RHA (band 1) is a member of the DEAH-box DNA/RNA helicases. It utilizes the energy of NTP binding and hydrolysis to remodel RNA or RNA-protein complexes, resulting in RNA duplex strand separation, displacement of proteins from RNA molecules, or both. RHA is ubiquitously expressed and participates in many cellular processes including transcription, mRNA splicing, nuclear export/translation and

RNA interference. However, it remains unclear how it targets to the right substrates *in vivo* and how its activity is regulated (Hahn and Beggs, 2010; Seeburg, 2000).

hnRNP H2 (band 4) belongs to the subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs), which are RNA binding proteins complex with heterogeneous nuclear RNA (hnRNA). While all of the hnRNPs are present in the nucleus, some of them seem to shuttle between the nucleus and the cytoplasm. hnRNP H2 has three repeats of quasi-RRM domains that bind to RNAs (Stark *et al.*, 2011). It's reported that it was associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA cleavage, polyadenylation and other aspects of mRNA metabolism and transport (Alkan *et al.*, 2006; Bagga *et al.*, 1998).

GRSF-1 (band 4) is a member of the RNP superfamily that binds RNAs containing the G-rich elements (Cassetti *et al.*, 2001). It is implicated in a wide variety of cellular processes, including RNA processing, nuclear export, trafficking, mRNA stability, and mRNA translation (Burd and Dreyfuss, 1994; Cuesta *et al.*, 2000; Siomi and Dreyfuss, 1997). The protein is localized in the cytoplasm, and has been shown to bind specifically to conserved viral 5'-UTR sequences and stimulated translation of viral 5'-UTR driven mRNAs *in vitro* (Park *et al.*, 1999). Multiple transcript variants encoding different isoforms have been found for *GRSF-1* gene (Hendrickson *et al.*, 2010).

Apart from classical proteins with clear involvement in RNA-protein interaction, proteins previously not known to be involved in RNA binding were also discovered (Table 4.1), such as X-ray repair cross-complementing protein 5 (XRCC5, band 2), Acetyl-CoA Carboxylase 1 (ACC-1, band 1) isoform 2, Glutathione S-transferase P (GSTP1, band 7) and so on.

Table 4.1 Identification of TIFm71/ M229 RNA binding proteins

Band	M/S result	Accession number	Peptides identified	Sequence coverage (%)	Actual protein size (kDa)	Estimated protein size (kDa)
1	Acetyl-CoA Carboxylase 1 (ACC-1) isoform 2	gi|38679967	14	8.44	267.1	~200
	RNA helicase A (RHA)	gi|307383	10	8.37	143.4	
	Acetyl-CoA Carboxylase 2 (ACACB)	gi|71051955	2	3.15	97.8	
2	X-ray repair cross-complementing protein 5 (XRCC5)	gi|10863945	16	22.13	83.2	~100
	M-phase phosphoprotein 4 (MPP4)	gi|1770458	9	21.93	66.6	
3	PI-3-kinase-related kinase SMG-1	gi|14132744	2	0.89	343.7	~100
	Interleukin enhancer binding factor 3 (ILF-3)	sp|Q12906|ILF3	11	14.99	95.4	
4	Heterogeneous nuclear ribonucleoprotein H2 (hnRNP H2)	sp|P55795|HNRH2	1	2.23	49.3	< 60
	G-rich sequence factor 1 (GRSF 1)	sp|Q12849|GRSF1	1	1.89	53.1	
5	N/A*	N/A	N/A	N/A	N/A	~50
6	N/A	N/A	N/A	N/A	N/A	~50
7	GTP-binding nuclear protein Ran	sp|P62826|RAN	4	20.83	24.4	~30
	ras-related protein Rab-6B/11B	sp|Q15907|RB11B	3/3	13.94/15.14	23.5/24.5	
	Glutathione S-transferase P (GSTP1)	sp|P09211|GSTP1	2	15.24	23.4	
	Thioredoxin-dependent peroxide reductase, mitochondrial	sp|P30048|PRDX3	2	8.98	27.7	
	proteasome subunit beta type-4	sp|P28070|PSB4	2	6.44	29.2	
8	N/A	N/A	N/A	N/A	N/A	~30
9	N/A	N/A	N/A	N/A	N/A	~80
10	N/A	N/A	N/A	N/A	N/A	~42
11	N/A	N/A	N/A	N/A	N/A	~36

RNA pull down proteins were separated on denatured SDS-PAGE and then subjected to in-gel trypsin digestion, followed with proteomics analysis by mass spectrometry. This table listed all the proteins identified from TIFm71/M229 RNA pull-down assay. Each protein was listed with the protein name, accession number, number of peptides identified, actual protein size (calculated based protein sequence) and estimated protein size (indicated by protein marker in SDS-PAGE gel). *N/A = no data available.

4.5 Discussion

The 3'-UTR of mRNA was believed to influence mRNA stability, localization, and translational efficiency by interacting with RNA-binding proteins and/or small regulatory RNAs (Mangone *et al.*, 2010). Our present study revealed a new function of 3'-UTR. We showed that a small ncRNA TIFm71 was released from the 3'-UTR of an exonic transcript TIF, which also encodes a functional CXC chemokine protein (Lin *et al.*, 2009).

Our finding was consistent with a recent study that a diversity of 5'-modified long and short RNAs were generated from processed mRNAs by deep sequencing (Affymetrix, 2009). The 5' cap is a specially altered nucleotide at the 5' end of precursor mRNA and some other primary RNA transcripts as found in eukaryotes. The process of 5'-capping is vital to create mature mRNA, it ensures the messenger RNA's stability while it undergoes translation in the process of protein synthesis (Kapp and Lorsch, 2004). The cap structure prevents 5' degradation and increases the half-life of the mRNA, mainly through blocking the degradation by 5' exonucleases and the access of decapping enzymes to the cap (Parker and Sheth, 2007). The processing of mature mRNAs through a yet unknown mechanism may generate complex populations of both long and short RNAs whose apparently capped 5' ends coincide (Affymetrix, 2009).

Smad binding element (SBE) is a consensus sequence (5'-CAGAC-3') found in the promoter regions of transforming growth factor-β (TGFβ) target genes, such as

germline Igα constant region, plasminogen activator inhibitor type 1 (PAI-1) and TGFβ1 (Dennler *et al.*, 1998; Massague *et al.*, 2005). A recent study reported that a set of miRNAs (T/B-miRNAs) which was regulated posttranscriptionally by TGFβ and BMP signaling contain a conserved sequence similar to SBE (R-SBE). Smad protein could directly associate with R-SBE and facilitate the maturation of T/B-miRNAs (Davis *et al.*, 2010). In our present study, the R-SBE-like (5'-ACAGAG-3') sequence was found within the 3' end and immediately in front of the 5' end of TIFm71. However, it was not essential for TIFm71 releasing in the transiently transfected HEK293 cells, as TIFm71 itself can be released from vector (Fig. 4.1B). We cannot be excludeteh possibility that R-SBE-like sequences are responsible for the inducible up-regulation in solid tumors.

As discussed in Chapter 3, AU-rich elements (ARE) are important *cis*-acting short sequences in the 3'-UTR that mediate the recognition of an array of RNA-binding proteins, affecting mRNA stability and translation (Khabar, 2010). Based on their sequence features, AREs can be broadly divided into three distinct classes. Class I AREs contain multiple independent repeats of the AUUUA pentamer coupled with a nearby U-rich region or U stretch; class II has at least two overlapping copies or closely juxtaposed the nonamer UUAUUUA(U/A)(U/A) motifs in a U-rich region; while class III has a AU or U-rich region but lacks a core AUUUA element (Chen and Shyu, 1995; Novotny *et al.*, 2005). As shown in Fig. 3.5A, a pair of type II AREs (5'-UAUUUAUUGGUAUUUUU -3') were found in -27 nt and +58 nt of TIFm71. We propose that it probably involved in TIFm71 processing or acted as a *cis*-element in the regulation of TIF mRNA stability or translational efficiency (Khabar, 2010; Piecyk *et al.*, 2000). Therefore, the pair of ARE sequences were included in M229

sequence as described in Chapter 2. However, our result showed that the AREs in the 3'-UTR of TIF were not essential for TIFm71 releasing within the transiently transfected HEK293 cells.

Our preliminary data have revealed that TIFm71 could fold into a canonical pre-miRNA-like stem-loop secondary structure. Therefore we initially hypothesized that TIFm71 could generate a mature miRNA. As discussed in Chapter 1, conserved biogenesis machinery governs the production of canonical miRNAs in invertebrate and vertebrate cells (Kim *et al.*, 2009). In nucleus, pri-miRNA transcripts containing one or more hairpin structures are first cleaved by the Drosha RNase III enzyme and associated factors 8 (DGCR8)/Pasha to yield ~55-80 nt pre-miRNA hairpins. Then the pre-miRNAs are cleaved again in the cytoplasm by the Dicer RNase III enzyme to yield a miRNA/miRNA* duplex, of which the mature miRNA is associated with Ago protein of the RISC complex. Results of Dicer *in vitro* processing assay showed that small RNA (~20 nt) was generated in TIFm71 and TIFm71 containing fragments, suggesting TIFm71, but not SY3 or S229, could serve as a substrate of Dicer.

A canonical mature miRNA should be expressed as a distinct transcript of ~22 nt in cells or tissues which is detectable by Northern blot analysis, real-time PCR or other experimental methods (Berezikov *et al.*, 2006). After transiently transfected into HEK293 cells, only M299 generated a small RNA in ~21 nt but not the full-length TIF transcript or TIFm71. Although the only signal derived from M229 was a weak one and was detected only once in three experiments, the miRNA expressed in *mas* oncongene over-expressing Mc0M80 cells proved that TIF-miR was a real mature

miRNA. Another evidence was the observation of small size RNAs (~21 nt) detected in the solid tumor tissues (Fig. 3.4B, tumor 3 and 4), indicating that TIF-miR maybe an inducible miRNA, which possibly contributes to tumorigenesis.

A key question remains as to whether the small RNA TIFm71 arising from the internal exon of TIF transcript represented evolutionary debris to be degraded after splicing excision. In our study, a striking change in cellular morphology by the over-expression of TIFm71 argues against the idea that TIFm71 only represents simple degradation intermediates. As mentioned in the Introduction section, ncRNAs usually elicit their biological responses through interactions with RNA-binding proteins (Charon *et al.*, 2010). Therefore, we hypothesized that TIFm71 might act as a scaffold to recruit EMT related proteins and affect their biological activities. In order to prove that TIFm71 can serve as a scaffold, we tried EMSA first. Smeared bands together with three major complexes were detected in EMSA, indicating that many proteins could potentially bind to TIFm71 RNA. Then we turned to biotinylated RNA pull-down assay, followed with protein identification by mass spectrometry. A battery of TIFm71 RNA binding proteins was identified, which probably facilitated TIFm71 in eliciting its biological functions. Among these proteins, Ran and MPP4 attracted our attention. Previous studies implied that both of them are involved in the control of cell cycle progression (Abaza *et al.*, 2003; Clarke and Zhang, 2008; Matsumoto-Taniura *et al.*, 1996). The deregulation of Ran expression may also have a significant effect in cell transformation or cancer progression (Azuma *et al.*, 2004; Sorokin *et al.*, 2007; Stewart, 2007). Therefore, we proposed that TIFm71 would influence the functions of these proteins through direct interaction to induce EMT and affect cell cycle.

Chapter 5

Functional Characterization of TIFm71

5.1 Introduction

An epithelial-to-mesenchymal transition (EMT) is a biologic process that allows a polarized epithelial cell, which normally interacts with basement membrane via its basal surface, to undergo multiple biochemical changes that enable it to assume a mesenchymal cell phenotype. As shown in Fig. 5.1, the changes in EMT include enhanced migratory capacity, invasiveness, elevated resistance to apoptosis and greatly increased production of ECM components (Kalluri and Neilson, 2003; Kalluri and Weinberg, 2009). EMT is a pivotal cellular program in regulating cellular plasticity in normal adult tissues and tumors, where multiple distinct cellular subpopulations can be generated to contribute to intratumoral heterogeneity. Also, EMT is aberrantly activated in cancer cells to promote their malignant and stem cell characteristics (Kang and Massague, 2004; Polyak and Weinberg, 2009).

One of the critical steps driving EMT is the repression of E-Cadherin, resulting in loss of cell-cell recognition and adhesion. E-Cadherin is expressed in most epithelial cells in which adherent junctions are formed to create the multicellular organization (Barnes *et al.*, 2010; Pertz *et al.*, 1999). Snail is a transcriptional factor that was found to play a central role in the regulation of EMT (Peinado *et al.*, 2007). A key underlying mechanism for this effect is that Snail suppresses the expression of several cellular adhesion proteins, including E-Cadherin (Batlle *et al.*, 2000; Cano *et al.*, 2000). Also, Snail is an important regulator for the development and progression of cancer (Nieto, 2002).

The cell cycle is a series of events that occur in dividing cells between the completion of one mitotic division and the completion of the next division (Fig. 5.2A). This cycle includes accurate duplication of the genome during the DNA synthesis phase (S phase), segregation of complete sets of chromosomes to each of the daughter cells (M phase). The somatic cell cycle also contains "Gap" phases, known as G1, which connects the completion of M phase to initiation of S phase in the next cycle. Dependent on environmental and developmental signals, cells in G1 may temporarily or permanently leave the cell cycle and enter a quiescent or arrested phase known as G0. Another "Gap" phase is G2, which separates the S phase and M phase (Ferrell *et al.*, 2011; Maiato, 2010). The effect of TIFm71 on cell cycle of CHO-K1 cells was evaluated by DNA flow cytometric analysis, and a typical result of flow analysis on cellular DNA content was shown in Fig. 5.2B.

In our present study, we observed a striking change in cellular morphology by the over-expression of TIFm71, whereby the cobblestone-like appearance of epithelial cells switched to a fibroblast-, spindle-like morphology with extensive cellular scattering, which are hallmarks of EMT. In addition, TIFm71 RNA binding proteins, including Ran and MPP4, were identified by RNA pull-down followed with proteomics analysis by mass spectrometry. Those proteins were reported to be involved in the control of cell cycle progression. Therefore, we hypothesized that the over-expression of TIFm71 could promote EMT and affect cell cycle distribution, probably by acting as a scaffold to recruit EMT related proteins and affect their biological activities.

In this chapter, we first examined the expression of E-Cadherin and Snail 1, and then the migrative and metastatic potential of CHO-K1 cells that overexpressing TIFm71 was exaimined. Finally, we evaluated the impact of TIFm71 on cell cycle regulation by DNA flow cytometric analysis.

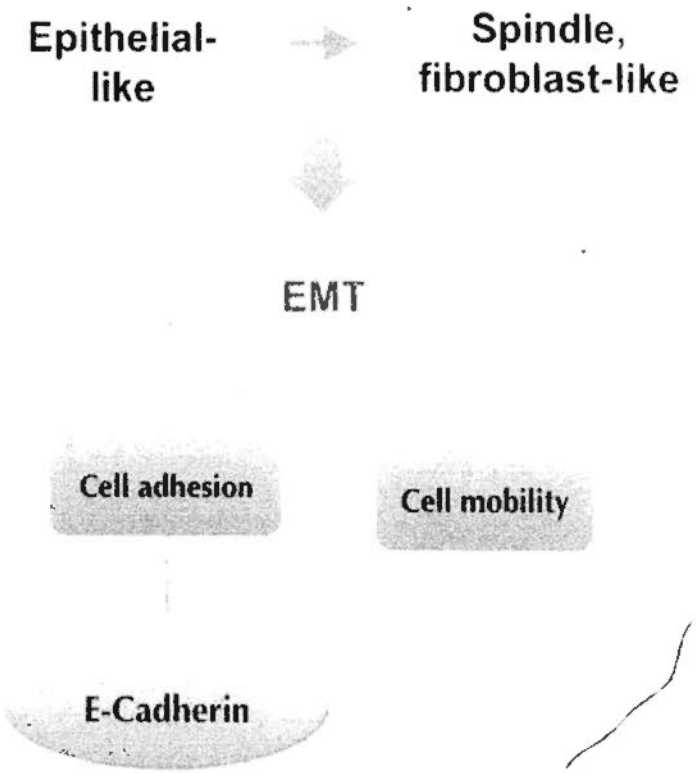


Figure 5.1 Epithelial-to-mesenchymal transition (EMT). EMT is a biologic process of cells characterized by loss of cell adhesion, repression of E-cadherin expression, and increased cell mobility.

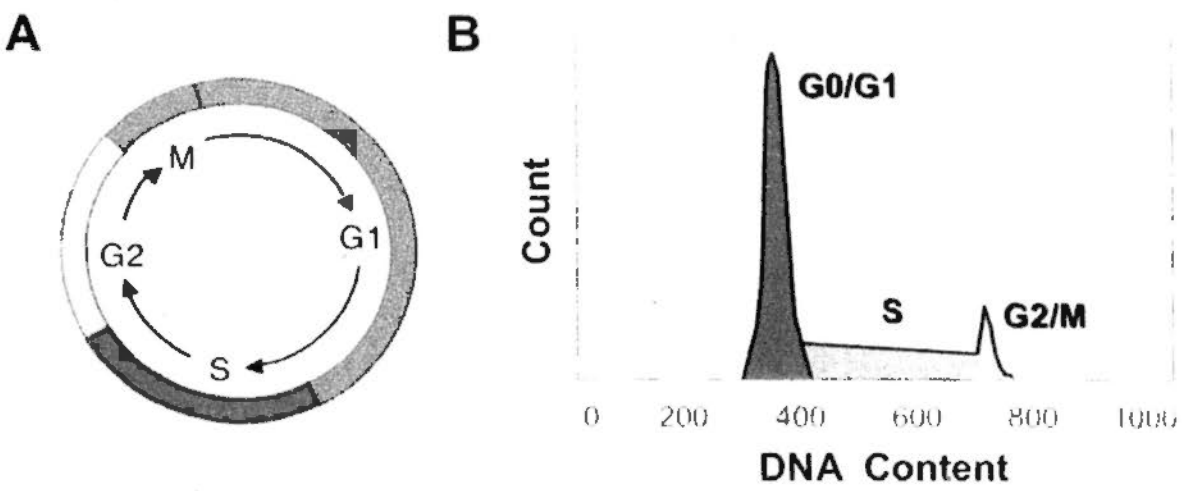


Figure 5.2 Simple representations of the cell cycle. (A) A typical (somatic) cell cycle, which can be divided in four sequential phases: G1, S, G2 and M. (B) Schematic diagram showing cell cycle distribution by DNA flow cytometric analysis.

5.2 Materials

5.2.1 Chemicals

F12-Nutrient powder (#21700-075), fetal bovine serum (FBS, #10270-106), penicillin/streptomycin (#15140-122), trypsin and DAPI (#D1306) were purchased from Invitrogen (Carlsbad, CA, USA). G418 were purchased from Merck Biosciences (Whitehouse Station, NJ, USA). Protease Inhibitor Cocktail Tablets (#04 693 116 001) and Hygromycin B (#843555) were purchased from Roche (Roche Diagnostics Corporation, IN, USA). BD Matrigel™ Matrix Growth Factor Reduced (#356230) was purchased from BD Bioscience (NJ, USA). Crystal violet (#C3886-25G) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Polycarbonate Transwell (24-well, #3422) was purchased from Corning (NY 14831, USA). miScript miRNA Mimics mmu-miR-1195 (#MSY0005856) and AllStars Negative Control siRNA (#1027280) were synthesized by Qiagen (Hilden, Germany). HiPerFect Transfection Reagent (#301704) was purchased from Qiagen (Valencia, CA, USA). All other chemicals and reagents of molecular biology grade were purchased from Sigma-Aldrich (St. Louis, MO, USA).

5.2.2 Antibodies

Mouse anti-E-Cadherin antibody (#610404) was purchased from BD Biosciences (Franklin Lakes, NJ, USA). Goat anti-Snail1 polyclone antibody (#17732) was purchased from Abcam (San Francisco, CA, USA), and mouse anti-β-Actin antibody (#1-800-244-1173) was purchased from Sigma (St. Louis, MO, USA).

5.2.3 Kits and instruments

Amersham ECL™ Western Blotting Detection reagent was purchased from GE Healthcare (UK). BCA™ Protein Assay Kit (#23225) was purchased from Thermo SCIENTIFIC Inc (Rockford, IL 61101, USA).

5.3 Methods

5.3.1 Western blot

Cells seeded in 60 mm dishes were cultured until confluence and lysed by RIPA buffer (50mM Tris-Cl pH 7.4, 150mM NaCl, 1% NP40, 0.25% Na-deoxycholate) containing protease inhibitor Cocktail (Roche). Protein concentration was measured with the bicinchoninic Acid (BCA) protein assay kit (Pierce) as described in section 4.3.3.2. Protein samples (20 µl/well) were typically analyzed on 12% SDS-PAGE gel (8 cm x 9 cm) in running buffer (25 mM Tris pH 8.3, 192 mM glycine and 0.1% SDS) at 120 V constant voltage for ~3 h, followed by transferring to nitrocellulose membrane in transfer buffer (25 mM Tris pH 8.3, 192 mM glycine and 20% methanol) using the Mini-PROTEAN® II Cell (Bio-Rad, USA) at a constant voltage of 70 V for 2 h. Non-specific protein binding sites were blocked by incubating the nitrocellulose membrane with immuno-blotting buffer (5% skimmed milk, 0.05% Tween-20, 150 mM NaCl and 10 mM Tris-HCl, pH 7.5) for 1 hour at room temperature. Primary antibodies used in this study were mouse anti-E-Cadherin antibody (BD Biosciences), goat anti-Snail1 polyclone antibody (Abcam) and mouse anti-β-Actin antibody (Sigma). The nitrocellulose membrane was incubated with primary antibodies in the immno-blotting buffer with gentle shaking at 4°C overnight. Unbound primary antibodies were removed by washing with TBST (50 mM Tris, 150 mM NaCl, 0.05% Tween-20, pH 7.6) for 10 min four times. Then the protein blot was probed with 1:2000 (v/v) HRP-conjugated anti-mouse/anti-goat secondary antibody at room temperature for 2 h with gentle shaking. Unbound secondary antibodies were removed by washing with TBST for 10 min four times. Immunoreative bands were visualized by incubating with the Amersham ECL™ Western Blotting Detection reagent (GE Healthcare, UK) and chemoluminescent signals were detected with FUJI

medical X-ray film. Data was normalized against the corresponding controls. The immunoreative bands were digitalized using a flatbed scanner (EPSON GT9500) at 300 dpi and were quantified using UN-SCAN-IT software (Silk Scientific). Differences between normalized data were analyzed with GraphPad Prism 5 Software.

5.3.2 Confocal fluorescence microscopy

Coverslips were sterilized and pretreated with 0.01% poly-D-Lysine (w/v, in autoclaved water) for 40 min at room temperature. After washing with PBS for three times, the coverslips were put into 35 mm culture dishes. Cells (2×10^5) were seeded onto the coated coverslips and cultured in a 37°C humidified incubator for 24-48 h. The cells on coverslips were washed once with PBS and fixed by adding 1 ml of 4% paraformaldehyde (w/v, in PBS) for 15 min at room temperature or with 1 ml of methanol for 10 min at -20°C. The cells were then washed with Tris buffer (50 mM Tris, pH 7.4) for 5 min three times. Nonspecific binding was blocked by incubating with 1% BSA (Sigma) in PBS containing 0.1% Triton X-100 at room temperature for 30 min. After blocking, cells were washed with Tris/Triton buffer (50 mM Tris, 0.1% Triton X-100, pH 7.4) for 5 min twice, followed by incubating with 1:200 diluted anti-E-Cadherin primary antibody (diluted with Tris/Triton buffer containing 10% goat serum) in a humidified air-tight box at 4°C overnight. After that, cells were washed with Tris/Triton buffer for 5 min three times and incubated with FITC conjugated goat-anti-mouse secondary antibody in a humidified air-tight box at room temperature for 1 h, followed by washing with Tris/Triton buffer for 5 min three times. Nuclei of the cells were stained with 300 nM of DAPI (#D1306, Invitrogen) at room temperature for 5 min followed by washing with Tris/Triton buffer for 5 min

three times. Cells on the coverslips were then air-dried, mounted with anti-fade mounting medium (90% glycerol, 2.3% 1, 4-diazabicyclo- [2.2.2]-octane, 0.02% sodium azide, 0.1 M Tris-HCl, pH 8.0) and sealed with nail polish.

Cells on coverslips prepared from immunofluorescence described above were examined with a Leica TCS SP5 confocal microscope. FITC was excited with wavelength at 488 nm by an Ar/Kr laser and detected at 525±50 nm. Fluorescent and transmission images were captured in separate channels and merged to create overlay images.

5.3.3 Wound-healing assay

CHO-K1 wild type cells, as well as CHO-K1-M229 and CHO-K1-S229 stable cells were cultured to near confluence (~90%) in 35 mm dish with F12 medium supplemented with 10% FBS and 500 μg/ml hygromycin B. A line was drawn on the bottom of the dish with a marker pen, followed by scratching three separate wounds through the monolayer cells moving perpendicular to the line using a 200 μl pipet tip. Suspended cells were removed by replacing the old medium with fresh one. Pictures were taken 0 h, 24 h and 48 h after scratching the wound, and cells that migrated into the scratched region were counted.

5.3.4 Matrigel invasion assay

BD Matrigel™ Matrix Growth Factor Reduced (#356230, BD) was thaw on ice for overnight and diluted to 1 mg/ml with ice-cold serum free F12 medium, 100 µg of the reconstituted Matrilgel was coated to the 8 µm, 24-well polycarbonate Transwell (#3422, Corning) inserts using a pre-cooled pipet tip. The Transwell inserts were kept in 37°C cell culture incubator overnight. The next day, CHO-K1, CHO-K1-M229 and CHO-K1-S229 stable cells were trypsinized and counted. Cells (50,000) of each type were resuspended in 100 µl of F12 medium supplemented with 1% FBS, and seeded in Transwell inserts. The lower compartment of Transwell was filled with 600 µl of F12 medium supplemented with 10% FBS. Cells were allowed to migrate across the Matrigel toward complete growth medium for 24 h, and non-invading cells were removed mechanically using cotton swabs. Cells were fixed with 4% freshly prepared paraformaldehyde for 5 min, and stained with 0.05% crystal violet for 30 min, followed by washing with tap water twice.

5.3.5 M229 knockdown assay

Shortly before transfection, 1.5-6 x 10^5 cells were seeded in 35 mm dish containing 2.3 ml of complete culture medium. Mmu-miR-1195 mimic single-stranded RNA oligo (Fig. 3.3E) or AllStars Negative Control siRNA (150 ng) was diluted in 100 µl of culture medium without serum to give a final RNA concentration of 5 nM in cells, and 12 µl of HiPerFect Transfection Reagent (#301704) was added to the diluted RNA oligos. The mixture was mixed by vortexing and incubated at room temperature for 5-10 min to allow the formation of transfection complexes. The mixture was then added onto the cells drop-wise, and the dish was gently swirled to ensure uniform distribution of the transfection complexes. Cells were incubated at 37°C in a 5% CO_2

humidified incubator for ~24 h. The sequence mmu-miR-1195 mimic for M229 knockdown was 5'- UGA GUU CGA GGC CAG CCU GCU CA -3'. The knockdown efficiency was determined by RT-PCR as described in section 2.3.2.4. PCR reaction was run for 23 cycles with annealing temperature at 60°C. The forward (F) and reverse (R) primers and length of amplicons (in parentheses) obtained were: M229 F: 5'-TGT CTT AGA ACT GAG AAG TT-3', R: 5'-TTC CAA ATT TTT TGT AAA AA-3' (229 bp); GAPDH F: 5'-CAA GGC TGA GAA TGG AAA GC-3', R: 5'-GGG GTT ATT GGA CAG GGA CT-3' (916 bp).

5.3.6 Cell cycle analysis

Wild type CHO-K1 cells, as well as cell clones with stable integration of M229 and S229 transgenes, were seeded in 60 mm dishes and cultured in a humidified 5% CO_2 incubator till ~90% confluence. Cells were typsinized, washed with PBS once and re-suspended in 300 µl of ice-cold PBS. To fix cells, 700 µl of ice-cold absolute ethanol (pre-cooled in -20°C) was added into the cells dropwise and mixed thoroughly. The cells were fixed in 4°C overnight. Prior to flow cytometry analysis, cells were spun down, washed once with ice-cold PBS and incubated with 1 ml of PBS containing 8 µg/ml of RNase A and 40 µg/ml of propidium iodide (#P4170, Sigma) for 30 min at 37°C in dark. FAC/SSC and PE-Texas Red-A/PE-Texas Red-W were recorded in liner scale. Data were analyzed with Graph Prism 5 software (GraphPad, Inc, CA, USA).

5.3.7 Statistical analysis

All data were represented as mean ± SEM from three or more experiments. Statistical difference between two groups was detected by t-test, while statistical difference between three or more groups was analyzed by One-way ANOVA using GraphPad Prism 5 Software.

5.4 Results

5.4.1 Down-regulation of E-Cadherin by TIFm71

To evaluate the potential functions of random insertion of the ncRNA TIFm71, cell clones with stable integration of M229 and S229 transgenes were constructed. The expression of integrated transgenes was confirmed by RT-PCR. As discussed in Chapter 3, it is of interest to note that a striking change in cellular morphology in cells over-expressing TIFm71, whereby the cobblestone-like appearance of epithelial cells switched to a fibroblast-, spindle-like morphology with extensive scattering of cells (Fig. 3.8). These are hallmarks of epithelial-to-mesenchymal transitions (EMT), a process in which cells lose epithelial characteristics to adopt a mesenchymal phenotype.

To determine if the molecular alterations typical of an EMT occurred in TIFm71 stable cells, we examined the expression level of E-Cadherin in the stable cell lines. E-Cadherin is a membrane protein characteristic of cell-cell tight contacts and currently thought to be a suppressor of invasion during carcinoma progression (Birchmeier and Behrens, 1994; Miettinen *et al.*, 1994). Loss of functional E-Cadherin expression appears to be a pivotal step in EMT (Nawijn *et al.*, 2011). Western blot analysis revealed that the expression level of E-Cadherin was significantly down-regulated in TIFm71 stable cells compared to wild type CHO-K1 cells or CHO-K1 cells with stable integration of S229 transgene (Fig. 5.3A). The immunoreative bands were digitalized and normalized against β-Actin. Quantitation of three independent Western blot results showed that the E-Cadherin expression decreased up to 60% by the over-expression of TIFm71 (Fig. 5.3B).

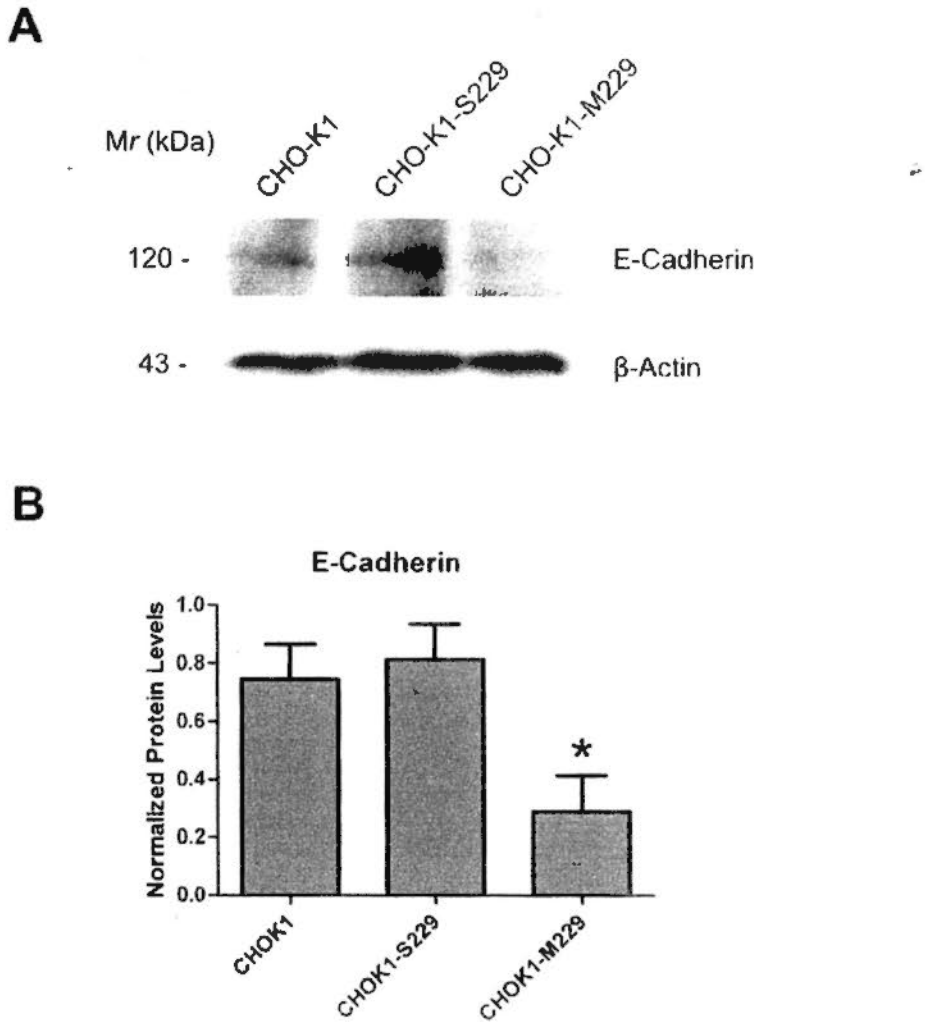
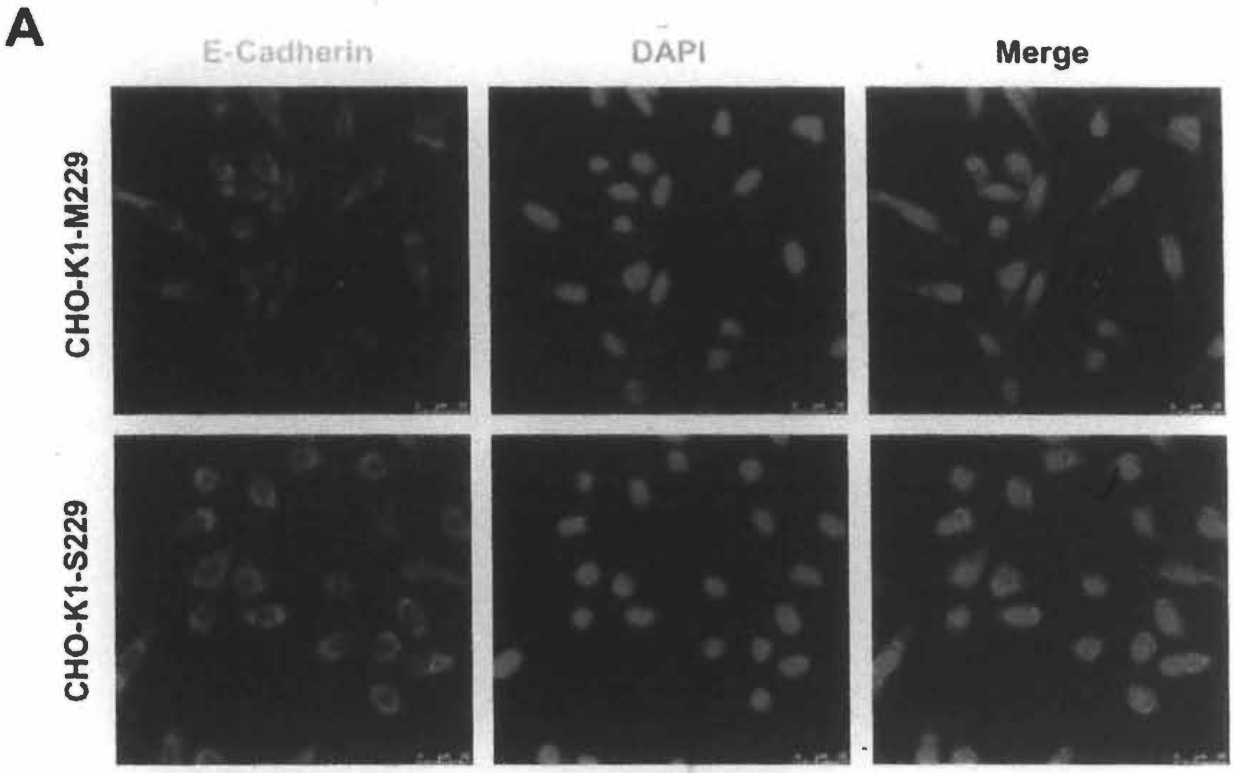
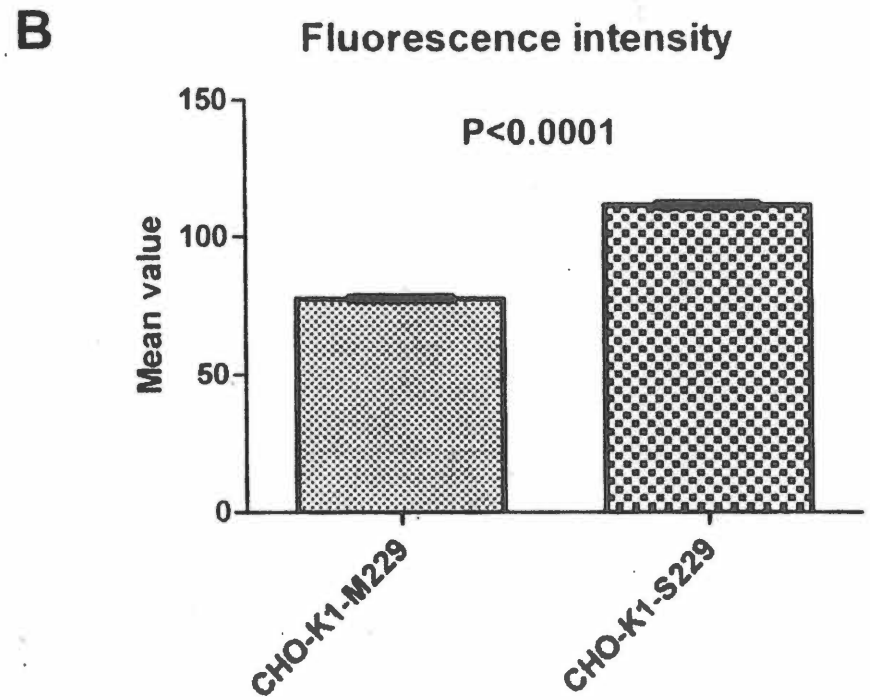


Figure 5.3 Expression of E-Cadherin analyzed by Western blot. (A) Cells were seeded in 60 mm dishes, cultured until confluence and lysed by RIPA buffer containing protease inhibitor Cocktail. The whole cell lysate protein samples were separated in a 12% SDS-PAGE and detected with anti-E-Cadherin and anti-β-Actin antibodies. Data shown was representative of 3 separate experiments with similar results. (B) Intensity of E-Cadherin immunoreative bands was normalized against the corresponding β-Actin. Differences among cell lines were expressed as mean ± SEM of 3 separate experiments and analyzed with one-way ANOVA by GraphPad Prism 5 Software (*$P < 0.05$).

To further confirm the Western blot result, the E-Cadherin protein level was determined with immunofluorescence assay as described in the Methods section. CHO-K1-M229 and CHO-K1-S229 cells were immunostained with anti-E-Cadherin antibody, and FITC-conjugated secondary antibody was used to visualize the primary antibody staining. Nuclei of the cells were stained with of DAPI (Invitrogen). Consistent with the Western blot results showed in Fig. 5.3, there was a significant decrease of FITC-fluorescence intensity in CHO-K1-M229 cells in comparison to that of CHO-K1-S229 cells (Fig. 5.4A). The fluorescence intensity of cells shown in Fig. 5.4A was quantified with LAS AF Lite software version 2.1.1 (Leica Microsystems) and analyzed by GraphPad Prism 5 (GraphPad Software). The result showed the E-Cadherin level decreased by 30-40% in CHO-K1-M229 cells compared to that of the CHO-K1-S229 cells (Fig. 5.4B).

Figure 5.4 Expression of E-Cadherin analyzed by immunofluorescence. (A) Representative immunostainning of E-Cadherin in CHO-K1-M229 cells (upper panel) and CHO-K1-S229 cells (lower panel) detected by confocal microscopy. Cells were grown on 25 mm^2 coverslips in 35 mm dishes. Immunofluorescent confocal microscopy was performed as described in the Methods section. Fluorescence images of FITC and DAPI were captured separately and merged together to form overlay images. Scale bar was 25 μM. The experiment was repeated for three times independently. (B) Cells from separate experiments were randomly selected, and FITC-fluorescence intensity of each cell was quantified with LAS AF Lite software version 2.1.1 (Leica Microsystems). Differences between CHO-K1-M229 and CHO-K1-S229 were expressed as mean ± SEM and analyzed with t-test by GraphPad Prism 5 Software.


A
E-Cadherin
DAPI
Merge
CHO-K1-M229
CHO-K1-S229


B
Fluorescence intensity
P<0.0001
Mean value
150
100
50
0
CHO-K1-M229
CHO-K1-S229

5.4.2 Up-regulation of Snail1 by TIFm71

The central role of Snail in the regulation of EMT has been underscored by recent studies (Peinado *et al.*, 2007). Snail plays a fundamental role in EMT by suppressing E-Cadherin expression in mammalian cells, and could be regarded as early markers of tumor malignancy (Barrallo-Gimeno and Nieto, 2005). It induces the expression of genes associated with a mesenchymal and invasive phenotype and also down-regulates the expression of other epithelial molecules such as Muc1, Claudins and Occludins (Barrallo-Gimeno and Nieto, 2005; Wu and Zhou, 2010).

To further determine whether the over-expression of TIFm71 could promote EMT, we examined the expression and Snail1 in CHO-K1-M229 cells by Western blot analysis. Our results showed a significant up-regulation of Snail1 in the cells over-expressing TIFm71 (Fig. 5.5A). Quantitation of three independent Western blot results showed that the Snail1 expression was increased by ~65% in CHO-K1-M229 cells compared to that in CHO-K1-S229 cells (Fig. 5.5B).

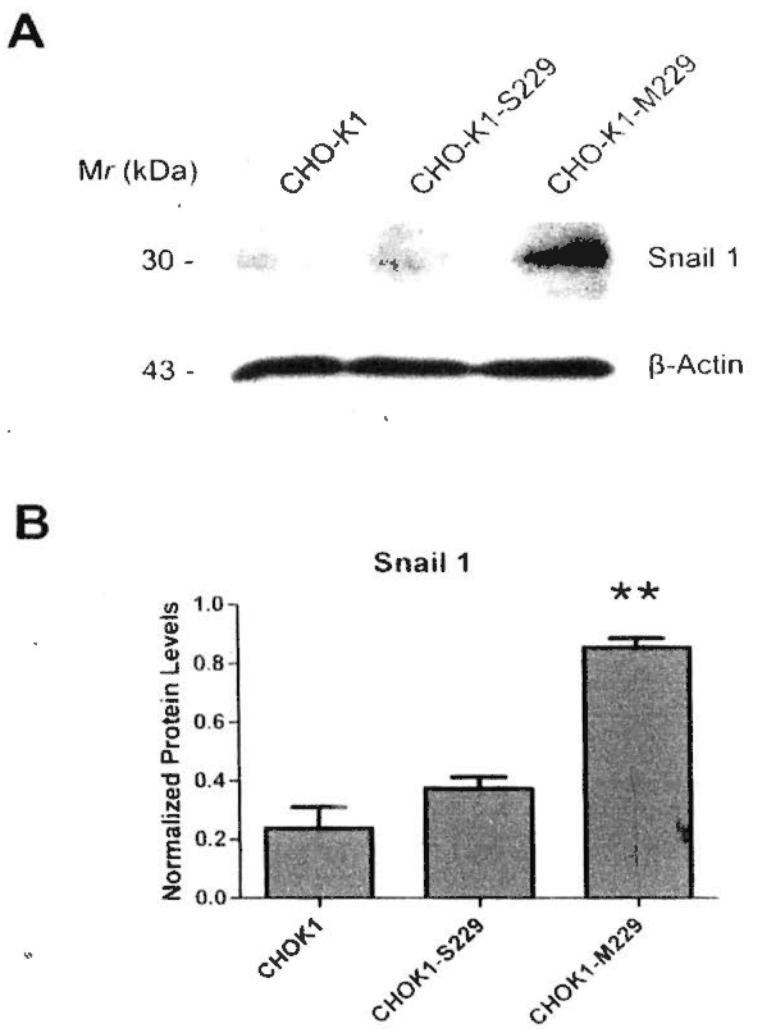
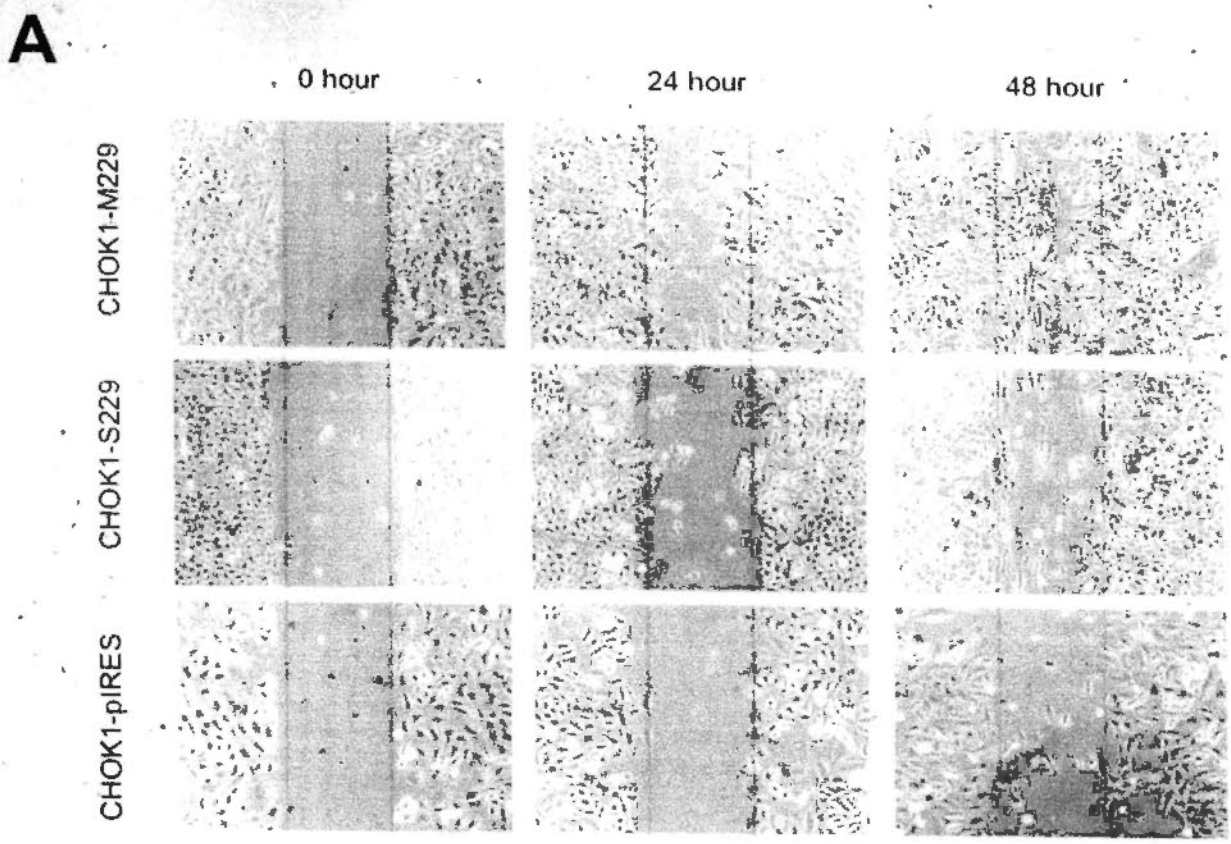


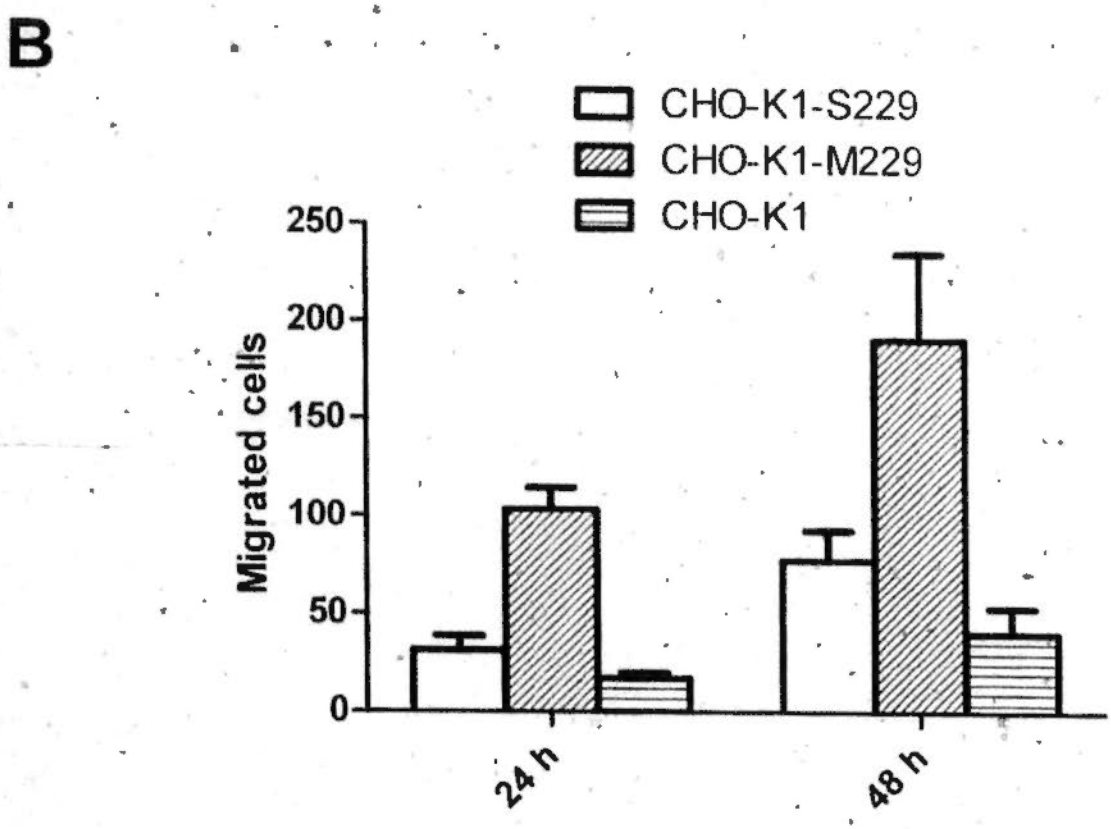
Figure 5.5 Expression of Snail1 analyzed by Western blot. (A) Cells were seeded in 60 mm dishes, cultured until confluence and lysed by RIPA buffer containing protease inhibitor Cocktail (Roche). The whole cell lysate protein samples were separated in a 12% SDS-PAGE and detected with anti-Snail1 and anti-β-Actin antibodies. Data shown was representative of at least 3 separate experiments with similar results. (B) The intensities of Snail1 immunoreative bands were normalized against the corresponding β-Actin. Differences among cell lines were expressed as mean ± SEM of 3 separate experiments and analyzed with one-way ANOVA by GraphPad Prism 5 Software (**$P < 0.01$).

5.4.3 Cell migration analysis by wound-healing assay

Next, we tested whether TIFm71 over-expressing CHO-K1 cells could increase cell migration through decreasing the expression of E-Cadherin. Migration of confluent monolayers of CHO-K1 wild type cells, as well as CHO-K1-M229 and CHO-K1-S229 stable cells were analyzed by wound-healing assay. The wound-healing assay is simple, inexpensive, and one of the earliest developed methods to study directional cell migration *in vitro* (Rodriguez *et al.*, 2005). This assay mimics cell migration during wound healing *in vivo* to estimate the migration and proliferation rate of different cells and culture conditions. The basic steps involve growing a confluent cell monolayer, creating a "wound" in the cell monolayer, capturing images at the beginning and at regular intervals during cell migration to close the wound, and comparing the images to quantify the migration rate of the cells. As showed in Fig. 5.6, raising the expression of TIFm71 in CHO-K1 cells increased cellular migratory rate by -2 to -3 folds, whereas over-expression of control S229 has no significant effect.

Figure 5.6 Wound-healing assay. (A) Cells were cultured to near confluence (~90%) in 35 mm dish with F12 medium supplemented with 10% FBS. A line was drawn on the bottom of the dish, followed by scratching three separate wounds through the cells moving perpendicular to the line with a 200 µl pipet tip. Suspended cells were removed by replacing the old medium with fresh one. Images were taken 0 h, 24 h and 48 h after scratch wounding. The data shown was representative of at least 3 independent experiments with similar results. (B) Data shown was mean ± SEM of 3 separate experiments analyzed with GraphPad Prism 5 Software.


A
0 hour
24 hour
48 hour
CHOK1-M229
CHOK1-S229
CHOK1-pIRES

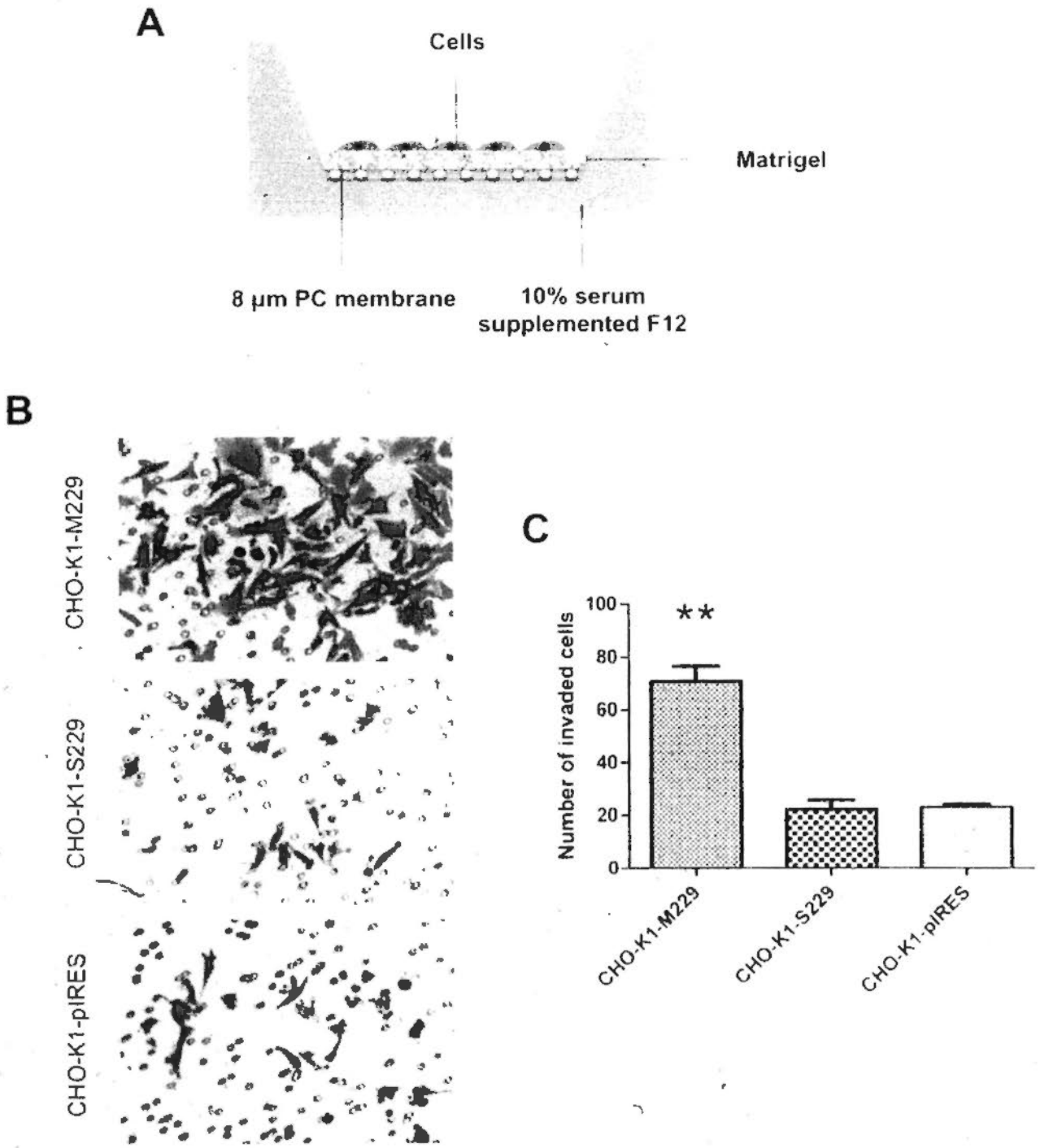

B
CHO-K1-S229
CHO-K1-M229
CHO-K1
Migrated cells
250
200
150
100
50
0
24 h
48 h

5.4.4 Cell invasive capacity analysis by Matrigel invasion assay

BD Matrigel™ Matrix Growth Factor Reduced (#356230, BD) is a solubilized basement membrane preparation extracted from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma, a tumor rich in extracellular matrix proteins. Its major component is laminin, followed by collagen IV, heparan sulfate proteoglycans, entactin/nidogen (Kleinman *et al.*, 1986; Kleinman *et al.*, 1982). Invasion of tumor cells into Matrigel has been used to characterize involvement of ECM receptors and matrix degrading enzymes, which play important roles in tumor progression.

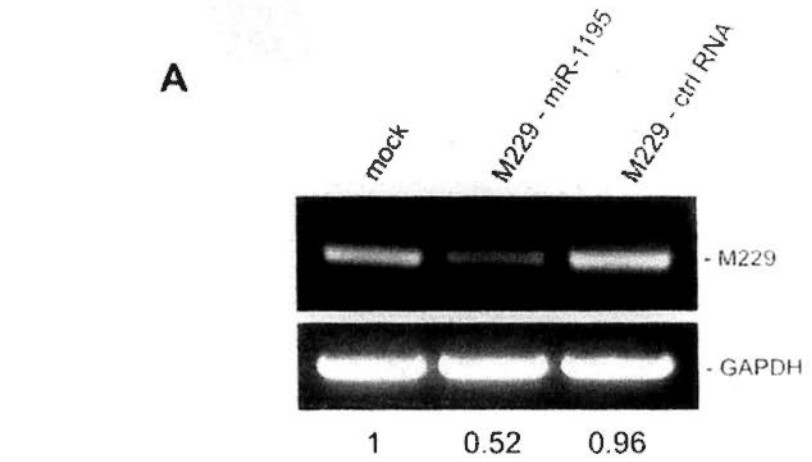
To further analyze metastatic potential, we measured the infiltration of these cells through Matrigel in an 8 μm, 24-well polycarbonate Transwells chamber using 10% of FBS as the chemoattractant (Fig. 5.7A). Cells were allowed to migrate toward complete growth medium for 24 h, and non-invading cells were removed. Migrated cells were fixed with 4% freshly prepared paraformaldehyde, followed by staining with 0.05% crystal violet. The result showed that TIFm71 promoted 3.5 folds of the invasive ability of CHO-K1 cells compare to control cells (Fig. 5.7B, C).

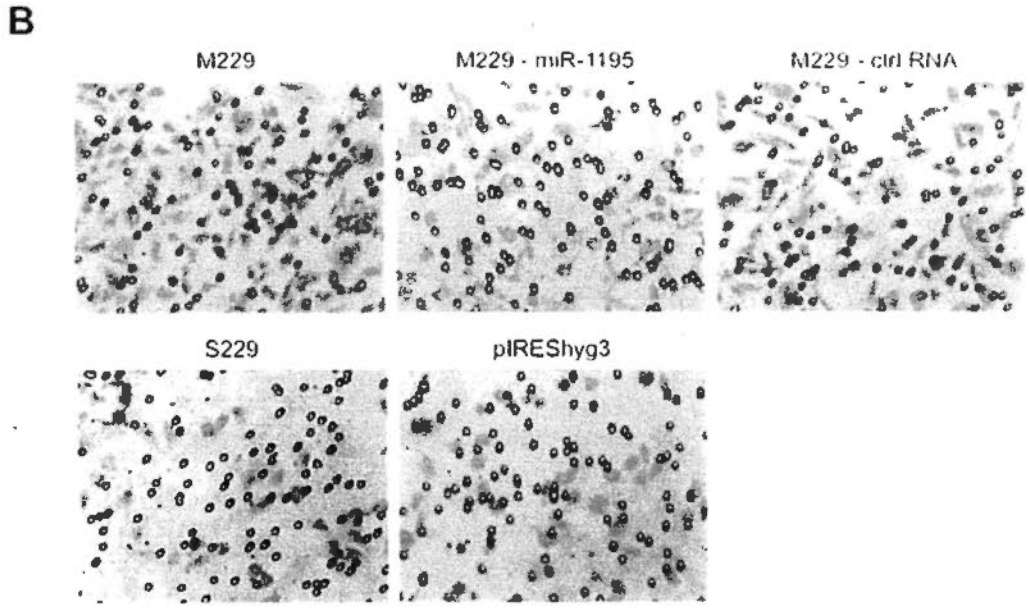
THIS IS A BLANK PAGE

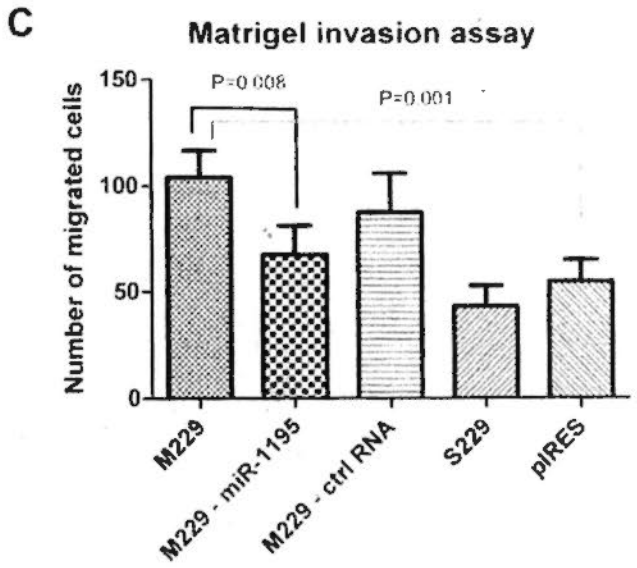

A
Cells
Matrigel
8 μm PC membrane
10% serum supplemented F12
B
CHO-K1-M229
CHO-K1-S229
CHO-K1-pIRES
C
Number of invaded cells
100
80
60
40
20
0
**
CHO-K1-M229
CHO-K1-S229
CHO-K1-pIRES

In light of the preceding data, we sought to test whether the invasive capacity of CHO-K1-M229 cells would be recovered after knocking down of M229. As shown in Fig. 3.3E, TIFm71 could be targeted by a mouse miRNA mmu-miR-1195 in the 3'-stem. Thus a single stranded mmu-miR-1195 miRNA mimic oligo, as well as AllStars Negative Control siRNA, was synthesized. Following transfection with mmu-miR-1195 mimic, the expression M229 decrease by ~50% determined by RT-PCR (Fig. 5.8A). Then the cells together with CHO-K1-S229 and CHO-K1-pIREShyg3 (empty vector) were subjected to Matrigel invasion assay. The result showed that number of CHO-K1-M229 cells infiltrated the Matrigel and Transwell chamber was significantly reduced after M229 knockdown (Fig. 5.8B. C). These data suggested that the gain of high invasive capacity was a result of TIFm71 over-expression.

THIS IS A BLANK PAGE


A
mock
M229 - miR-1195
M229 - ctrl RNA
- M229
- GAPDH
1
0.52
0.96


B
M229
M229 - miR-1195
M229 - ctrl RNA
S229
pIREShyg3


C
Matrigel invasion assay
Number of migrated cells
150
100
50
0
P=0.008
P=0.001
M229
M229 - miR-1195
M229 - ctrl RNA
S229
pIRES

5.4.5 Cell cycle analysis by flow cytometry

As was mentioned in Chapter 4, a number of TIFm71 binding proteins were identified by RNA pull-down followed with proteomics analysis by mass spectrometry. Among the proteins identified, Ras-related GTP-binding nuclear protein (Ran) and M-phase phosphoprotein 4 (MPP4) were reported to be involved in the control of cell cycle progression. Thus we hypothesized that TIFm71 would affect cell cycle distribution through direct interaction with these proteins. For this, we evaluated the impact of TIFm71 on cell cycle regulation by DNA flow cytometric analysis.

Wild type CHO-K1 cells, as well as cell clones with stable integration of M229 and S229 transgenes were fixed, stained with propidium iodide, and subjected to flow cytometry analysis. As showed in Fig. 5.9, the proportion of CHO-K1-M229 cells in G2/M phase was decreased compared to that of the wild type CHO-K1 cells or CHO-K1-S229, whereas the proportion of G0/G1 phase and S phase cells was not significantly affected. These data implicated TIFm71 might block S phase cells from entering G2/M phase or it may shorten the G2/M phase. The underline mechanisms need further investigation.

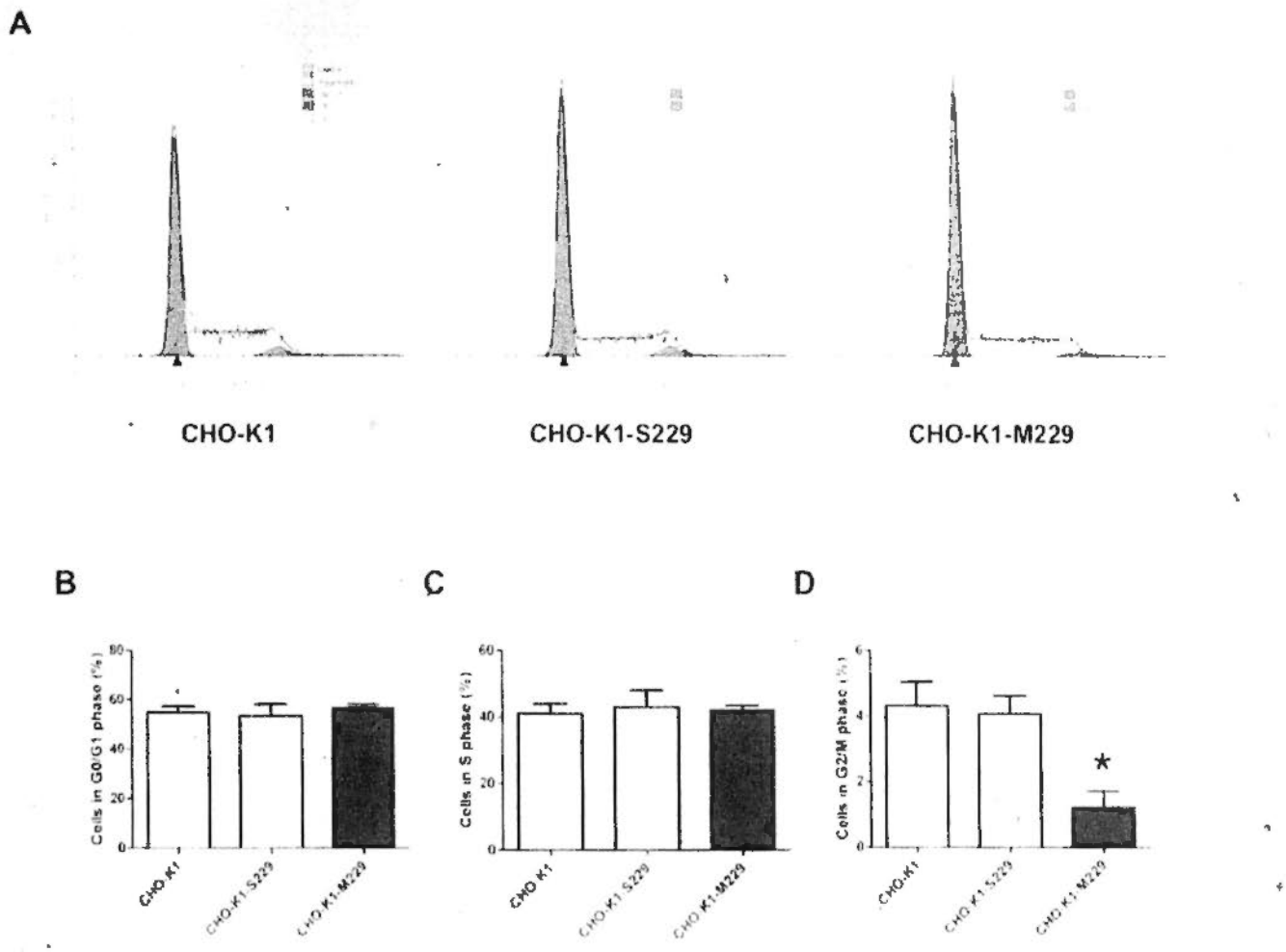


Figure 5.9 Effect of TIFm71 over-expression on cell cycle distribution of CHO-K1 cells. (A) Representative figures showing cell cycle distribution as determined by DNA content. Cells cultured in 60 mm dishes were harvested when reached ~90% confluence, followed by fixing with ice-cold 70% ethanol. After RNase treatment and propidium iodide staining, DNA contents of the cell samples were analyzed by DNA flow cytometric analysis. Data in (B-D) showed mean ± SEM of the proportion of cells in G0/G1 phase (B), S phase (C) and G2/M phase (D) from 3 separate experiments. Differences of means were analyzed with one-way ANOVA by GraphPad Prism 5 Software (*$P < 0.05$).

5.5 Discussion

In previous chapters, we proved that TIFm71 was a small ncRNA originating from exonic 3'-UTR of TIF transcript. Endogenous expression of TIFm71 was detected in various normal tissues, and its expression level was greatly up-regulated in solid tumors. These observations indicated that TIFm71 may not only represent a simple degradation intermediate, but is probably involved in cancer progression. In order to understand the potential biological roles, we subcloned TIFm71 and TIFm71 related sequences and thereafter established a serial of cell clones with a stable expression of TIFm71.

A significant finding here was indeed the association of the TIFm71 redundant with EMT, which has emerged as one of the hottest spots in clinical research during the past few years (Gomes *et al.*, 2011). The EMT program is used to assist epithelial plasticity in normal tissues (Polyak and Weinberg, 2009). Also, it is associated with cancer progression, metastasis, as well as certain cases of drug resistance and relapses after treatment. The activation of an EMT program is critical for the acquisition of malignant phenotypes by epithelial cancer cells (Thiery, 2002). Here, we showed that high level of TIFm71 induced epithelial CHO-K1 cells toward a less-differentiated mesenchymal fate to foster metastasis, evidenced by a decrease in the expression of cell adherent marker E-Cadherin, an increase in the expression of Snail1, and a morphologic change from cobblestone-like appearance of epithelial cells to a spindle-, fibroblast-like morphology.

There are two other ncRNAs reported previously that could promote cancer cell metastatic progression, HOTAIR and MALAT-1. HOTAIR is a large intervening non-coding RNAs (lincRNAs). It recruits the PRC2 complex to specific target genes genome-wide, leading to H3K27 trimethylation and epigenetic silencing of metastasis suppressor genes. Within the context of cancer cells, ectopic expression of HOTAIR seems to re-impose the chromatin state, thereby enabling gene expression programs that are conducive to cell motility and matrix invasion (Gupta *et al.*, 2010). MALAT-1, also a long ncRNA, was found to play a pivotal role in colorectal cancer metastasis (Ji *et al.*, 2003; Xu *et al.*, 2011), but the underline mechanism was not known. Based on our current understanding of the molecular mechanisms by which ncRNAs functions, ncRNAs commonly act as adaptors that position a target molecule, which is a nucleic acid or proteins to from a ribonucleoprotein unit for enzyme activity (Huttenhofer and Schattner, 2006; Wilusz *et al.*, 2009).

Different from the above-mentioned two lincRNAs, TIFm71 was small in size (71 nt) and could fold into a canonical stem-loop secondary structure, which usually provides binding sites for RNA-protein interaction (Tian *et al.*, 2004). Generally, ncRNAs elicit their biological responses through interactions with RBPs within ribonucleoprotein particles (RNPs) (Charon *et al.*, 2010), we therefore hypothesized that TIFm71 might act as a scaffold to recruit EMT related proteins and affects their biological activities. Clearly, our focus on EMT does not exclude the possibility that TIFm71 may regulate other biological processes or functions in other cellular contexts.

In our present study, we proved that TIFm71 could promote cell migration *in vitro*. Cell migration is essential for multiple biological processes such as embryonic development, immune response, wound healing and the tissue renewal. Alterations in cellular migration processes are the cause of various pathologies such as chronic inflammatory diseases, mental disorders, vascular diseases and metastasis formation (Ridley *et al.*, 2003; Vicente-Manzanares *et al.*, 2005). Increased ability of tumor cells to migrate is a characteristic property of metastasis formation (Gerlitz and Bustin, 2011). By wound-healing assay, it was shown that raising the expression of TIFm71 in CHO-K1 cells increased the migratory capacity of CHO-K1 cells by 2 to 3 folds.

TIFm71 also possibly plays an important role in cell invasion and metastasis. In the Matrigel invasion assay, localized degradation of the basement membrane requires the expression of proteases that are specific for the resident ECM, and the concomitant expression of protease inhibitors to prevent excess matrix degradation (Hood and Cheresh, 2002). The result showed that TIFm71 greatly promoted the invasive ability of CHO-K1 cells comparing to control cells, and the phenotype was recovered after TIFm71 knockdown.

Cancer progression has been suggested to involve the loss of cell cycle checkpoint controls that regulate the passage through the cell cycle. These checkpoints monitor the integrity of the DNA and make sure the genes are expressed in a coordinated manner (Schlaepfer *et al.*, 1999; Wang *et al.*, 2000b). In the DNA flow cytometric analysis, we observed that TIFm71 over-expressing CHO-K1 cells were found to have a reduction in the G2/M phase cell population. There are two possible

explanations. Firstly, TIFm71 might accelerate mitosis and shorten the G2/M phase; and secondly, TIFm71 possibly block the cells in S phase from entering G2/M phase. In both cases, the ratio of cells in G2/M will be relatively lower than normal cells when examined in a particular time point. Two potential TIFm71 RNA binding proteins, Ran and MPP4, are probably responsible or at least partially responsible for the deregulation of cell cycle distribution. Previous studies implied that these two proteins are involved in the control of cell cycle progression (Abaza *et al.*, 2003; Clarke and Zhang, 2008; Matsumoto-Taniura *et al.*, 1996). However, much more studies will be required to clearly reveal the underlying mechanisms.

Taken together, these results proved that ectopic expression of TIFm71 promoted the non-invasive, epithelial like ovary cells to undergo mesenchymal transition, which was consistent with previous study that TIFm71 expression was greatly up-regulated in xenografts (Lin *et al.*, 2009). It implied that there was a cause and effect relationship between TIFm71 over-expression and tumor formation. In addition, we observed a G2/M phase reduction in TIFm71 stably over-expressing cells.

Chapter 6

Alu-like Sequences in Human Chemokine Genes

6.1 Introduction

As introduced in Chapter 1 and Chapter 2, chemokines are a family of small, structurally related proteins (8-14 kDa) that may participate in immune and inflammatory responses through the chemoattraction and activation of leukocytes (Sharma, 2010). They are subdivided into four groups based on the number and spacing of the cysteine residues conserved in the N-terminal position and are named CXC, CC, CX3C and C, in agreement with the systematic nomenclature (Rostene *et al.*, 2011). To date, more than 50 chemokines have been identified in human. CCL5 is a member of human chemokine belonging to the CC subfamily. Gene structure of CCL5 is similar to TIF in several ways: ~1000 bp in length, with long 3'-UTRs (larger than the ORF regions), consisting of a stem-loop structure near the end of 3'-UTR, and most interestingly, the sequences of the stem-loop region of both genes are related to Alu elements. TIFm71 is complementary to Alu-J (Fig. 3.3F), while the stem-loop of CCL5 shares high similarity with Alu-sb.

Alu elements are the most abundant repetitive elements in the human genome. They emerged 65 million years ago from a 5' to 3' fusion of the 7SL RNA gene and amplified throughout the human genome by retrotransposition to reach more than one million copies (Hasler and Strub, 2006). Fragmented and/or full-length Alu elements have been found in the internal exons of human genome (Moolhuijzen *et al.*, 2010). Although it has once been considered to be nothing more than 'genome parasites', these Alu elements are useful for many purposes in cell metabolism. For instance, an enrichment of biosynthesis genes with Alu-associated binding sites was observed in

developmental transcription factors, which possibly play inhibitory roles, suppressing proliferation during differentiation (Hasler and Strub, 2006).

In the previous chapters, we have shown that TIFm71 was originated from an exonic 3'-UTR of the TIF transcript which is a new pathway for ncRNA generation. One question arises as to whether this is only an exceptional case? Therefore, it is of great interest to find other human genes that potentially generate ncRNA from their untranslated regions.

6.2 Materials

6.2.1 Chemicals

Lipofactamine 2000 (Cat. 11668-019), TRIzol reagent, 1 kb plus DNA ladder and all DNA oligonucleotides were ordered from Invitrogen (Carlsbad, CA, USA). GeneRuler™ Ultra Low Range DNA Ladder (#SM 1211) was purchased from Fermentas (Maryland, USA). ULTRAhyb®-Oligo Hybridization Buffer (#AM8663) was purchased from Ambion (CA, USA). Nytran® SPC membrane (#10416296) was ordered from Whatman (Dassel, Germany). All other chemicals and reagents of molecular biology grade were purchased from Sigma-Aldrich (St. Louis, MO, USA).

6.2.2 Enzymes

T4-Polynucleotide Kinase (T4-PNK, #M0201S) and all restriction enzymes including *Bam*HI and *Xba*I were purchased from New England Biolabs (Ipswich, UK). SuperScript™ II reverse transcriptase (#18064-014) was obtained from Invitrogen (Carlsbad, CA, USA).

6.2.3 Vectors and cells

pcDNA3 cloning and expression vector was kindly provided by Prof. TT, Kwok (CUHK). MCF-7, HepG2, A431, Raji, LNCap-FGC, MDA-MB-231 and HEK293 cells were ordered from ATCC (Manassas, VA, USA). CaCo and Pheonix Ampho cells were kindly provided by Prof. Y, Ke (CUHK).

6.3 Methods

6.3.1 CCL5 expression in human cell lines examined by RT-PCR

RNA samples were isolated from a variety of human cell lines including mammalian adenocarcinoma MCF-7, hepatocarcinoma HepG2, colorectal adenocarcinoma CaCo, embryonic kidney HEK293 and Pheonix Ampho, epidermoid carcinoma A431, lymphoma Raji and prostate carcinoma LNCap-FGC cells. A similar RT-PCR protocol was described in section 2.3.2. Full-length CCL5 was amplified by PCR using oligonucleotides pair 5'-CTC TCC CAC AGG TAC CAT GAA-3' and 5'-ACG GAG TCT CGC TCT GTC G-3'. The PCR reaction was run for 35 cycles at denaturing temperature 94°C for 30 sec, an annealing temperature 55°C for 30 sec and an extension temperature 72°C for 30 sec. At the end of the amplification cycles, an extended incubation at 72°C for 7 min was added.

6.3.2 Subcloning of CCL5-FL and CCL5-SL

Full-length CCL5 (CCL5-FL) and stem-loop region in the 3'-UTR of CCL5 (CCL5-SL) were amplified by PCR and constructed into the multiple cloning sites of pcDNA3 between *Bam*HI and *Xba*I restriction enzyme cutting sites (Fig. 3.1). A detailed expression vector construction protocol was described in section 3.3.2. The CCL5-FL fragment (726 bp) was PCR-amplified from the cDNA of A431 cancer cells using oligonucleotides pair F1538-*Bam*HI-CCL5: 5'-CGG GAT CCC GCT CTC CCA CAG GTA CCA TGA A-3' and R1540-*Xba*I-CCL5: 5'-GCT CTA GAG CAC GGA GTC TCG CTC TGT CG-3' (Fig. 6.2A). PCR reaction was run for 35 cycles with denaturing temperature at 94°C for 30 sec, an annealing temperature at 55°C for

30 sec and an extension temperature at 72°C for 1 min. At the end of the amplification cycles, an extended incubation at 72°C for 7 min was added. CCL5-SL fragment was amplified by PCR from CCL5-FL using oligonucleotides pair F1539-*Bam*HI-CCL5: 5'-CGG GAT CCC GTG ATG AAA CCC CAT CTC TAC TAA A-3' and R1540-*Xba*I-CCL5: 5'-GCT CTA GAG CAC GGA GTC TCG CTC TGT CG-3' (Fig. 6.2A). PCR conditions were the same as that used in CCL5-FL amplification.

6.3.3 CCL5-SL processing analyzed by polyacrylamide gel Northern blot

6.3.3.1 Cell transfection and total RNA isolation

HEK293 cells were transfected with pcDNA3 empty vector, pcDNA3-CCL5-FL or pcDNA3-CCL5-SL constructs using lipofectamine 2000 as described in section 4.3.1.1. In brief, 4 μg of each plasmid and 10 μl of lipofectamine 2000 were incubated separately in 500 μl of serum free DMEM for 5 min. Then the diluted plasmids and lipofectamine 2000 reagent were mixed at room temperature for another 15 min. The mixture together with 4 ml of serum free DMEM were added to the HEK293 cells. After incubation for 5 h, medium was changed to DMEM supplemented with 10% FBS. Total RNA of transiently transfected cells was isolated with TRIzol reagent (Invitrogen) 24 h post transfection as described in section 2.3.2.2.

6.3.3.2 Polyacrylamide gel Northern blot analysis

To detect the CCL5-SL processing, 20 μg of each total RNA sample was resolved on a 12% urea-polyacrylamide gel and transferred to Nytran SPC membrane (Whatman) using a semidry electroblotting apparatus. Membranes were crosslinked and

hybridized overnight at 30°C in ULTRAhyb-Oligo hybridization buffer (Ambion) with [γ-^{32}P] ATP-labeled oligonucleotides complementary to the 3' -stem of CCL5-SL (5'-TCG CCC AGG CTG GAG TGC AGT GGC GCG ATC TCG GCT-3'). The relative position of the probe to the stem-loop of CCL5 was shown in Fig. 6.3A (indicated by a dashed line). The membrane was exposed to Kodak Phosphor Screen SD230.

6.4 Results

6.4.1 TIFm71-like structures in the 3'-UTR of human chemokines

To date, approximately 50 human chemokines have been discovered (Allen *et al.*, 2007). Complete sequence of cDNAs of each chemokine was obtained from the NCBI database. The secondary structure of all 50 human chemokines was analyzed by RNAfold WebServer in Vienna RNA Websuite (Gruber *et al.*, 2008), a comprehensive collection of tools for folding, design and analysis of RNA sequences, aiming to find out whether there are human chemokine genes that have similar structure with TIF in their 3'-UTRs.

The result showed that 24 human chemokine genes have a long 3'-UTR tail that larger than 300 bp. It is of interest to note that 13 chemokines could fold into at least one hairpin secondary structure (50-100 nt) in the 3'-UTR region, including CCL2, CCL3, CCL5, CCL6, CCL7, CCL13, CCL16, CCL22, CCL28, CXCL5, CXCL6, CXCL13 and CXCL16 (Fig. 6.1).

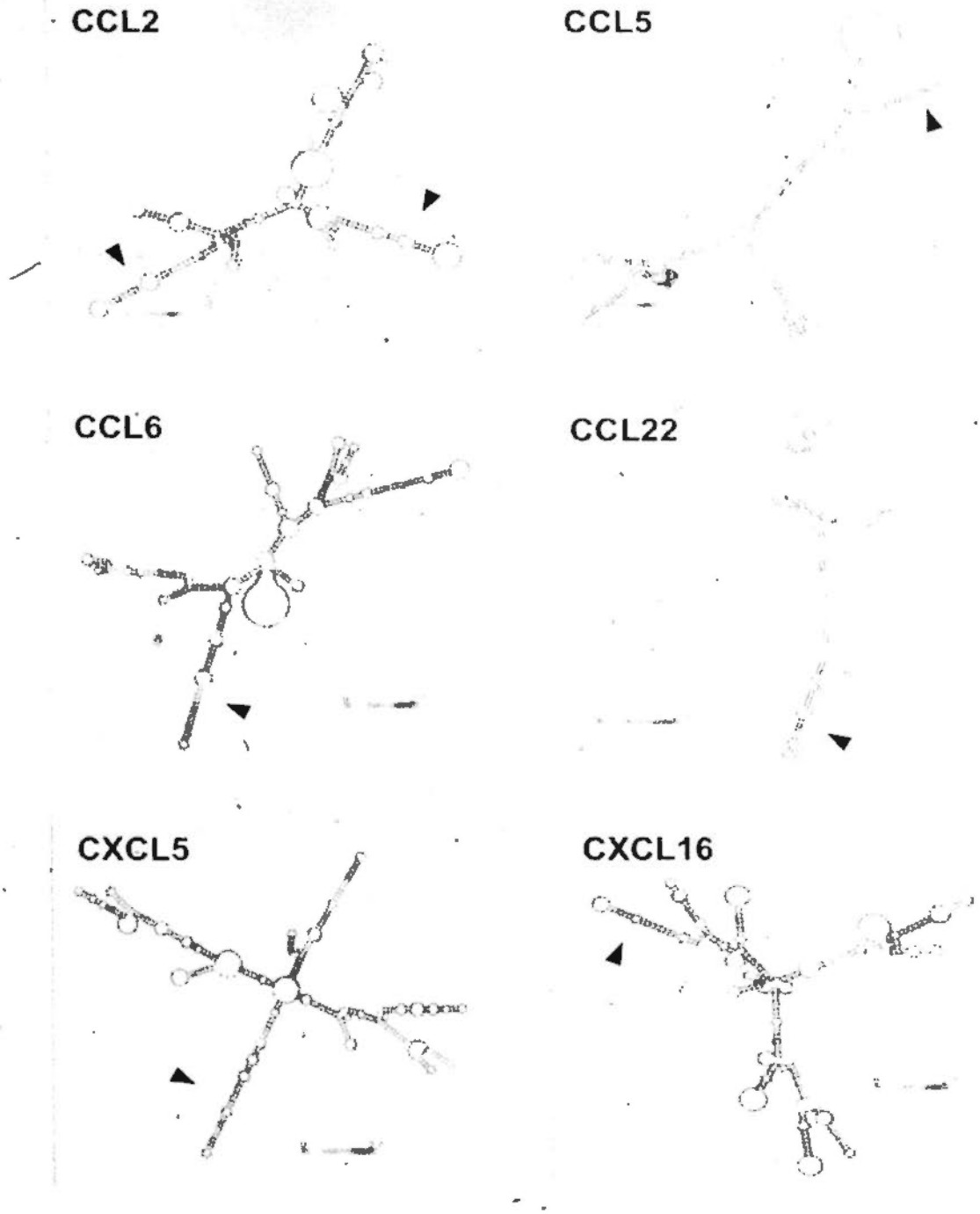


Figure 6.1 Stem-loop structures in the 3'-UTR of human chemokines. RNA secondary structures shown were predicted with RNAfold WebServer. Stem-loop regions were indicated with arrows. Different colors of nucleotides indicate values of positional entropy.

6.4.2 Subcloning of CCL5-FL and CCL5-SL

Among the human chemokines that have one or more stem-loop structure in the 3'-UTR, CCL5 shared a similar gene structure with TIF. CCL5 was 1237 bp in length, with a 56 bp stem-loop structure (CCL5-SL) closed to the 3'-end (Fig. 6.2A). Intriguingly, CCL5-SL shared a high sequence similarity with the consensus part of a human Alu sequence (Fig. 6.2B). However, no ARE or R-SBE-like element was found. Previous studies revealed that CCL5 is one of several CC cytokine genes clustered on the q-arm of chromosome 17, functions as a chemoattractant for blood monocytes, memory T helper cells and eosinophils (Mi *et al.*, 2011).

In order to investigate whether CCL5-SL could be released from the CCL5 transcript, constructs of full-length CCL5 (CCL5-FL) and CCL5-SL were established as described in the Methods section. To obtain the full-length sequence of CCL5 cDNA, RT-PCR was carried out using the cDNA derived from various human cell lines. CCL5 mRNA was found to be expressed in cancer cell lines, and the most abundant expression was detected in human epidermoid carcinoma A431 cells (Fig. 6.2C).

CCL5-FL fragment tailed with *Bam*HI and *Xba*I restriction enzyme cutting sites were PCR-amplified using the cDNA of A431 cell as template and subcloned into the multiple cloning sites of pcNDA3. CCL5-SL fragment was PCR-amplified from CCL5-FL and subcloned using the same methods.

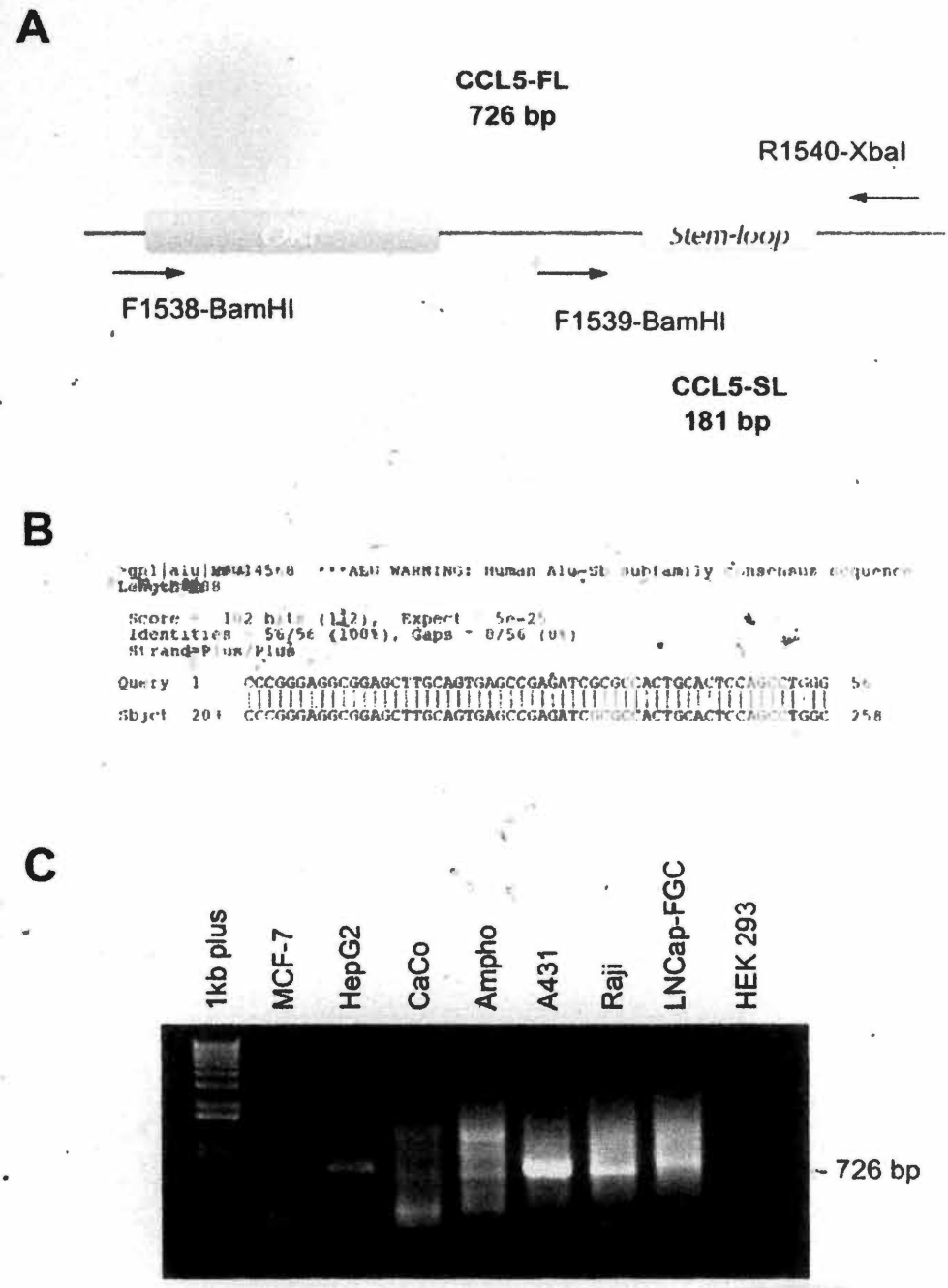


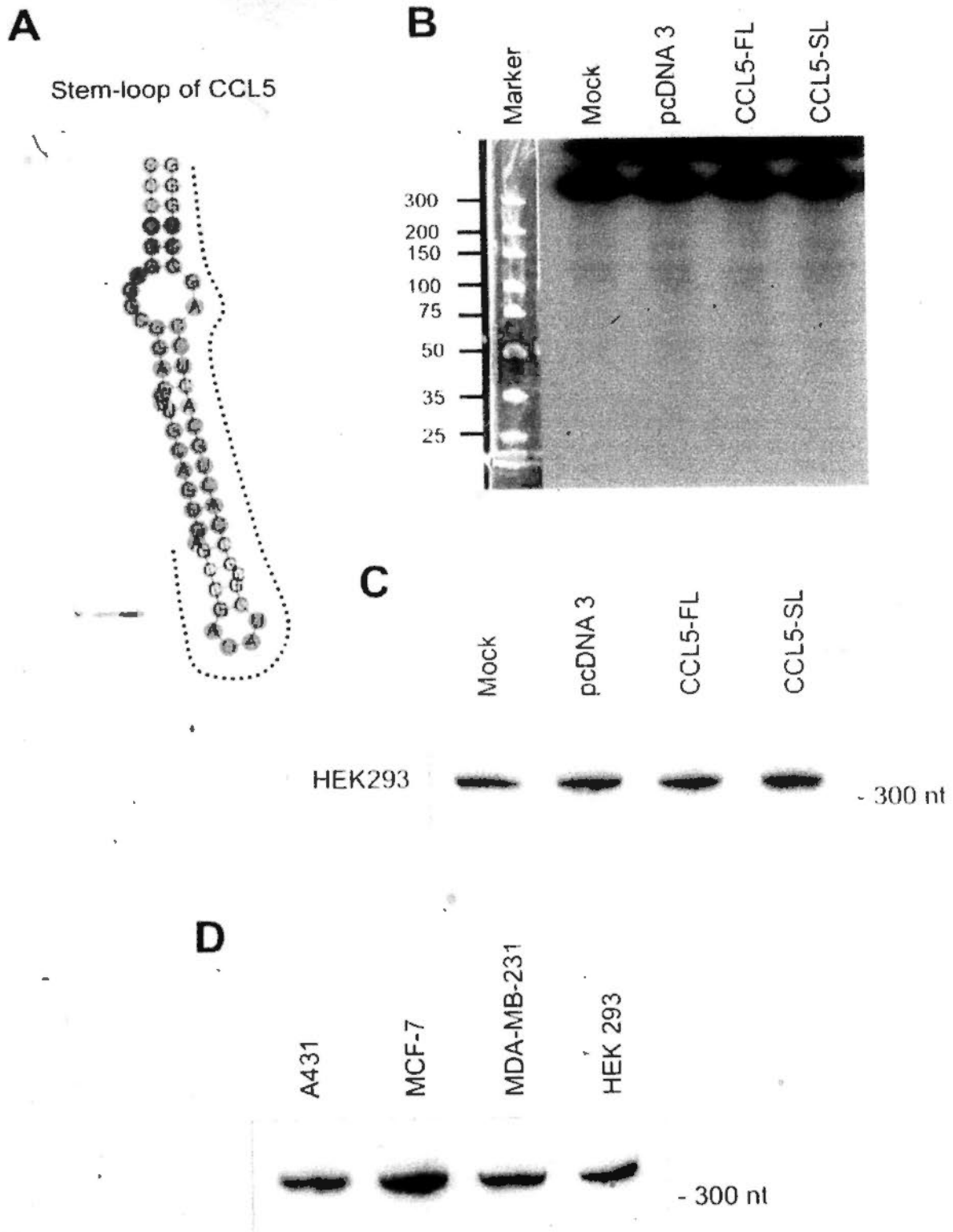
Figure 6.2 Subcloning of CCL5-FL and CCL5-SL. (A) Schematic diagram of CCL5 gene. Dark black arrows showed positions of primers used in the RT-PCR as described in the Methods section. The open reading frame region was labeled with "ORF" and colored in purple; the stem-loop region was labeled with light blue. (B) BLASTEN search of CCL5-SL sequence against Human ALU repeats element (alu-_repeats) database in NCBI. (C) CCL5 expression in human cell lines was examined by RT-PCR.

6.4.3 The 3'-UTR of CCL5 did not generate ncRNA or miRNA

After the establishment of expression vectors of CCL5-FL and CCL5-SL, we investigated whether CCL5-SL would be released as an ncRNA or even a pre-miRNA. Following the transient transfection of HEK293 cells with these expression vectors, total RNA was subsequently extracted for polyacrylamide gel Northern blot analysis using a probe complementary to the CCL5 stem-loop sequence indicated by dashed line in Fig. 6.3A. The result showed that no ncRNA was detected in around or less than 60 nt region, indicating the 3'-UTR of CCL5 could not generate ncRNA or miRNA (Fig. 6.3B).

Indeed, an endogenous large RNA transcript in HEK293 cells (>300 nt) was strongly hybridized with [γ-^{32}P] ATP-labeled oligonucleotides complementary to the 3'-stem of CCL5-SL (Fig. 6.3B-C), regardless of the transfected expression vectors. Also, the RNA transcript was highly conserved as it could be detected in different cell lines besides HEK293, including epidermoid carcinoma A431, mammary adenocarcinoma MCF-7 and MDA-MB-231 cells (Fig. 6.3D). Moreover, BLASTN search showed a high sequence similarity (95%) between CCL5-SL and consensus part of a human Alu sequence. We therefore concluded that the large RNA transcript detected by CCL5-SL probe was likely to be Alu sequence(s).

THIS IS A BLANK PAGE


A
Stem-loop of CCL5
B
Marker
Mock
pcDNA 3
CCL5-FL
CCL5-SL
300
200
150
100
75
50
35
25
C
Mock
pcDNA 3
CCL5-FL
CCL5-SL
HEK293
- 300 nt
D
A431
MCF-7
MDA-MB-231
HEK 293
- 300 nt

6.4.4 Stem-loops in human chemokine genes 3'-UTR and Alu sequences

We next tested whether other stem-loop regions in the 3'-UTR of human chemokines share sequence similarities with Alu sequence. For this, the sequence of stem-loop regions were BLASTN searched against Human ALU repeats element (alu_repeats) database in NCBI. To our surprise, a lot of chemokine genes has an Alu sequence-like 3'-UTR in the stem-loop region, including CCL5, CCL6 (all of the three stem-loops), CCL13, CCL22, CCL28, CXCL5, CXCL13 and CXCL16. Among these chemokines, high similarities were found in the stem-loop regions of CCL5, CCL6 (3rd stem-loop), CCL28 and CXCL16 (Fig. 6.4), implying that Alu-like sequences in 3'-UTR commonly present, and might play important roles in host gene regulation, or function independently as an ncRNA.

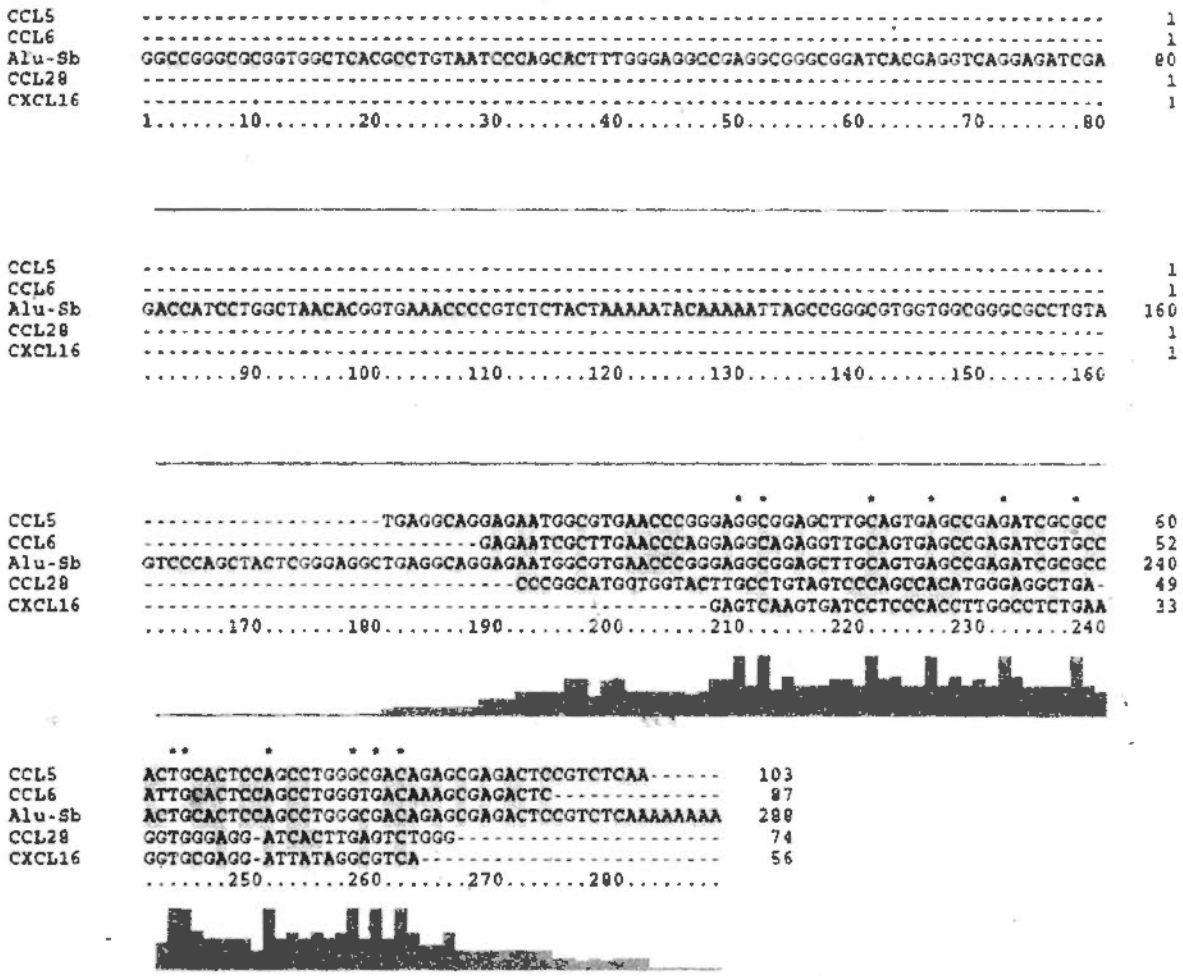


Figure 6.4 Multiple sequences alignment. The stem-loop regions in the 3'-UTR of human chemokines were aligned with Alu-sb by ClustalW 2.0 software (developed by the European Bioinformatics Institute).

6.5 Discussion

In a recent study, the authors reported that a diversity of 5'-modified long and short RNAs were generated from at least partially processed mRNAs by deep sequencing (Affymetrix, 2009). Our data showed that TIFm71 was an example of mRNA-derived ncRNA. One question we would like to ask is whether there are other human genes that potentially generate ncRNA from their untranslated regions. TIF is a member of chemokines family, therefore we started from analyzing the secondary structure of the 3'-UTR of chemokines in human by RNAfold WebServer in Vienna RNA Websuite. The results revealed that 13 out of 50 known chemokines could fold into at least one TIFm71-like hairpin secondary structure (50-100 nt) in the 3'-UTR region.

Among the 13 chemokines having one or more stem-loop secondary structures in the 3'-UTR, CCL5 was chosen for further analysis because its gene structure was similar to that of TIF as mentioned in the Introduction section. Both full-length and stem-loop region with flanking sequence of CCL5 were subcloned downstream of CMV promoter of pcDNA3 cloning and expression vector (Fig. 6.2A). By polyacrylamide gel Northern blot analysis, no ncRNA was detected around or less than 60 nt region after transient expression of either CCL5-FL or CCL5-SL in HEK293 cells, indicating the 3'-UTR of CCL5 could not generate ncRNA or miRNA. Obviously, our focus on chemokine family does not exclude other genes that are capable of generating ncRNAs. For this, further investigations will be required.

Although we did not see the processing of CCL5-SL, an endogenous large RNA transcript in HEK293 cells (>300 nt) was strongly hybridized in the polyacrylamide gel Northern blot analysis with a [γ-^{32}P] ATP-labeled probe complementary to the 3'-stem of CCL5-SL. Because of the high sequence similarity between CCL5-SL and consensus part of a human Alu element, we believe that the large RNA transcript detected by CCL5-SL probe was Alu element(s). This observation promoted us to check whether other stem-loop regions in the 3'-UTR of human chemokines also share sequence similarities with Alu elements. The result showed that many chemokine genes indeed have an Alu sequence-like 3'-UTR in the stem-loop region.

The Alu sequences within the 3'-UTR may affect mRNA stability or degradation by contributing adenine and uracil-rich elements to the transcript (An *et al.*, 2004) or by adenosine to inosine (A-to-I) editing in Alu sequences (Levanon *et al.*, 2004). They might also affect translational efficiency through providing secondary and tertiary structures to hinder translational editing or translational rates (Hagan *et al.*, 2003; Moolhuijzen *et al.*, 2010). In addition, it was hypothesized that Alu elements within human mRNAs probably act as miRNA targets (Smalheiser and Torvik, 2006).

Nevertheless, the amplitude of the "Alu phenomenon" in both human genome and transcriptome has been only partially uncovered (Hasler and Strub, 2006). It is of interest to investigate the potential biological functions of the highly abundant Alu RNA in the transcriptome, both independent transcripts and fragmented Alu elements embedded in the untranslated region of host mRNA. However, much more work will be required in the future to understand their roles in cell physiology and signaling.

Chapter 7

General Discussion

It has been widely acknowledged that ncRNAs are implicated in almost all known cellular processes and regulate diverse cellular functions (Mattick, 2009). However, the number of ncRNA type and their biological functions were not yet fully characterized. In the present study, we identified a new type of ncRNA, TIFm71, which was released from the 3'- untranslated region (3'-UTR) of an exonic TIF mRNA. TIFm71 was 71 nt in size and could fold into a canonical precursor microRNA-like stem-loop secondary structure. TIFm71 was found to be expressed in normal tissues at a low level but was greatly elevated in solid tumors. A high level of TIFm71 induced epithelial CHO-K1 cells transformation toward a less-differentiated mesenchymal fate to foster metastasis, namely epithelial-to-mesenchymal transitions (EMT). In addition, a reduction of cell population in G2/M phase was observed in TIFm71 stably over-expressing cells as examined by flow cytometric assay. Potential binding proteins of TIFm71 were identified by RNA-pull down and proteomics assay. We believed that the underlying mechanisms of TIFm71-induce EMT and the deregulation of cell cycle distribution are at least partially contributed by the proteins that bind with TIFm71.

7.1 Tumor induced factor (TIF)

A novel chemokine gene TIF was identified in xenograft tumors induced by oncogene *mas*-stably expressing CHO-K1 cells (Lin *et al.*, 2009). It encodes a 101 amino acid protein with a 28 amino acids signal peptide at its N-terminus, indicating TIF belongs to ELR$^+$ CXC chemokine subfamily. Preliminary functional studies revealed that TIF acted as a chemoattractant to neutrophil and also promoted blood vessel formation in an aortic ring assay. An inhibitory effect on tumorigenesis was found when co-

injecting TIF-stably over-expressing cells together with embryonic fibroblasts into nude mice to induce xenograft tumor, implying that TIF have an anti-tumor effect, possibly by inhibiting fibroblast proliferation (Zhoufang, unpublished data).

It was reported that a wide variety of cell types secret chemokines, which exert their effects on one or more target cell populations (Visser *et al.*, 1998). Several other cell types, including monocytes (Olsnes *et al.*, 2009) and macrophages (Goodman *et al.*, 1998) can produce chemokines without stimulation. Tumor cells can also secrete chemokines that act as autocrine growth factors or enhance metastasis (Kershaw *et al.*, 2002). As far as we know, there is no report on the chemokine expression in normal tissues. Here we showed that TIF is a ubiquitously expressed gene as examined by RT-PCR and Northern blot analysis. The expression of TIF mRNA was detected in a wide variety of normal tissues including kidney, liver, testis, heart, blood cells and others. The blood cells expressed TIF mRNA in a relatively high level in comparison to other tissues. These results indicated that the functions of chemokines might not restrict to inflammation or tumor, it probably has a fundamental function under physiological conditions.

7.2 Identification of TIFm71

An isoform of TIF, SY3, was also found to be up-regulated in xenograft tumors induced by oncogene *mas*-stably expressing CHO-K1 cells (Lin *et al.*, 2009). SY3 gene shares identical sequence with TIF except lacking a 71 nt string in the 3'-UTR. This 71 nt string was referred as TIFm71. It locates in the fourth exon, and GT/AG

exon splicing signal was observed at both ends of TIFm71, indicating that TIFm71 might be a potential intron. Interestingly, TIFm71 was predicted to fold into a pre-miRNA-like stem-loop structure by RNA folding programs (Gruber *et al.*, 2008). Moreover, sequence homologs were found using BLAST against miRNA database. Therefore, we hypothesized that TIF transcript was more than a protein-coding gene, and it may also encode a small ncRNA or a pre-miRNA TIFm71.

7.2.1 Expression of TIFm71

We examined the expression of TIFm71 by Northern blot analysis. Although the expression level was low in normal tissues, endogenous TIFm71 was detectable by Northern blot. However, the expression of TIFm71 was greatly up-regulated in solid tumors, which usually contains poorly vascularized regions characterized by severe hypoxia (oxygen deprivation), low pH, and nutrient starvation (Keith and Simon, 2007). Meanwhile, smaller RNAs that less than 71 nt was detected in solid tumors. The mechanisms of the formation of TIFm71 related small RNAs were unclear.

7.2.2 Processing of TIFm71

By deep sequencing, it was shown that a diversity of 5'-modified long and short RNAs were generated from processed mRNAs (Affymetrix, 2009). Our data is consistent with the idea that some protein-coding mRNAs could release ncRNAs. In order to investigate whether TIFm71 was a pre-miRNA or an ncRNA released from TIF mRNA, several fragments were subcloned downstream of CMV promoter of pcDNA3 and pIREShyg3 expression vectors. By polyacrylamide gel Northern blot

analysis, we showed that TIFm71 was released from the 3'-UTR of exonic transcript TIF in transiently transfected HEK293 cells. RNA processing is usually mediated by RNA binding proteins including serine/arginine rich (SR) proteins (Huang and Steitz, 2005), hnRNPs (Rossbach *et al.*, 2009), or some regulatory RNA molecules such as MALAT-1 (Lin *et al.*, 2011). In transiently transfected HEK293 cells, TIFm71 without flanking sequence can release from vector by itself, indicating TIFm71 have a self-splicing activity. However, we could not conclude that the flanking sequences are functionless, because the situation *in vivo* is different and more complicated than that in the transiently over-expressing cells. A possibility is that some inducible factors could bind to the flanking sequence of TIFm71 and facilitates TIFm71 processing under particular stimulus, and that would be why TIFm71 was greatly up-regulated in solid tumors.

7.2.3 Is TIFm71 a pre-miRNA?

Structure of pre-miRNAs is quite diverse. For example, different sites of stem in pre-miRNA may have different numbers of unpaired nucleotides, and the length of terminal loop by bioinformatic prediction could range from 3 nt to more than 20 nt (Zhu *et al.*, 2011). The lowest energy structures of pre-miRNAs such as mismatches, bulges, symmetrical and asymmetrical internal loops were analyzed for the presence of various secondary structure motifs (Zuker, 2003). Of the 1,243 motifs examined, 612 were asymmetrical (bulges and internal loops) and 631 were symmetrical (1-5 nt mismatches and internal loops). Single nucleotide mismatches and bulges accounted for most of the findings. The number of structural motifs in the pre-miRNA structures analyzed ranged from zero to seven, with an average of 2.7 motifs per precursor

(Starega-Roslan *et al.*, 2011; Zuker, 2003). Predicted secondary structure of TIFm71 contains 5 motifs, 2 single nucleotide mismatches, 2 asymmetrical bulges and one symmetrical bulge. It also has a 6 nt terminal loop structure (Fig. 3.3D), which was proved to influence Dicer's cleavage efficiency during miRNA maturation (Zhang and Zeng, 2010). These observations suggested that TIFm71 was likely to act as a pre-miRNA.

An alternative pathway of miRNA biogenesis was reported by Ruby *et al* (2007) in which debranced introns mimic the structure feature of pre-miRNA and entered Dicer processing pathway (Ruby *et al.*, 2007). However, there is no experimental evidence proving that a pre-miRNA could originate from the untranslated region of a protein-coding gene like an intron and then matured by the Dicer processing pathway. By Dicer *in vitro* processing assay, we showed that small RNA (~20 nt) was generated in TIFm71 and TIFm71 containing fragments, suggesting TIFm71 could serve as a substrate of Dicer.

A canonical mature miRNA should be expressed as a distinct transcript of ~22 nt in cells or tissues which is detectable by Northern blot analysis, real-time PCR or other experimental means (Berezikov *et al.*, 2006). After transient transfection into HEK293 cells, M299 generated a small RNA in ~21 nt but not the full-length TIF transcript or TIFm71. Although the only signal derived from M229 was a weak one, the miRNA expressed in *Mas* oncongene over-expressing Mc0M80 cells proved that TIF-miR was a real mature miRNA. Typically, a pre-miRNA hairpin contains a long imperfect stem of approximately 30 bp with flanking single-stranded RNA segments

at its base (Han *et al.*, 2006; Zeng and Cullen, 2005). This structure is recognized and cleaved by the microprocessor complex containing the ribonuclease Drosha (RNase III enzyme), DGCR8 and other proteins (Starega-Roslan *et al.*, 2011). This would explain why TIFm71 without flanking sequences did not generate TIF-miR. As to the full-length TIF, machinery of translation proteins probably occupied the binding sites of the microprocessor complex, therefore hindered TIF-miR processing.

The same miRNA was also detected in the solid tumor tissues, indicating that TIF-miR maybe an inducible miRNA and involved in tumorigenesis or cancer metastasis. Our data showed that hypoxic condition was not adequate to induce TIF-miR formation. Intriguingly, a pair of short repetitive RNA motif termed R-SBE-like (5'-ACAGAG-3'), was found in both 5'- and 3'-end of TIFm71. It was reported recently that a group of miRNAs (T/B-miRNAs), which contain a conserved sequence similar to Smad binding element (SBE, 5'-CAGAC-3'), could be regulated posttranscriptionally by TGFβ and BMP signaling. These conserved sequences were named R-SBE. Smad protein could directly associate with R-SBE and facilitate the maturation of T/B-miRNAs (Davis *et al.*, 2010). Therefore, a possibility arises that TGFβ and BMP signaling pathways may be also responsible for TIF-miR formation.

7.3 Characterization of TIFm71

A key question remains as to whether the small RNA TIFm71 is arising from degradation of the TIF transcript. As mentioned in the Introduction section, ncRNAs usually elicit their biological responses through interactions with RNA-binding

proteins (Charon *et al.*, 2010). Using biotinylated RNA pull-down and downstream proteomics analysis with mass spectrometry, a battery of TIFm71 RNA binding proteins was identified, suggesting that TIFm71 probably not only represents a simple degradation intermediate. In order to understand the potential functions of TIFm71 non-coding gene, we subcloned TIFm71 and its related sequences and established a series of CHO-K1 cell clones with stable random integration of TIFm71.

7.3.1 TIFm71 promoted EMT

We observed a striking change in cellular morphology of CHO-K1 cells by the over-expression of TIFm71, whereby the cobblestone-like appearance of epithelial cells switched to a fibroblast-, spindle-like morphology with extensive cellular scattering, indicating an EMT was induced. Besides CHO-K1, we established TIFm71 stable cell lines in MCF-7, which were derived from human mammalian adenocarcinoma and retained several characteristics of differentiated mammary epithelium (Levenson and Jordan, 1997). A small proportion of MCF-7 cells also tended to show a fibroblast-like morphology induced by TIFm71 over-expression (data not shown). However, in human embryonic kidney HEK293 cells that were transiently over-expressing TIFm71, no significant phenotypic change was observed (data not shown). These results imply the promotion of EMT by TIFm71 is cell type dependent or maybe also time dependent: persistent TIFm71 over-expression is more likely to induce EMT than transient expression.

Recently, Liao *et al* reported that miRNA cluster 302-367 could enhance somatic cell reprogramming by accelerating mesenchymal-to-epithelial transition (MET), a reverse process of EMT (Liao *et al.*, 2011). TGFβR2 could bind TGFβ cytokines and induces EMT through phosphorylation of TGFβR1 and activation of the Smad signaling pathway (Li *et al.*, 2010). The authors reported that TGFβR2 mRNA was targeted by cluster 302-367 to promote E-Cadherin expression, resulted in mesenchymal-to-epithelial changes necessary for colony formation (Liao *et al.*, 2011). As the R-SBE-like motifs found in TIFm71, it possible that TIFm71 also uses the Smad signaling pathway to promote EMT. In contrast to suppressive effect of cluster 302-367 on TGFβR2, TIFm71 might promote TGFβR2 expression or facilitate the binding of TGFβR2 and TGFβ.

Two lincRNAs, HOTAIR and MALAT-1, were demonstrated to promote cancer cell metastatic progression. HOTAIR is a large intervening ncRNAs (lincRNAs). It recruits the PRC2 complex to specific target genes genome-wide, leading to H3K27 trimethylation and epigenetic silencing of metastasis suppressor genes. Within the context of cancer cells, ectopic expression of HOTAIR seems to re-impose the chromatin state, thereby enabling gene expression programs that are conducive to cell motility and matrix invasion (Gupta *et al.*, 2010). MALAT-1, also a long ncRNA, was found to play a pivotal role in colorectal cancer metastasis (Ji *et al.*, 2003; Xu *et al.*, 2011), but the underlying mechanism was not unclear. TIFm71 is not a lincRNA, thus the mechanism of TIFm71 promoted EMT might be different from that of HOTAIR and MALAT-1. The canonical stem-loop secondary structure usually provide binding sites for RNA binding proteins (Tian *et al.*, 2004). Based on our current understanding of the molecular mechanisms by which ncRNAs functions, ncRNAs

commonly act as adaptors that position target molecules, which are nucleic acids or proteins to from a ribonucleoprotein unit for enzyme activity (Huttenhofer and Schattner, 2006; Wilusz *et al.*, 2009). We therefore hypothesized that TIFm71 might act as a scaffold to recruit EMT related proteins and affect their biological activities.

Among the TIFm71 RNA binding proteins we identified, Ran is potentially responsible for the EMT promotion. Ran is a small GTP binding protein belonging to the RAS superfamily. It was proposed that Ran may contribute to malignant transformation and/or enhanced proliferation in cancer cells (Lavia *et al.*, 2003), and the deregulation of Ran expression may lead to genomic instability (Azuma *et al.*, 2004). In addition, Ran is essential for the translocation of RNAs and proteins through the nuclear pore complex, and also involved in the control of DNA synthesis (Sorokin *et al.*, 2007; Stewart, 2007). Like other small GTPases, Ran cycles between a GTP- and GDP bound state; however, the conversion between these two forms occurs slowly in the absence of accessory factors (Weis, 2003). In our study, TIFm71 might promote the activity of Ran by direct binding or by recruiting accessory factors of Ran, such as the Ran guanine nucleotide exchange factor (RanGEF) RCC1, RanGTPase activating protein RanGAP and so on, resulting in malignant transformation (Weis, 2003).

7.3.2 TIFm71 and cell cycle progression

Cancer progression has been suggested to involve the loss of cell cycle checkpoint control that regulates the passage through the cell cycle. These checkpoints monitor the integrity of the DNA and make sure the genes are expressed in a coordinated

manner (Schlaepfer *et al.*, 1999; Wang *et al.*, 2000b). In the DNA flow cytometric analysis, we observed that there was a reduction of TIFm71 over-expressing CHO-K1 cells in G2/M phase. There are two possible explanations. First, TIFm71 might accelerate mitosis and shorten the G2/M phase; and second, TIFm71 possibly blocks the cells in S phase from entering G2/M phase. In both cases, the ratio of cells in G2/M will be relatively lower than normal cells when examined in a particular time point.

Two putative TIFm71 RNA binding proteins, Ran and MPP4, are probably responsible for the deregulation of cell cycle distribution. Previous studies implied that Ran controls the cell cycle through the regulation of nucleocytoplasmic transport, mitotic spindle organization, and nuclear envelope formation (Azuma *et al.*, 2004; Clarke and Zhang, 2008; Weis, 2003). MPP4 is a double-stranded RNA binding protein, localized to the nucleus and undergoes phosphorylation in M phase. It may be involved in the control of cell cycle progression (Abaza *et al.*, 2003; Matsumoto-Taniura *et al.*, 1996). We proposed that TIFm71 binding to Ran or MPP4, or both, possibly affected the activities of the proteins and result in a deregulation of cell cycle distribution by a yet to be identified pathway.

7.3.3 Relationship between EMT and cell cycle regulation

Although both EMT and deregulation of cell cycle were induced by the over-expression of TIFm71, a question still remained unanswered is that whether the two cellular events are mechanistically tied together. In a previous study, it was shown

that TGFβ induced apoptosis and EMT in a cell cycle-related manner, in which apoptosis and EMT took place at G2/M and G1/S phases, respectively (Yang *et al.*, 2006). In this regard, TIFm71-induced EMT shared the same characteristics as TGFβ-induced EMT/apoptosis.

On the other hand, the principal tumor-suppressor protein, p53, is best characterized as a transcription factor that binds to specific DNA sequences and transactivates a number of genes with a variety of functions including cell cycle arrest, apoptosis, leading to suppression of tumor progression, metastasis and other changes (Green and Kroemer, 2009; Riley *et al.*, 2008). Several months ago, two research groups made a similar discovery independently concerning the regulation of EMT by p53. They showed that p53 suppressed EMT by up-regulating miRNAs including miR-200 and miR-192 through direct binding to miRNA promoters (Chang *et al.*, 2011; Kim *et al.*, 2011). Meanwhile, p53 participates in the regulation of multiple cell cycle checkpoints (Giono and Manfredi, 2006). For instance, cell-cycle arrest depends on the ability of p53 to induce the transcription of target genes such as $p21^{CDKN1A}$, which is a CDK inhibitor (Chipuk and Green, 2006; Schlereth *et al.*, 2010). Therefore, a possibility arises that TIFm71 might counteract p53, either directly or indirectly, to affected the downstream signaling pathway and contribute to both EMT and reduction of cell population in G2/M phase.

Taken together, EMT and cell cycle regulation could be coordinated by cell signaling processes. However, how TIFm71 and its binding proteins regulate these processes

would require further studies on the signaling pathways in TIFm71 over-expressing cells.

7.4 Alu elements

7.4.1 Alu-like sequences in human chemokines

Since TIFm71 is the first example of mRNA-derived ncRNA, we tried to find other human chemokine genes that potentially generate ncRNA from their untranslated regions. The result showed that 13 of them could fold into at least one TIFm71-like hairpin secondary structure (50-100 nt) in their 3'-UTR region. Unexpectedly, stem-loop regions of CCL5, CCL6 (3rd stem-loop), CCL28 and CXCL16 share high similarity with human Alu sequence. These observations implied that Alu-like sequence in 3'-UTR is commonly present.

Later on, CCL5 was chosen for further analysis because its gene structure was similar to that of TIF in several aspects as mentioned in Introduction of Chapter 6. However, we did not detect any processing product from either CCL5-FL or CCL5-SL, suggesting the 3'-UTR of CCL5 could not generate ncRNA or miRNA. It should be noted that there are lacking of R-SBE-like motifs and AU-rich sequences in CCL5. Moreover, TIFm71 is complementary to Alu-J while the stem-loop of CCL5 shares similarity with Alu-sb. Thus we would like to ask whether the conserved motifs and the orientation of the stem-loop region would affect processing? Actually, the CXCL16 contains a stem-loop whose sequence is reverse complement to Alu

elements like TIFm71. Therefore, further studies will be required to investigate the processing of the stem-loop in CXCL16.

7.4.2 Alu elements and human diseases

Alu elements are non-autonomous retrotransposons that mobilize in a copy-and-paste fashion. They are known to create genetic instability and disease in a number of different ways, including insertional mutagenesis and then deletions/duplications through Alu/Alu non-allelic homologous recombination (Belancio *et al.*, 2010; Callinan and Batzer, 2006). Alu elements are commonly found in introns and do not alter gene function significantly. Alu insertion into coding regions account for several diseases including neurofibromatosis, haemophilia, leukemia, breast cancer and ovarian cancer (Callinan and Batzer, 2006; Deininger and Batzer, 1999). Both independent transcripts and fragmented Alu elements embedded in the 3'-UTR appear to have few negative affects, such as regulating mRNA stability or degradation by contributing AU-rich elements to the transcript (An *et al.*, 2004) or by adenosine to inosine (A-to-I) editing in Alu sequences (Levanon *et al.*, 2004). Also, they might affect translational efficiency through providing secondary and tertiary structures to hinder translational editing or translational rates (Hagan *et al.*, 2003; Moolhuijzen *et al.*, 2010).

However, the amplitude of the "Alu phenomenon" in both human genome and transcriptome has been only partially uncovered (Hasler and Strub, 2006). It would be of great interest to investigate the potential biological functions of the highly

abundant Alu RNA in the transcriptome, both independent transcripts and fragmented Alu elements embedded in the untranslated region of host mRNA.

7.5 Conclusions

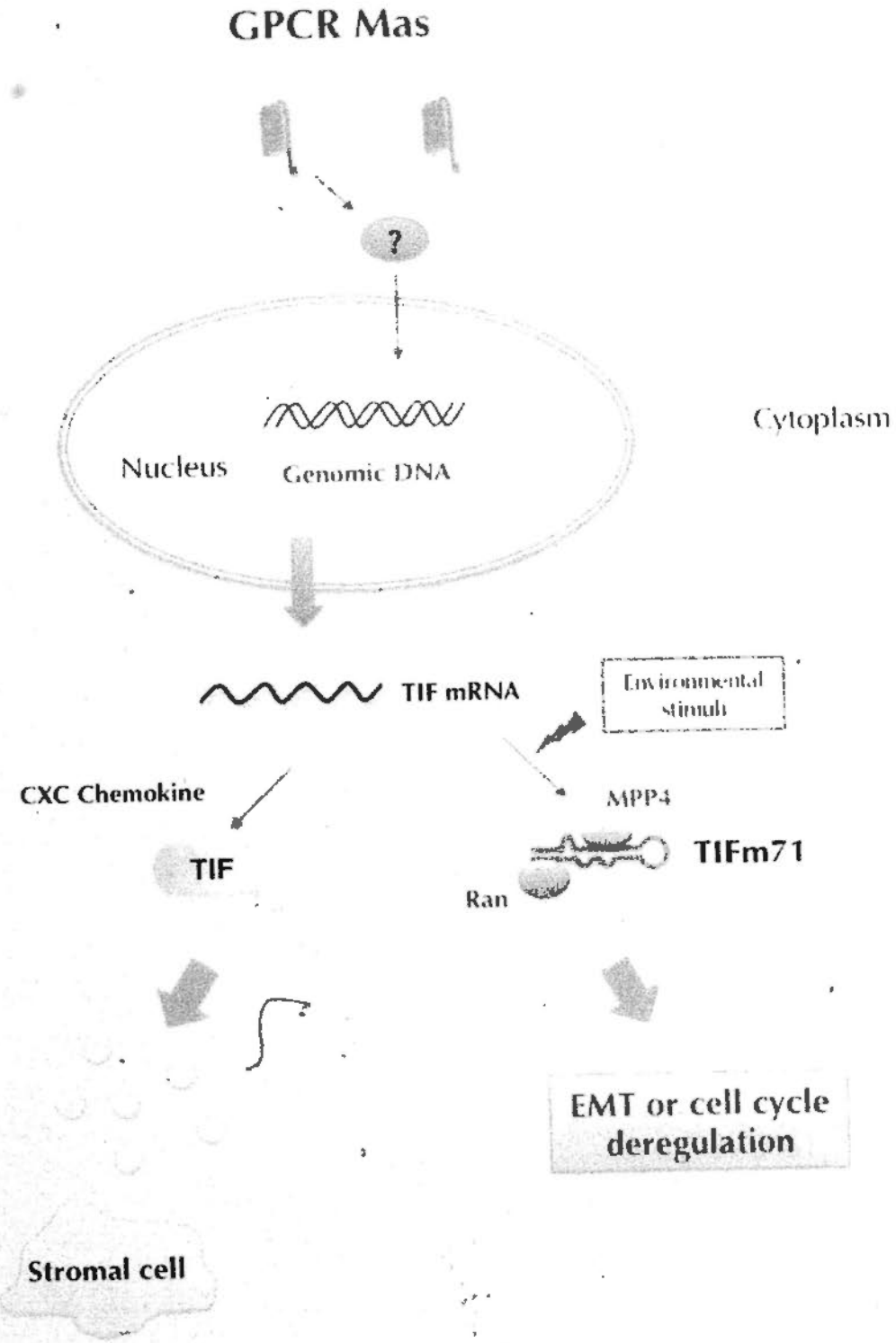
This thesis contains two major discoveries: first, we presented a novel ncRNA TIFm71 which was originated from the untranslated region of a protein-coding mRNA, and second, functional assay showed that ectopic expression of TIFm71 promoted the non-invasive, epithelial like ovary cells to undergo mesenchymal transition, and also caused a deregulation of cell cycle distribution.

A cellular model of TIFm71 biogenesis and functions was summarized in Fig. 7.1. On the one hand, the secretory chemokine encoded by TIF gene could be secreted outside the host cells and interacts with stromal cells, affecting its proliferation and trans-differentiation, modifying the microenvironment both *in vitro* and *in vivo*. On the other hand, TIFm71 was released to induce the host epithelial CHO-K1 cells toward a less-differentiated mesenchymal fate to foster metastasis. Also, it altered the cell cycle parttern, probably by binding to and affecting protein activities such as MPP4 and Ran. The two pathways may be alternatively activated. Under physiological condition, it is expected that TIF transcript is tightly associated with translation-related proteins to produce chemokine TIF. However, under stress stimuli, translation-related proteins might be replaced with TIFm71 processing/metastasis-related proteins, resulting in the generation of TIFm71 RNA, which may increase malignancy

of tumor cells. Therefore, our model provided some implications for the understanding of the multiple functions of a gene.

By the end of this thesis, several questions remain to be answered. What kind of stimulus will be responsible for TIF-miR formation? What will be the target(s) of this miRNA and how could it be involved in tumorigenesis or metastasis? Regarding the TIF mRNA that does not release TIFm71 ncRNA, would TIFm71 sequence act as a targeting site for miRNA? Besides binding to proteins, does TIFm71 target RNAs or DNAs to elicit other biological functions (Matsui *et al.*, 2010)? What are the roles of the highly abundant Alu elements in the transcriptome? This project is just a beginning for the characterization of the novel ncRNA TIFm71 and for the exonic ncRNA/miRNA discovery. Much more work will be required in the future to answer these questions.

THIS IS A BLANK PAGE


GPCR Mas
?
Nucleus
Genomic DNA
Cytoplasm
TIF mRNA
Environmental stimuli
CXC Chemokine
MPP4
TIF
TIFm71
Ran
EMT or cell cycle deregulation
Stromal cell

References

Abaza, A., Soleilhac, J. M., Westendorf, J., Piel, M., Crevel, I., Roux, A., and Pirollet, F. (2003). M phase phosphoprotein 1 is a human plus-end-directed kinesin-related protein required for cytokinesis. The Journal of biological chemistry *278*, 27844-27852.

Affymetrix (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature *457*, 1028-1032.

Alkan, S. A., Martincic, K., and Milcarek, C. (2006). The hnRNPs F and H2 bind to similar sequences to influence gene expression. Biochem J *393*, 361-371.

Allen, S. J., Crown, S. E., and Handel, T. M. (2007). Chemokine: receptor structure, interactions, and antagonism. Annu Rev Immunol *25*, 787-820.

Allen, T. A., Von Kaenel, S., Goodrich, J. A., and Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat Struct Mol Biol *11*, 816-821.

Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell *113*, 673-676.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., *et al.* (2003). A uniform system for microRNA annotation. RNA *9*, 277-279.

An, H. J., Lee, D., Lee, K. H., and Bhak, J. (2004). The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3' untranslated regions. BMC genomics *5*, 97.

Azuma, K., Sasada, T., Takedatsu, H., Shomura, H., Koga, M., Maeda, Y., Yao, A., Hirai, T., Takabayashi, A., Shichijo, S., and Itoh, K. (2004). Ran, a small GTPase gene, encodes cytotoxic T lymphocyte (CTL) epitopes capable of inducing HLA-A33-restricted and tumor-reactive CTLs in cancer patients. Clin Cancer Res *10*, 6695-6702.

Bagga, P. S., Arhin, G. K., and Wilusz, J. (1998). DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. Nucleic Acids Res *26*, 5343-5350.

Baggiolini, M., Dewald, B., and Moser, B. (1997). Human chemokines: an update. Annu Rev Immunol *15*, 675-705.

Barnes, E. A., Kenerson, H. L., Jiang, X., and Yeung, R. S. (2010). Tuberin regulates E-cadherin localization: implications in epithelial-mesenchymal transition. Am J Pathol *177*, 1765-1778.

Barrallo-Gimeno, A., and Nieto, M. A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development (Cambridge, England) *132*, 3151-3161.

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

Baskerville, S., and Bartel, D. P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA (New York, NY) *11*, 241-247.

Batlle, E., Sancho, E., Franci, C., Dominguez, D., Monfar, M., Baulida, J., and Garcia De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol *2*, 84-89.

Baum, B., Settleman, J., and Quinlan, M. P. (2008). Transitions between epithelial and mesenchymal states in development and disease. Semin Cell Dev Biol *19*, 294-308.

Belancio, V. P., Roy-Engel, A. M., and Deininger, P. L. (2010). All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol *20*, 200-210.

Berezikov, E., Cuppen, E., and Plasterk, R. H. (2006). Approaches to microRNA discovery. Nature genetics *38 Suppl*, S2-7.

Birchmeier, W., and Behrens, J. (1994). Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. Biochim Biophys Acta *1198*, 11-26.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem *72*, 291-336.

Borchert, G. M., Lanier, W., and Davidson, B. L. (2006). RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol *13*, 1097-1101.

Brendle, A., Lei, H., Brandt, A., Johansson, R., Enquist, K., Henriksson, R., Hemminki, K., Lenner, P., and Forsti, A. (2008). Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. Carcinogenesis *29*, 1394-1399.

Buratti, E., and Baralle, F. E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol *24*, 10505-10514.

Burd, C. G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science (New York, NY) *265*, 615-621.

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 (New York, NY) *10*, 1957-1966.

Callinan, P. A., and Batzer, M. A. (2006). Retrotransposable elements and human disease. Genome Dyn *1*, 104-115.

Cano, A., Perez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., and Nieto, M. A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol *2*, 76-83.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Cassetti, M. C., Noah, D. L., Montelione, G. T., and Krug, R. M. (2001). Efficient translation of mRNAs in influenza A virus-infected cells is independent of the viral 5' untranslated region. Virology *289*, 180-185.

Chang, C. J., Chao, C. H., Xia, W., Yang, J. Y., Xiong, Y., Li, C. W., Yu, W. H., Rehman, S. K., Hsu, J. L., Lee, H. H., *et al.* (2011). p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. Nat Cell Biol *13*, 317-323.

Charo, I. F., and Ransohoff, R. M. (2006). The many roles of chemokines and chemokine receptors in inflammation. N Engl J Med *354*, 610-621.

Charon, C., Moreno, A. B., Bardou, F., and Crespi, M. (2010). Non-protein-coding RNAs and their interacting RNA-binding proteins in the plant cell nucleus. Mol Plant *3*, 729-739.

Chen, C. Y., and Shyu, A. B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci *20*, 465-470.

Chipuk, J. E., and Green, D. R. (2006). Dissecting p53-dependent apoptosis. Cell Death Differ *13*, 994-1002.

Choi, P. S., Zakhary, L., Choi, W. Y., Caron, S., Alvarez-Saavedra, E., Miska, E. A., McManus, M., Harfe, B., Giraldez, A. J., Horvitz, R. H., *et al.* (2008). Members of the miRNA-200 family regulate olfactory neurogenesis. Neuron *57*, 41-55.

Christensen, B. C., Moyer, B. J., Avissar, M., Ouellet, L. G., Plaza, S. L., McClean, M. D., Marsit, C. J., and Kelsey, K. T. (2009). A let-7 microRNA-binding site

polymorphism in the KRAS 3' UTR is associated with reduced survival in oral cancers. Carcinogenesis *30*, 1003-1007.

Clark, F., and Thanaraj, T. A. (2002). Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Hum Mol Genet *11*, 451-464.

Clarke, P. R., and Zhang, C. (2008). Spatial and temporal coordination of mitosis by Ran GTPase. Nat Rev Mol Cell Biol *9*, 464-477.

Clement, J. Q., Maiti, S., and Wilkinson, M. F. (2001). Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem *276*, 16919-16930.

Clement, J. Q., Qian, L., Kaplinsky, N., and Wilkinson, M. F. (1999). The stability and fate of a spliced intron from vertebrate cells. RNA (New York, NY) *5*, 206-220.

Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Costa, F. F. (2008). Non-coding RNAs, epigenetics and complexity. Gene *410*, 9-17.

Cuesta, R., Xi, Q., and Schneider, R. J. (2000). Adenovirus-specific translation by displacement of kinase Mnk1 from cap-initiation complex eIF4F. The EMBO journal *19*, 3465-3474.

Davis, B. N., Hilyard, A. C., Nguyen, P. H., Lagna, G., and Hata, A. (2010). Smad Proteins Bind a Conserved RNA Sequence to Promote MicroRNA Maturation by Drosha. Mol Cell *39*, 373-384.

Deininger, P. L., and Batzer, M. A. (1999). Alu repeats and human disease. Mol Genet Metab *67*, 183-193.

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.

Dennis, C. (2002). The brave new world of RNA. Nature *418*, 122-124.

Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J. M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. The EMBO journal *17*, 3091-3100.

Di Leva, G., and Croce, C. M. (2010). Roles of small RNAs in tumor formation. Trends Mol Med *16*, 257-267.

Espinoza, C. A., Allen, T. A., Hieb, A. R., Kugel, J. F., and Goodrich, J. A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat Struct Mol Biol *11*, 822-829.

Espinoza, C. A., Goodrich, J. A., and Kugel, J. F. (2007). Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. RNA (New York, NY) *13*, 583-596.

Esquela-Kerscher, A., and Slack, F. J. (2006). Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer *6*, 259-269.

Ferrell, J. E., Jr., Tsai, T. Y., and Yang, Q. (2011). Modeling the cell cycle: why do certain circuits oscillate? Cell *144*, 874-885.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res *36*, 2257-2267.

Gerlitz, G., and Bustin, M. (2011). The role of chromatin structure in cell migration. Trends Cell Biol *21*, 6-11.

Giono, L. E., and Manfredi, J. J. (2006). The p53 tumor suppressor participates in multiple cell cycle checkpoints. J Cell Physiol *209*, 13-20.

Gkirtzou, K., Tsamardinos, I., Tsakalides, P., and Poirazi, P. (2010). MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. PLoS One *5*, e11843.

Gomes, L. R., Terra, L. F., Sogayar, M. C., and Labriola, L. (2011). Epithelial-Mesenchymal Transition: Implications in Cancer Progression and Metastasis. Curr Pharm Biotechnol, in press.

Goodman, R. B., Strieter, R. M., Frevert, C. W., Cummings, C. J., Tekamp-Olson, P., Kunkel, S. L., Walz, A., and Martin, T. R. (1998). Quantitative comparison of C-X-C chemokines produced by endotoxin-stimulated human alveolar macrophages. Am J Physiol *275*, L87-95.

Goodrich, J. A., and Kugel, J. F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. Nat Rev Mol Cell Biol *7*, 612-616.

Grayson, M. H., and Holtzman, M. J. (2006). Chemokine signaling regulates apoptosis as well as immune cell traffic in host defense. Cell Cycle *5*, 380-383.

Green, D. R., and Kroemer, G. (2009). Cytoplasmic functions of the tumour suppressor p53. Nature *458*, 1127-1130.

Grosso, A. R., Gomes, A. Q., Barbosa-Morais, N. L., Caldeira, S., Thorne, N. P., Grech, G., von Lindern, M., and Carmo-Fonseca, M. (2008). Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res *36*, 4823-4832.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. Nucleic Acids Res *36*, W70-74.

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Tsai, M. C., Hung, T., Argani, P., Rinn, J. L., *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071-1076.

Hagan, C. R., Sheffield, R. F., and Rudin, C. M. (2003). Human Alu element retrotransposition induced by genotoxic stress. Nat Genet *35*, 219-220.

Hahn, D., and Beggs, J. D. (2010). Brr2p RNA helicase with a split personality: insights into structure and function. Biochem Soc Trans *38*, 1105-1109.

Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T., and Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Hasler, J., and Strub, K. (2006). Alu elements as regulators of gene expression. Nucleic Acids Res *34*, 5491-5497.

Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2011). Mass spectrometry-driven proteomics: an introduction. Methods in molecular biology (Clifton, NJ) *753*, 1-27.

Hendrickson, S. L., Lautenberger, J. A., Chinn, L. W., Malasky, M., Sezgin, E., Kingsley, L. A., Goedert, J. J., Kirk, G. D., Gomperts, E. D., Buchbinder, S. P., *et al.*

(2010). Genetic variants in nuclear-encoded mitochondrial genes influence AIDS progression. PLoS ONE *5*, e12862.

Herrin, G. L., McCurdy, H. H., and Wall, W. H. (2005). Investigation of an LC-MS-MS (QTrap) method for the rapid screening and identification of drugs in postmortem toxicology whole blood samples. J Anal Toxicol *29*, 599-606.

Holden, N. S., and Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. J Pharmacol Toxicol Methods *63*, 7-14.

Hood, J. D., and Cheresh, D. A. (2002). Role of integrins in cell invasion and migration. Nat Rev Cancer *2*, 91-100.

Huang, C., Bi, E., Hu, Y., Deng, W., Tian, Z., Dong, C., and Sun, B. (2006). A novel NF-kappaB binding site controls human granzyme B gene transcription. J Immunol *176*, 4173-4181.

Huang, Y., and Steitz, J. A. (2005). SRprises along a messenger's journey. Mol Cell *17*, 613-615.

Hugo, H., Ackland, M. L., Blick, T., Lawrence, M. G., Clements, J. A., Williams, E. D., and Thompson, E. W. (2007). Epithelial--mesenchymal and mesenchymal--epithelial transitions in carcinoma progression. J Cell Physiol *213*, 374-383.

Huttenhofer, A., and Schattner, P. (2006). The principles of guiding by RNA: chimeric RNA-protein enzymes. Nat Rev Genet *7*, 475-482.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Inui, M., Martello, G., and Piccolo, S. (2010). MicroRNA control of signal transduction. Nat Rev Mol Cell Biol *11*, 252-263.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene *22*, 8031-8041.

Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science (New York, NY) *302*, 2141-2144.

Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kalluri, R., and Neilson, E. G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest *112*, 1776-1784.

Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. J Clin Invest *119*, 1420-1428.

Kamal, M., Xie, X., and Lander, E. S. (2006). A large family of ancient repeat elements in the human genome is under strong selection. Proc Natl Acad Sci U S A *103*, 2740-2745.

Kang, Y., and Massague, J. (2004). Epithelial-mesenchymal transitions: twist in development and metastasis. Cell *118*, 277-279.

Kapp, L. D., and Lorsch, J. R. (2004). The molecular mechanics of eukaryotic translation. Annu Rev Biochem *73*, 657-704.

Keith, B., and Simon, M. C. (2007). Hypoxia-inducible factors, stem cells, and cancer. Cell *129*, 465-472.

Kershaw, M. H., Wang, G., Westwood, J. A., Pachynski, R. K., Tiffany, H. L., Marincola, F. M., Wang, E., Young, H. A., Murphy, P. M., and Hwu, P. (2002). Redirecting migration of T cells to chemokine secreted from tumors by genetic modification with CXCR2. Hum Gene Ther *13*, 1971-1980.

Khabar, K. S. (2010). Post-transcriptional control during chronic inflammation and cancer: a focus on AU-rich elements. Cell Mol Life Sci *67*, 2937-2955.

Kim, T., Veronese, A., Pichiorri, F., Lee, T. J., Jeon, Y. J., Volinia, S., Pineau, P., Marchio, A., Palatini, J., Suh, S. S., *et al.* (2011). p53 regulates epithelial-mesenchymal transition through microRNAs targeting ZEB1 and ZEB2. J Exp Med *208*, 875-883.

Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kim, V. N., and Nam, J. W. (2006). Genomics of microRNA. Trends Genet *22*, 165-173.

Kleinman, H. K., McGarvey, M. L., Hassell, J. R., Star, V. L., Cannon, F. B., Laurie, G. W., and Martin, G. R. (1986). Basement membrane complexes with biological activity. Biochemistry *25*, 312-318.

Kleinman, H. K., McGarvey, M. L., Liotta, L. A., Robey, P. G., Tryggvason, K., and Martin, G. R. (1982). Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. Biochemistry *21*, 6188-6193.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597-610.

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., and Gage, F. H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. Cell *116*, 779-793.

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA (New York, NY) *9*, 175-179.

Lam, Y. W., Evans, V. C., Heesom, K. J., Lamond, A. I., and Matthews, D. A. (2010). Proteomics analysis of the nucleolus in adenovirus-infected cells. Mol Cell Proteomics *9*, 117-130.

Lavia, P., Mileo, A. M., Giordano, A., and Paggi, M. G. (2003). Emerging roles of DNA tumor viruses in cell proliferation: new insights into genomic instability. Oncogene *22*, 6508-6516.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal *23*, 4051-4060.

Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S. R., Sztybel, D., *et al.* (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol *22*, 1001-1005.

Levenson, A. S., and Jordan, V. C. (1997). MCF-7: the first hormone-responsive breast cancer cell line. Cancer Res *57*, 3071-3078.

Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A *100*, 189-192.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., *et al.* (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51-63.

Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X., Qin, B., *et al.* (2011). MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. J Biol Chem *286*, 17359-17364.

Lin, R., Roychowdhury-Saha, M., Black, C., Watt, A. T., Marcusson, E. G., Freier, S. M., and Edgington, T. S. (2011). Control of RNA processing by a large non-coding RNA over-expressed in carcinomas. FEBS letters *585*, 671-676.

Lin, W. Z., Li, Z. F., Tsang, S. Y., Lung, L. K., Wang, D. K., Chan, W. Y., Zhu, Y. K., Lee, S. S., and Cheung, W. T. (2009). Identification and characterization of a novel CXC chemokine in xenograft tumor induced by mas-overexpressing cells. Int J Cancer *125*, 1316-1327.

Lowe, C. B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci U S A *104*, 8005-8010.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in Alternative Pre-mRNA Splicing. Cell *144*, 16-26.

Luco, R. F., and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr Opin Genet Dev, in press.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science (New York, NY) *303*, 95-98.

Luster, A. D. (1998). Chemokines--chemotactic cytokines that mediate inflammation. N Engl J Med *338*, 436-445.

Maiato, H. (2010). Mitosis: wisdom, knowledge, and information. Cell Mol Life Sci *67*, 2141-2143.

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S., Mis, E., Zegar, C., Gutwein, M. R., Khivansara, V., *et al.* (2010). The Landscape of C. elegans 3'UTRs. Science *329*, 432-5.

Mantovani, A., Savino, B., Locati, M., Zammataro, L., Allavena, P., and Bonecchi, R. (2010). The chemokine system in cancer biology and therapy. Cytokine Growth Factor Rev *21*, 27-39.

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Matlin, A. J., Clark, F., and Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

Matsui, M., Sakurai, F., Elbashir, S., Foster, D. J., Manoharan, M., and Corey, D. R. (2010). Activation of LDL Receptor Expression by Small RNAs Complementary to a Noncoding Transcript that Overlaps the LDLR Promoter. Chem Biol *17*, 1344-1355.

Matsumoto-Taniura, N., Pirollet, F., Monroe, R., Gerace, L., and Westendorf, J. M. (1996). Identification of novel M phase phosphoproteins by expression cloning. Mol Biol Cell *7*, 1455-1469.

Mattick, J. S. (2009). The genetic signatures of noncoding RNAs. PLoS genetics *5*, e1000459.

Mattick, J. S., and Makunin, I. V. (2005). Small regulatory RNAs in mammals. Hum Mol Genet *14 Spec No 1*, R121-132.

McDonald, B., and Kubes, P. (2010). Chemokines: Sirens of Neutrophil Recruitment-but Is It Just One Song? Immunity *33*, 148-149.

Mi, Z., Bhattacharya, S. D., Kim, V. M., Guo, H., Talbot, L. J., and Kuo, P. C. (2011). Osteopontin promotes CCL5-mesenchymal stromal cell-mediated breast cancer metastasis. Carcinogenesis *32*, 477-487.

Miettinen, P. J., Ebner, R., Lopez, A. R., and Derynck, R. (1994). TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. J Cell Biol *127*, 2021-2036.

Moolhuijzen, P., Kulski, J. K., Dunn, D. S., Schibeci, D., Barrero, R., Gojobori, T., and Bellgard, M. (2010). The transcript repeat element: the human Alu sequence as a component of gene networks influencing cancer. Funct Integr Genomics *10*, 307-319.

Nawijn, M. C., Hackett, T. L., Postma, D. S., van Oosterhout, A. J., and Heijink, I. H. (2011). E-cadherin: gatekeeper of airway mucosa and allergic sensitization. Trends Immunol *32*, 248-255.

Nelson, D. M., Ye, X., Hall, C., Santos, H., Ma, T., Kao, G. D., Yen, T. J., Harper, J. W., and Adams, P. D. (2002). Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. Mol Cell Biol *22*, 7459-7472.

Nelson, H. H., Christensen, B. C., Plaza, S. L., Wiencke, J. K., Marsit, C. J., and Kelsey, K. T. (2010). KRAS mutation, KRAS-LCS6 polymorphism, and non-small cell lung cancer. Lung Cancer *69*, 51-53.

Nieto, M. A. (2002). The snail superfamily of zinc-finger transcription factors. Nat Rev Mol Cell Biol *3*, 155-166.

Nishihara, H., Smit, A. F., and Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res *16*, 864-874.

Novotny, M., Datta, S., Biswas, R., and Hamilton, T. (2005). Functionally independent AU-rich sequence motifs regulate KC (CXCL1) mRNA. J Biol Chem *280*, 30166-30174.

Nyberg, K. A., Michelson, R. J., Putnam, C. W., and Weinert, T. A. (2002). Toward maintaining the genome: DNA damage and replication checkpoints. Annu Rev Genet *36*, 617-656.

Ohler, U., Yekta, S., Lim, L. P., Bartel, D. P., and Burge, C. B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA *10*, 1309-1322.

Olsnes, C., Stavang, H., Brokstad, K., Olofsson, J., and Aarstad, H. J. (2009). Chemokines are secreted by monocytes following OK-432 (lyophilized Streptococcus pyogenes) stimulation. BMC Immunol *10*, 6.

Park, Y. W., Wilusz, J., and Katze, M. G. (1999). Regulation of eukaryotic protein synthesis: selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. Proc Natl Acad Sci U S A *96*, 6694-6699.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol Cell *25*, 635-646.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer *7*, 415-428.

Pertz, O., Bozic, D., Koch, A. W., Fauser, C., Brancaccio, A., and Engel, J. (1999). A new crystal structure, Ca2+ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. The EMBO journal *18*, 1738-1747.

Piecyk, M., Wax, S., Beck, A. R., Kedersha, N., Gupta, M., Maritim, B., Chen, S., Gueydan, C., Kruys, V., Streuli, M., and Anderson, P. (2000). TIA-1 is a translational silencer that selectively regulates the expression of TNF-alpha. The EMBO journal *19*, 4154-4163.

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. Clin Biochem Rev *30*, 19-34.

Polyak, K., and Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer *9*, 265-273.

Ponicsan, S. L., Kugel, J. F., and Goodrich, J. A. (2010). Genomic gems: SINE RNAs regulate mRNA production. Curr Opin Genet Dev *20*, 149-155.

Ponjavic, J., and Ponting, C. P. (2007). The long and the short of RNA maps. Bioessays *29*, 1077-1080.

Prasanth, K. V., and Spector, D. L. (2007). Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev *21*, 11-42.

Raman, D., Baugher, P. J., Thu, Y. M., and Richmond, A. (2007). Role of chemokines in tumor growth. Cancer letters *256*, 137-165.

Reichman, T. W., and Mathews, M. B. (2003). RNA binding and intramolecular interactions modulate the regulation of gene expression by nuclear factor 110. RNA (New York, NY) *9*, 543-554.

Reichman, T. W., Parrott, A. M., Fierro-Monti, I., Caron, D. J., Kao, P. N., Lee, C. G., Li, H., and Mathews, M. B. (2003). Selective regulation of gene expression by nuclear factor 110, a member of the NF90 family of double-stranded RNA-binding proteins. J Mol Biol *332*, 85-98.

Ren, B. (2010). Transcription: Enhancers make non-coding RNA. Nature *465*, 173-174.

Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., Parsons, J. T., and Horwitz, A. R. (2003). Cell migration: integrating signals from front to back. Science (New York, NY) *302*, 1704-1709.

Riley, T., Sontag, E., Chen, P., and Levine, A. (2008). Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol *9*, 402-412.

Rodriguez, L. G., Wu, X., and Guan, J. L. (2005). Wound-healing assay. Methods in molecular biology (Clifton, NJ) *294*, 23-29.

Rollins, B. J. (1997). Chemokines. Blood *90*, 909-928.

Rossbach, O., Hung, L. H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., and Bindereif, A. (2009). Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. Molecular and cellular biology *29*, 1442-1451.

Rossi, D., and Zlotnik, A. (2000). The biology of chemokines and their receptors. Annual review of immunology *18*, 217-242.

Rostene, W., Guyon, A., Kular, L., Godefroy, D., Barbieri, F., Bajetto, A., Banisadr, G., Callewaere, C., Conductier, G., Rovere, C., *et al.* (2011). Chemokines and chemokine receptors: new actors in neuroendocrine regulations. Front Neuroendocrinol *32*, 10-24.

Rot, A., and von Andrian, U. H. (2004). Chemokines in innate and adaptive host defense: basic chemokinese grammar for immune cells. Annu Rev Immunol *22*, 891-928.

Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Ryan, B. M., Robles, A. I., and Harris, C. C. (2010). Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer *10*, 389-402.

Saj, A., and Lai, E. C. (2011). Control of microRNA biogenesis and transcription by cell signaling pathways. Curr Opin Genet Dev, in press.

Satyanarayana, A., and Kaldis, P. (2009). Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. Oncogene *28*, 2925-2939.

Schickel, R., Park, S. M., Murmann, A. E., and Peter, M. E. (2010). miR-200c regulates induction of apoptosis through CD95 by targeting FAP-1. Mol Cell *38*, 908-915.

Schlaepfer, D. D., Hauck, C. R., and Sieg, D. J. (1999). Signaling through focal adhesion kinase. Prog Biophys Mol Biol *71*, 435-478.

Schlereth, K., Beinoraviciute-Kellner, R., Zeitlinger, M. K., Bretz, A. C., Sauer, M., Charles, J. P., Vogiatzi, F., Leich, E., Samans, B., Eilers, M., *et al.* (2010). DNA binding cooperativity of p53 modulates the decision between cell-cycle arrest and apoptosis. Mol Cell *38*, 356-368.

Seeburg, P. H. (2000). RNA helicase participates in the editing game. Neuron *25*, 261-263.

Sharma, M. (2010). Chemokines and their receptors: orchestrating a fine balance between health and disease. Crit Rev Biotechnol *30*, 1-22.

Shi, H., Zhang, Z., Wang, X., Liu, S., and Teng, C. T. (1999). Isolation and characterization of a gene encoding human Kruppel-like factor 5 (IKLF): binding to

the CAAT/GT box of the mouse lactoferrin gene promoter. Nucleic Acids Res *27*, 4807-4815.

Shivdasani, R. A. (2006). MicroRNAs: regulators of gene expression and cell differentiation. Blood *108*, 3646-3653.

Siomi, H., and Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. Curr Opin Genet Dev *7*, 345-353.

Smalheiser, N. R., and Torvik, V. I. (2006). Alu elements within human mRNAs are probable microRNA targets. Trends Genet *22*, 532-536.

Sorokin, A. V., Kim, E. R., and Ovchinnikov, L. P. (2007). Nucleocytoplasmic transport of proteins. Biochemistry (Mosc) *72*, 1439-1457.

Stamatakos, M., Palla, V., Karaiskos, I., Xiromeritis, K., Alexiou, I., Pateras, I., and Kontzoglou, K. (2010). Cell cyclins: triggering elements of cancer or not? World J Surg Oncol *8*, 111.

Starega-Roslan, J., Koscianska, E., Kozlowski, P., and Krzyzosiak, W. J. (2011). The role of the precursor structure in the biogenesis of microRNA. Cell Mol Life Sci.

Stark, M., Bram, E. E., Akerman, M., Mandel-Gutfreund, Y., and Assaraf, Y. G. (2011). Heterogeneous nuclear ribonucleoprotein H1/H2-dependent unsplicing of thymidine phosphorylase results in anticancer drug resistance. J Biol Chem *286*, 3741-3754.

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

Stewart, M. (2007). Molecular mechanism of the nuclear protein import cycle. Nat Rev Mol Cell Biol *8*, 195-208.

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., *et al.* (2001). Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res *11*, 677-684.

Szymanski, M., Barciszewska, M. Z., Zywicki, M., and Barciszewski, J. (2003). Noncoding RNA transcripts. J Appl Genet *44*, 1-19.

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. J Pathol *220*, 126-139.

Tang, G., Reinhart, B. J., Bartel, D. P., and Zamore, P. D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev *17*, 49-63.

Tang, R. B., Wang, H. Y., Lu, H. Y., Xiong, J., Li, H. H., Qiu, X. H., and Liu, H. Q. (2005). Increased level of polymerase III transcribed Alu RNA in hepatocellular carcinoma tissue. Mol Carcinog *42*, 93-96.

Thiery, J. P. (2002). Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer *2*, 442-454.

Thiery, J. P., Acloque, H., Huang, R. Y., and Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thiery, J. P., and Sleeman, J. P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol *7*, 131-142.

Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., and Mathews, M. B. (2004). The double-stranded-RNA-binding motif: interference and much more. Nat Rev Mol Cell Biol *5*, 1013-1023.

Tong, W. Y., Liang, Y. M., Tam, V., Yip, H. K., Kao, Y. T., Cheung, K. M., Yeung, K. W., and Lam, Y. W. (2010). Biochemical characterization of the cell-biomaterial interface by quantitative proteomics. Mol Cell Proteomics *9*, 2089-2098.

Trujillo, M. A., Sakagashira, M., and Eberhardt, N. L. (2006). The human growth hormone gene contains a silencer embedded within an Alu repeat in the 3'-flanking region. Mol Endocrinol *20*, 2559-2575.

Vicente-Manzanares, M., Webb, D. J., and Horwitz, A. R. (2005). Cell migration at a glance. J Cell Sci *118*, 4917-4919.

Visone, R., and Croce, C. M. (2009). MiRNAs and Cancer. Am J Pathol *174*, 1131-1138.

Visser, C. E., Tekstra, J., Brouwer-Steenbergen, J. J., Tuk, C. W., Boorsma, D. M., Sampat-Sardjoepersad, S. C., Meijer, S., Krediet, R. T., and Beelen, R. H. (1998). Chemokines produced by mesothelial cells: huGRO-alpha, IP-10, MCP-1 and RANTES. Clin Exp Immunol *112*, 270-275.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, W., Furneaux, H., Cheng, H., Caldwell, M. C., Hutter, D., Liu, Y., Holbrook, N., and Gorospe, M. (2000a). HuR regulates p21 mRNA stabilization by UV light. Mol Cell Biol *20*, 760-769.

Wang, W., Heideman, L., Chung, C. S., Pelling, J. C., Koehler, K. J., and Birt, D. F. (2000b). Cell-cycle arrest at G2/M and growth inhibition by apigenin in human colon carcinoma cell lines. Mol Carcinog *28*, 102-110.

Wang, Z., and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA (New York, NY) *14*, 802-813.

Wassarman, K. M., and Storz, G. (2000). 6S RNA regulates E. coli RNA polymerase activity. Cell *101*, 613-623.

Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell *112*, 441-451.

Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev *23*, 1494-1504.

Wright, M. W., and Bruford, E. A. (2011). Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics *5*, 90-98.

Wu, Y., and Zhou, B. P. (2010). Snail: More than EMT. Cell Adh Migr *4*, 199-203.

Xu, C., Yang, M., Tian, J., Wang, X., and Li, Z. (2011). MALAT-1: A long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. Int J Oncol *39*, 169-175.

Yang, J., and Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev Cell *14*, 818-829.

Yang, Y., Pan, X., Lei, W., Wang, J., and Song, J. (2006). Transforming growth factor-beta1 induces epithelial-to-mesenchymal transition and apoptosis via a cell cycle-dependent mechanism. Oncogene *25*, 7235-7244.

Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science (New York, NY) *304*, 594-596.

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

Young, D., Waitches, G., Birchmeier, C., Fasano, O., and Wigler, M. (1986). Isolation and characterization of a new cellular oncogene encoding a protein with multiple potential transmembrane domains. Cell *45*, 711-719.

Zamore, P. D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science (New York, NY) *309*, 1519-1524.

Zeng, Y., and Cullen, B. R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. J Biol Chem *280*, 27595-27603.

Zhang, W., Winder, T., Ning, Y., Pohl, A., Yang, D., Kahn, M., Lurje, G., Labonte, M. J., Wilson, P. M., Gordon, M. A., *et al.* (2011). A let-7 microRNA-binding site polymorphism in 3'-untranslated region of KRAS gene predicts response in wild-type KRAS patients with metastatic colorectal cancer treated with cetuximab monotherapy. Ann Oncol *22*, 104-109.

Zhang, X., and Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res *38*, 7689-7697.

Zhou, H., Hu, H., and Lai, M. (2010). Non-coding RNAs and their epigenetic regulatory mechanisms. Biol Cell *102*, 645-655.

Zhu, S., Jiang, Q., Wang, G., Liu, B., Teng, M., and Wang, Y. (2011). Chromatin structure characteristics of pre-miRNA genomic sequences. BMC genomics *12*, 329.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.