



# VCU

Virginia Commonwealth University  
VCU Scholars Compass

---

Theses and Dissertations

Graduate School

---

2017

## TAF2: A potential oncogene for hepatocellular carcinoma

Saranya Chidambaranathan Reghupaty  
*Virginia Commonwealth University*

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Molecular Genetics Commons](#)

© The Author

---

Downloaded from

<https://scholarscompass.vcu.edu/etd/4924>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

# **TAF2: A potential oncogene for hepatocellular carcinoma**

A thesis submitted in partial fulfillment of the requirements for the degree  
of Master of Science Virginia Commonwealth University

By

**SARANYA CHIDAMBARANATHAN REGHUPATY, B.TECH.**

Department of Biotechnology, Sathyabama University, India, 2014

**ADVISOR: DR. DEVANAND SARKAR, M.B.B.S., Ph.D.**

Associate Professor, Department of Human and Molecular Genetics  
Harrison Foundation Distinguished Professor in Cancer Research  
Blick Scholar  
Associate Scientific Director, Cancer Therapeutics  
VCU Institute of Molecular Medicine  
Associate Director of Education and Training  
Massey Cancer Center

**Virginia Commonwealth University**

**Richmond, Virginia**

**August 2017**

## ACKNOWLEDGEMENT

I extend my sincere gratitude to my advisor, Dr. Devanand Sarkar, for providing me with the opportunity to be his student and work under his guidance. I thank him for his motivation, encouragement and immense knowledge. He is an inspirational scientist and person without whose guidance and supervision this research would not have been possible.

Besides my advisor, I also thank the other members of my committee: Dr. Keith Baker and Dr. Andrei Ivanov for their valuable comments and insights.

I thank my lab members: Nidhi Jariwala, for always being available whenever I had doubts or concerns about my research. Rachel Mendoza, for teaching me most of the techniques I have used in this research. I also thank Dr. Chadia L Robertson and Dr. Maaged Akiel for having been a constant support.

Huge thanks to each and every one of my family members for believing in me. The liberty I have received from you means a lot. I thank my friend Harini for being a great companion during these two years as a student at VCU. I also thank the faculty members and my fellow students in the Department of Human Genetics for having been supportive of my efforts.

## TABLE OF CONTENTS

1. List of tables	v
2. List of figures	Vi
3. List of abbreviations and symbols	Vii
4. Abstract	Ix
5. Introduction	1
5.1. Transcription	1
5.2. Transcription factors	2
5.3. TBP Associated Factors	5
5.4. Assembly of TAFs	6
5.5. TAF2	7
5.6. Role of transcription factors in cancer	8
5.7. Hepatocellular Carcinoma	10
5.7.1. HCC	10
5.7.2. HCC risk factors	10
5.7.3. HCC incidence	12
5.7.4. HCC treatment	13
5.8. Gene amplification in cancer	13
5.9. AEG-1 in HCC	14
5.10. TAF2 in HCC	15

5.11. Cancer hallmarks	16
5.11.1. Self sufficiency in growth signals	16
5.11.2. Insensitivity to anti-growth signals	16
5.11.3. Evasion of apoptosis	17
5.11.4. Limitless replicative potential	17
5.11.5. Sustained angiogenesis	17
5.11.6. Tissue invasion and metastasis	18
5.12. Epithelial-Mesenchymal transition	20
6. Materials and methods	21
7. Results	28
8. Figures and table	33
9. Discussion and future directions	43

## LIST OF TABLES

<b>Table 2.1:</b> TAF2 mRNA expression in HCC patients and normal subjects	29
--	----

## LIST OF FIGURES

Figure 1.1 Model for holo-TFIID assembly	27
Figure 1.2 Hallmarks of cancer	27
Figure 2.1 Correlation of TAF2 with AEG1 expression	28
Figure 2.2 Amplification and copy number gain of TAF2 in HCC patients	28
Figure 2.3 Kaplan-Meier survival curve of HCC patients	29
Figure 2.4 Gene clustering with TAF2	30
Figure 2.5 Overexpression of TAF2 in HCC cell lines and HCC liver	31
Figure 2.6 Generation of stable TAF2 knockdown clones	32
Figure 2.7 Effect of TAF2 knockdown on cell proliferation	33
Figure 2.8 Effect of TAF2 knockdown on migration	34
Figure 2.9 Effect of TAF2 knockdown on invasion	35
Figure 2.10 Western blot of cyclins and EMT markers	36

**LIST OF ABBREVIATIONS AND SYMBOLS**

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
RNAP	RNA polymerase
mRNA	messenger RNA
snRNA	small nuclear RNA
CPE	Core Promoter Element
DPE	Downstream Core Promoter Element
Inr	Initiator
GTF	General Transcription Factor
TAF	TBP-Associated Factor
TBP	TATA binding protein
PIC	Preinitiation Complex
TFII	Transcription Factor II
ISGF-3	Interferon Stimulated Gene Factor-3
HCC	Hepatocellular carcinoma
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
NASH	Non alcoholic steatohepatitis



SHARP	Study of Heart and Renal Protection
AEG-1	Astrocyte Elevated Gene-1
EMT	Epithelial-Mesenchymal Transition
TGF-beta	Transforming Growth Factor Beta
E-cadherin	Epithelial cadherin
N-cadherin	Neural cadherin
FBS	Fetal Bovine Serum
shRNA	short hairpin RNA
CD24	Cluster of differentiation 24
MAP4K4	Mitogen-activated protein kinase
HGF	Hepatocyte Growth Factor
VEGF	Vascular Endothelial growth Factor
RB1	Retinoblastoma 1

## **ABSTRACT**

### **TAF2: A POTENTIAL ONCOGENE FOR HEPATOCELLULAR CARCINOMA**

By **SARANYA CHIDAMBARANATHAN REGHUPATY, B.Tech.**

A thesis submitted in partial fulfillment of the requirements for the degree of  
Master of Science Virginia Commonwealth University, 2017.

**ADVISOR: DR. DEVANAND SARKAR, M.B.B.S., Ph.D.**

Associate Professor, Department of Human and Molecular Genetics  
Harrison Foundation Distinguished Professor in Cancer Research  
Blick Scholar  
Associate Scientific Director, Cancer Therapeutics  
VCU Institute of Molecular Medicine  
Associate Director of Education and Training  
Massey Cancer Center

Astrocyte Elevated Gene 1 (AEG1) is an oncogene for hepatocellular carcinoma (HCC). Its role in HCC pathogenesis has been well studied. A pan cancer analysis of gene expression in multiple databases identified TATA-box binding protein associated factor 2 (TAF2) as the gene that is most frequently co-expressed with AEG1. TAF2 is a protein that is involved in transcription of genes by RNA polymerase II. It is a factor that is dispensable for basal transcription but, required for activated transcription. It has also been shown to be involved in regulating cyclin levels and hence cell cycle progression. Bioinformatic analysis on data from different cancer databases confirmed the positive correlation of TAF2 expression with AEG1 expression, the over expression

of TAF2 in HCC patients and poor survival of HCC patients with increasing TAF2. We confirmed the over expression of TAF2 in HCC cell lines using western blotting and HCC liver using immunohistochemistry. We established cell lines with stable knockdown of TAF2 expression. These clones showed significant decrease in their ability to invade and migrate but not their proliferation ability. This is in contrast to what has been observed in previous studies. We hypothesize that the knockdowns do not show any decrease in cellular proliferation since the remaining TAF2 in the cells is sufficient to produce cyclins and keep cell cycle undisturbed. The knockdown of TAF2 causes an increase in E-cadherin level and decrease in Snail protein expression which is a known negative regulator of E-cadherin. Knockdown of TAF2 causes cells to become more epithelial leading to a decrease in their ability to migrate and invade. This study shows that TAF2 is a potential oncogene that needs to be further studied.

## **CHAPTER 1**

### **INTRODUCTION**

#### **THE PROCESS OF TRANSCRIPTION**

All cells in our body, except germ cells and red blood cells, have the same set of DNA fragments enclosed within their nucleus. But, brain cells are structurally and functionally different from liver cells. What makes two cells with the same genetic composition different from each other is the way their genes are expressed. A gene is a part of a DNA segment that provides instructions in the form of a linear sequence of nucleotides to code for an mRNA. This mRNA is then translated to produce a protein. Gene expression is thus a two-step process involving transcription and translation. In some cases, such as ribosomal RNAs, RNA is the final product of the gene. Hence, translation of the RNA to a polypeptide does not occur. Cells do not require all of their gene products, be it RNA or proteins, at all times. Only a subset of its genes is actively being expressed at any given point of time. This difference in gene expression differentiates one cell type from the other.

Transcription is the first step in gene expression, wherein RNA is synthesized from a DNA template. Tight regulation of transcription is essential to controlling gene expression<sup>1</sup>. Transcription is carried out by the nuclear enzyme RNA polymerase (RNAP). Prokaryotes have a single RNAP enzyme that catalyzes the transcription of all of its genes whereas

eukaryotes have three RNAPs each involved in the synthesis of different types of RNA<sup>2</sup>. RNAPI transcribes ribosomal RNAs (rRNAs)<sup>3</sup>. RNAPII is responsible for the synthesis of mRNA precursors and non-coding RNAs like snRNA<sup>4</sup> and microRNA<sup>5</sup>. RNAPIII transcribes small RNAs like transfer RNAs (tRNAs)<sup>6</sup>. RNAPII is the extensively studied of all three RNAPs since it is involved in the transcription of protein-coding genes.

Transcription occurs in three stages-Initiation, elongation and termination. Initiation of transcription takes place when RNA polymerase binds to the promoter region of a gene and unwinds the double stranded DNA at the transcription start site. During elongation, RNA polymerase travels along the template DNA strand in the 5' to 3' direction, adding nucleotides to the 3' end of the growing RNA strand<sup>7</sup>. The steps involved in termination of transcription differ for each RNA polymerase. Pre-mRNA transcripts produced by RNAPII have a unique polyadenylation signal (AAUAAA) at the 3' end. As RNAPII creates this polyadenylation signal, it is recognized by two protein complexes CPSF (Cleavage and Polyadenylation Specificity Factor)<sup>8</sup> and CSTF (Cleavage Stimulation Factor)<sup>9,10</sup>. These protein complexes then recruit other proteins to carry out RNA cleavage.

## **TRANSCRIPTION FACTORS**

Transcription factors (TFs) are molecules that initiate and regulate the process of transcription. They are protein complexes with DNA binding domains that allow them to

bind to specific DNA sequences like promoters. They can also bind to enhancers or repressors resulting in activation or repression of gene expression<sup>11</sup>. Certain genes have varying levels of transcription. One is the basal transcription, which is usually a low level of expression. It is also called minimal or constitutive expression. General transcription factors are required for basal transcription and necessary for all genes. This low expression level can be enhanced to reach a high level of transcription, called activated transcription<sup>12</sup>. Activated transcription factors, unlike basal transcription factors, are gene specific.

The initiation of transcription requires the binding of TFs to the core promoter element resulting in the recruitment of the appropriate RNAP. Several core promoter elements (CPEs) can bind these TFs to initiate transcription. The most important of these are TATA box, Initiator (Inr) element and downstream core promoter element (DPE). The TATA box was the first eukaryotic core promoter element to be identified. In metazoans, it is typically located about 25-30 nucleotides upstream of the transcription start site. The consensus sequence for the TATA box is 5'-TATAAA-3'<sup>13</sup>. The Initiator element is another core promoter element that is located -6 basepairs upstream of the transcription start site and continues to around +11 basepairs downstream. The consensus sequence for the Inr element is YYANWYY<sup>14</sup>. The DPE is another core promoter element located about 28–33 nucleotides downstream of the transcription start site. The DPE consensus sequence

is RGWYV(T)<sup>15</sup>. These CPEs can be present by themselves or in combination with other CPEs. For example, DPE and Inr elements are found in many promoters that lack TATA box. There are also promoters that contain these elements in addition to the TATA box. Approximately twenty four percent of human genes have a TATA-like element in their promoter. But, only ten percent of these TATA-containing promoters have the canonical TATA box. Forty six percent of human core promoters contain the consensus INR and thirty percent are INR-containing TATA-less genes. Forty six percent of human promoters lack both TATA-like and consensus INR elements<sup>16</sup>.

The RNAP II transcription machinery is extremely complex made up of more than 70 polypeptides. It involves, in addition to transcription factors, several cofactors and mediators that aid RNAPII enzyme in transcription<sup>17-19</sup>. RNA polymerase II preinitiation complex is a 31 protein 1.5 megadalton protein complex. The formation of this complex is important to initiate the process of transcription<sup>20-22</sup>. It is a sequential process involving the binding of several General Transcription Factors (GTFs) to the promoter sequence<sup>23</sup>. The first of these steps is the recognition and binding of the GTF Transcription Factor IID (TFIID) to the core promoter element. Once TFIID is bound, it recruits TFIIA and then TFIIB. Once these three GTFs are bound, RNA polymerase and TFIIF are recruited. Finally TFIIE and TFIIH are recruited completing the preinitiation complex formation<sup>24</sup>. TFIID is the largest of the GTFs involved in eukaryotic transcription. It is a complex made

up of TATA binding protein (TBP) and several associated factors called TBP associated factors (TAFs). TBP is the specific subunit that binds to the TATA box whereas TAFs are coactivators of transcription.

### **TATA-BOX BINDING PROTEIN (TBP) ASSOCIATED FACTORS (TAFs)**

When GTFs were identified, TFIID was found to bind unambiguously to the TATA box. It was the only GTF that had sequence-specific DNA binding activity. TBP was the first constituent of TFIID to be cloned. Purified TBP was able to replace TFIID and assemble the PIC *in vitro*<sup>25</sup>. It was initially believed that TBP and TFIID were identical. However, it was soon found that TBP was sufficient for basal transcription, but unable to support activated transcription<sup>26,27</sup>. Biochemical fractionation of TFIID revealed that TBP had coactivators associated with it and formed a large, stable multi-subunit complex. When this whole complex was used, activated transcription was achieved. It was thus concluded that TFIID comprises not only TBP, but also TBP-associated factors (TAFs). TAFs are not a requirement for basal transcription but are necessary for activated transcription. They are dispensable for the activation of specific genes<sup>28</sup>. For example, eighty four percent of yeast genes are dependent upon one or more TAFs for their transcription whereas the remaining sixteen percent are TAF independent<sup>29</sup>. Till date, approximately 13 TAFs have been identified so far that vary greatly in size from as low as 15 KDa to as high as 250 KDa<sup>30</sup>.



TAFs have been cloned and characterized in Yeast<sup>31,32</sup>, *Drosophila*<sup>33-35</sup> and Humans<sup>27,36,37</sup> revealing that they are evolutionary conserved<sup>35,38</sup>. All three species have various shared sequence motifs, such as WD40 repeats (yTAF<sub>II</sub>90/dTAF<sub>II</sub>80/hTAF<sub>II</sub>100), as well as limited sequence homology to histones H3 (yTAF<sub>II</sub>17/dTAF<sub>II</sub>40/hTAF<sub>II</sub>31), H4 (yTAF<sub>II</sub>60/dTAF<sub>II</sub>60/hTAF<sub>II</sub>70), and H2B (yTAF<sub>II</sub>61/dTAF<sub>II</sub>30 $\alpha$ /hTAF<sub>II</sub>20)<sup>39</sup>. Interestingly, human TAF<sub>II</sub>130 and *Drosophila* TAF<sub>II</sub>110 lack an obvious yeast homolog, suggesting that the yeast and metazoan TFIID complexes, despite their many similarities, may nevertheless differ in functionally important respects.

### **ASSEMBLY OF TAFs**

TFIID is a twenty-subunit complex made up of 14 different polypeptides - TBP and 13 TAFs. TAFs are present in single copies or two copies in the TFIID complex. The complex is composed of one copy of TBP and TAFs 1, 2, 3, 7, 8, 11 and 13 and two copies of TAFs 4, 5, 6, 9, 10 and 12. Five of the six TAFs that are present in two copies (TAF4, 5, 6, 9 and 12) form a two-fold symmetric scaffold. TAF8 dimerises with TAF10 to form a TAF8–TAF10 complex in the cytoplasm. This is imported into the nucleus by importins. Binding of one copy of TAF8–TAF10 breaks the symmetry in core-TFIID, resulting in an asymmetric complex. It exhibits two distinct halves and new binding surfaces for further

subunits. Accretion of remaining TAFs and TBP in single copy, results in asymmetric clamp shaped holo-TFIID that nucleates the preinitiation complex (Fig. 1.1)<sup>40</sup>. A key feature in TAFs is the histone fold domain (HFD), which is present in nine out of 13 TAFs in TFIID. The HFD is a strong protein-protein interaction motif that mediates specific dimerization<sup>41</sup>.

### **TATA BOX BINDING PROTEIN ASSOCIATED FACTOR 2 (TAF2)**

TAF2 is one of these several TAFs and has been shown to be homologous between Yeast (TAF150 or TMS1), *Drosophila* (dTAFII150) and Mammals (TAF2, CIF150 or hTAFII150)<sup>42,43</sup>. This 150 KDa protein forms a trimer with TAF1 and TBP and binds specifically to the Initiator (Inr) element found at the transcription start site of many genes<sup>42,44</sup>. An endogenous trimer made up of TAF2, TAF8 and TAF10 has been identified suggesting that this sub complex subsequently binds with other TAFs to form the complete PIC<sup>45</sup>. Tomography experiments have suggested that TFIID complex can alter between four different conformations and binding of TAF2 stabilizes TFIID in a particular conformation<sup>46</sup>. As with other TAFs, TAF2 is not universally required for transcription. Only 3 percentage of yeast genome is dependent on TAF2 for transcription<sup>29</sup>. Some of the potential TAF2 targets that have been identified include ribosomal proteins L44, S10 and L7a, ISGF-3, metallothionein II and lipid kinase<sup>47</sup>. It is also involved in the expression of cyclins A and B1. Temperature sensitive TAF2 mutations in yeast and transient

knockout of TAF2 in mammalian cell lines both led to cell cycle arrest in the G2/M phase<sup>47,48</sup>.

## **ROLE OF TRANSCRIPTION FACTORS IN CANCER**

Cells are the basic building blocks that make up every living system. For an organism to grow and function normally, it is important that their cells divide and duplicate in an organized manner. Cell division refers to the series of events that a cell goes through to produce two daughter cells. It is made up of the interphase and the mitotic phase. Interphase is the phase in which cells spend most of their life. It is made of G1, S and G2 phases. During the G1 phase, the cells prepare to duplicate its DNA. DNA gets duplicated during S phase and the cells prepare for mitosis during G2 phase. Once the cells pass through interphase the mitotic phase begins. This is when cells share the duplicated DNA equally between two daughter cells. Mitosis is made up of prophase, prometaphase, metaphase, anaphase and telophase. After mitosis, some cells enter the G0 phase where they are in a quiescent state. Some cells like nerve cells and red blood cells that are terminally differentiated stay in this phase since there is no need for them to divide again. Adult mature hepatocytes are unique in that they stay in G0 phase until a regenerative process is initiated due to tissue loss or intoxication.

The three checkpoints - G1, G2 and M make sure that the cell stalls its division if the

conditions are not right. Once the cells go through the G1 checkpoint, they are irreversibly committed to division. The G2 checkpoint makes sure that the DNA produced after replication is free of damage. If there is any kind of damage to the DNA, the DNA damage repair mechanism fixes the damage before the cell moves to mitosis. The M checkpoint makes sure that all sister chromatids are correctly attached to the spindle microtubules before they are pulled towards the pole during anaphase. Several proteins are involved in keeping the cell cycle under control. An aberration in the level of these proteins can disrupt this control. When this process of cell cycle is perturbed, abnormal cell growth occurs. This is the primary difference between a normal cell and a tumor and explains the simplest definition of cancer, which is uncontrolled cell division.

Several proteins are involved in this process of transforming a normal cell into a tumor. A broad way of classifying these proteins is as oncoproteins and tumor suppressor proteins. Oncoproteins enhance cell division and viability. In a cancer cell, these proteins are overexpressed causing abnormal cell division. On the other hand, tumor suppressor proteins, that are required to slow down cell division and promote cell death are down regulated in a cancerous state. Transcription factors have an important role to play in producing these RNAs and proteins at optimum levels. Changes in levels of these transcription factors can lead to abnormal cell division causing cancer.

## **HEPATOCELLULAR CARCINOMA**

The liver is made up of four different types of cells namely hepatocytes, kupffer cells, stellate cells and liver sinusoidal cells. Of these four cell types, hepatocytes are the major functional cells and make up 75 percent of the liver. A cancer caused due to malignant tumors originating in these cells is called hepatocellular carcinoma (HCC). HCC is the most common primary liver cancer. Approximately 70 to 90 percent of liver cancers are HCC<sup>49</sup>. According to Globocan 2012, it is the fifth most common cancer in men and the ninth common cancer in women worldwide. It is the second most common cause of death from cancer worldwide. The overall ratio of mortality to incidence is 0.95. The similarity between the incidence and mortality reflects the poor prognosis of this disease.

### **HCC risk factors**

Most patients with HCC have liver cirrhosis<sup>49</sup>. It is a condition in which the hepatocytes slowly lose their proliferation capacity, preventing the liver from regenerating. As a result, scar tissue replaces the healthy liver tissue and partially blocks the flow of blood to the liver. The development of cirrhosis is a slow and gradual process. In the early stages of cirrhosis, the liver continues to function. But with worsening cirrhosis, the liver fails. HCC can develop anytime during the progression of liver cirrhosis<sup>50</sup>. Telomerase dysfunction and alterations in the micro and macro environment that stimulate cellular proliferation are some of the ways in which liver cirrhosis can lead to HCC<sup>51</sup>.

HBV infection is responsible for more than 50 percent of HCC cases worldwide<sup>52</sup>. HBV is a partially double-stranded DNA-containing virus belonging to the Hepadnaviridae family. In most cases, HBV infection can cause cirrhosis and eventually HCC. However, HBV can also cause HCC in the absence of cirrhosis. HBV can integrate its DNA into host cells and act as a mutagenic agent causing HCC<sup>53</sup>.

HCV infection is another risk factor of HCC and accounts for approximately 30 percent of HCC cases globally<sup>52</sup>. HCV is an RNA virus and unable to integrate into the host genome<sup>54</sup>. Thus, HCV causes HCC exclusively through liver cirrhosis. HCV core protein enters the host cell, where it localizes in the outer mitochondrial membrane and endoplasmic reticulum and promotes oxidative stress. This results in the activation of key signaling pathways leading to HCC<sup>55</sup>. Chronic hepatitis C is more aggressive in HIV positive subjects, leading to cirrhosis and liver failure<sup>56</sup>.

Chronic alcohol consumption is another important risk factor for HCC. It leads to liver damage through endotoxins, oxidative stress or inflammation causing HCC. It can also have genotoxic effect by enhancing the expression of certain oncogene expression or impairing the ability of cells to mend their DNA<sup>57</sup>. Chronic alcohol consumption can also act as a cofactor for HCC development synergistically with other risk factors like viral infections.

Another important HCC risk factor, especially in developing countries is Nonalcoholic steatohepatitis (NASH). The major cause of NASH is obesity<sup>58</sup>. Obesity leads to non-

alcoholic fatty liver disease, which in turn causes the inflammatory form non-alcoholic steatohepatitis and in turn promotes HCC<sup>59</sup>. Other risk factors of HCC include aflatoxin B1, pesticides, diabetes mellitus and diet<sup>57,60-62</sup>.

### **Trends in HCC incidence**

There is a striking variation in HCC incidence rates between various countries. The incidence rate is highest in developing countries particularly Eastern Asia and South-Eastern Asia, intermediate in Southern Europe and Northern America and lowest in Northern Europe and South-Central Asia<sup>63</sup>. Although at a global level, the major burden of HCC still falls on developing countries, HCC incidence has been decreasing in Asian countries<sup>64</sup>. This decrease can be attributed to programs to reduce aflatoxin B1 exposure and HBV transmission, declining rates of HCV infection in the population and other public health efforts<sup>65</sup>. In contrast, developed countries like the US have been experiencing an increase in HCC incidence<sup>66,67</sup>. Most of the increase in HCC cases can be attributed to the aging cohort with chronic hepatitis C infection. Increasing incidence of obesity and non-alcoholic fatty liver disease is another factor responsible for the rise in HCC incidences in the US<sup>68,69</sup>.

## **HCC treatment**

Liver transplantation is the best treatment option for patients with late cirrhosis<sup>70</sup>. Surgical resection is another treatment option available for patients with a single nodule, no cirrhosis and good liver function<sup>71,72</sup>. Local ablation with radiofrequency is the standard of care for patients with very early and early stage tumors that are unsuitable for surgery<sup>73</sup>. Transcatheter arterial chemoembolization is a minimally invasive procedure that involves administration of chemotherapy and embolization materials directly to a liver tumor via a catheter to restrict a tumor's blood supply. This is recommended for patients with intermediate-stage HCC<sup>74</sup>. Sorafenib is the only approved chemotherapy drug to treat HCC. It is an oral bi-aryl urea, which inhibits multiple cell surface and downstream kinases involved in tumor progression. Two phase III randomized placebo-controlled trials, the SHARP trial conducted mainly in America and Europe and a similar trial conducted in Asia reported improved overall survival with sorafenib<sup>75,76</sup>. In the SHARP trial, the median overall survival increased to 10.7 months with sorafenib compared to 7.9 months with placebo<sup>76</sup>. HCC is usually detected at late stages and treatment options for late stage HCC are minimum. This reveals the importance of finding new drug targets to develop HCC drugs.

## **GENE AMPLIFICATION IN CANCER**

Gene amplification is the increase in copy number of a restricted region of a chromosome



arm. It is a mechanism that cells use to overexpress certain genes for survival under stress such as during exposure to cytotoxic drugs. Some of the proposed mechanisms for the occurrence of gene amplification include extra replication and recombination, replication fork stalling and template switching, the breakage and fusion bridge cycle and double rolling circle replication<sup>77</sup>. It is frequently observed in some solid tumors and contributes to tumor evolution. MYC was the first oncogene that was proven to be amplified in a variety of tumor cells<sup>78-80</sup>. Later on, few chromosomal regions were found to be amplified to a great extent in many cancers. One such chromosomal segment that is frequently amplified in many cancers is the chromosomal region 8q<sup>47,48,81-84</sup>. This is the chromosomal region that houses the MYC oncogene. The MYC oncogene is present in the chromosomal region 8q24. Increasing evidence has shown that there are other neighboring regions in the long arm of chromosome 8 that are amplified<sup>85,86</sup>. 8q22 is one such region that was identified<sup>87</sup>. One of the many oncogenes present in this gene is the MTDH gene.

### **ASTROCYTE ELEVATED GENE-1 IN HCC**

AEG-1/MTDH is an oncogene which is overexpressed in many common cancers<sup>88-92</sup>, and its expression level negatively correlates with poor survival and overall adverse prognosis<sup>93-96</sup>. AEG-1 overexpression induces an aggressive, angiogenic and metastatic phenotype whereas knockdown of AEG-1 inhibits these phenotypes in all cancers studied

so far. AEG-1 is transcriptionally regulated by c-Myc, an oncogene frequently upregulated in HCC<sup>97</sup>. The tumor suppressor miRNA miR-375, which is downregulated in HCC patients, targets AEG-1<sup>98</sup>. In vitro and in vivo studies have shown that AEG-1 modulates expression of genes associated with proliferation, invasion, chemoresistance, angiogenesis and metastasis, and activates multiple pro-tumorigenic signaling pathways<sup>99</sup>. These studies indicate that AEG-1 is a valid target for HCC therapeutic development.

## **TAF2 IN CANCER**

The gene encoding the protein TAF2 is present in the chromosomal region 8q24.12. This region is a mutational hotspot and is frequently amplified in a variety of cancers. The first association between TAF2 and cancer emerged in 2014 when cBioPortal for cancer genomics was used to analyse TAF alterations in cancer. It was found that TAF2 exhibits copy number increases or mRNA overexpression in 73% of high grade serous ovarian cancer<sup>100</sup>. Other than overexpression, TAF2 is also hypomethylated in breast carcinoma<sup>101</sup>. Though TAF2 has been hypothesized as an oncogene solely based on bioinformatics data, no studies have looked at TAF2 in the context of cancer progression so far. Research has been going for decades now to better understand the role of TAF2 in transcription. Though, the role of TAF2 in transcription at a molecular level has been better understood, its role in cancer progression, the downstream targets, pathways

regulated and interacting partners are yet to be unraveled.

## **CANCER HALLMARKS**

Tumorigenesis refers to the production or formation of tumors. It is a multistep process, the progression of which depends on a sequential accumulation of mutations within tissue cells. Cancer is a manifestation of six essential alterations in cell physiology that collectively dictate malignant growth (Fig. 1.2)<sup>102</sup>.

### **Self-sufficiency in growth signals**

Cells of the body require molecules that act as signals for them to grow and divide. Cancer cells, however, grow in the absence of growth stimulatory signals that normal cells require from their environment. Cancer cells can do this by autocrine signaling, cell surface receptor overexpression and integrin switching<sup>102</sup>.

### **Insensitivity to anti-growth signals**

Cells have tight control over cell division. Tumor suppressor genes are important proteins in this process. These genes halt cell division if the cell fails any of the three checkpoints. In cancer, these tumour suppressor proteins are altered so that they don't effectively prevent cell division, even when the cell has abnormalities. Cancer cells also do not have contact inhibition, and will continue to grow and divide, regardless of their surroundings<sup>103</sup>.

### **Evasion of apoptosis**

Apoptosis or programmed cell death is initiated by cells when they are damaged or infected. This is required for organisms to grow and develop properly and for maintaining tissues of the body. Cancer cells, however, lose their ability to undergo apoptosis even though cells may be abnormal. The cancer cells may do this by altering the signals that detect the damage or abnormalities and activate apoptosis. They may also have defects in the downstream signaling itself, or the proteins involved in apoptosis<sup>104</sup>.

### **Limitless replicative potential**

Cells of the body have a limited number of divisions before the cells stop division. This is due to the DNA at the end of chromosomes, known as telomeres shortening with every cell division. Once it reaches a particular length, it activates senescence and the cell stops dividing. Cancer cells manipulate telomerases and maintain telomeres at a length above the critical length<sup>105</sup>. Thus, they divide indefinitely, without initiating senescence.

### **Sustained angiogenesis**

Normal tissues of the body require blood vessels running through them to deliver oxygen from the lungs. Cells must be close to the blood vessels to get enough oxygen for them to survive. An expanding tumor requires new blood vessels to deliver adequate oxygen to the cancer cells, and produces new vasculature by activating the angiogenic switch<sup>106</sup>.

In doing so, they control non-cancerous cells that are present in the tumor that can form blood vessels by reducing the production of factors that inhibit blood vessel production, and increasing the production of factors that promote blood vessel formation.

### **Tissue invasion and metastasis**

Cancer cells have the ability to invade neighboring tissues. Tissue invasion is the reason for their dissemination around the body. Cancer cells have to undergo a multitude of changes in order to acquire the ability to metastasize. It is a multistep process that starts with local invasion of the cells into the surrounding tissues, invasion of blood vessels, enter the circulatory system and exit in a different site and start dividing<sup>107</sup>.

### **MIGRATION AND INVASION IN CANCER**

Cancer metastasis is the cause of 90 percent of cancer related deaths<sup>108</sup>. It is the spread of cancer cells to tissues and organs beyond the primary site of tumor formation. Cancer cells can exploit their intrinsic migratory ability to invade nearby tissues and the vasculature, and ultimately metastasize. Metastasis is a multi-step process that includes dissociation-the process of a single tumor cell detaching from the primary tumor, invasion-the infiltration of the tumor cell to the stroma and invasion through the basement membrane, intravasion-the entry of the tumor cell into the vasculature and extravasation-the exit of the tumor cell from the vasculature and entry into remote organs<sup>108</sup>. Once the

tumor cell enters the secondary organ, it can proliferate resulting in clinically detectable tumor or remain dormant for years<sup>109</sup>. The cytoskeletal properties of the disseminating tumor cells play an important role in its ability to metastasize successfully.

Migration begins with polarization and extension of a leading front in the direction of the movement. Lamellipodia, filopodia, pseudopodia and invadopodia are different types of cell protrusions formed by migrating cells<sup>110</sup>. These protrusions are formed by actin polymerization. The leading front then binds to extracellular matrix proteins. This attachment occurs by means of cell adhesion molecules (CAMs) such as integrins and cadherins<sup>111</sup>. This binding causes the cell body to shrink which results in a traction force. This force allows the cell body to slide behind the migrating front resulting in cell movement. Actin-myosin contraction plays a major role in shrinking of the cell along the long axis. Organization of the actin-myosin skeleton is controlled by different enzymes. Myosin Light Chain kinase (MLCK) and Myosin Light Chain phosphatase (MLCP) are two enzymes that act on the light chains of myosin. The activity of these enzymes, in turn, is regulated by another set of enzymes, the Rho-GTPases. This group includes several members like Rho, Rac and CDC42 proteins<sup>112</sup>.

As opposed to cell migration, invasion by tumor cells require the cells to degrade the extracellular matrix. Matrix Metalloproteases are proteases that play an important role in degrading the basement membrane and extra cellular matrix. Cytokines and growth factors regulate the expression of MMPs through the mitogen-activated protein kinases

pathway that includes proteins such as ERK 1/2 (extracellular regulated kinase 1/2), JNK/SAPK1/2 (c-Jun N-terminal kinase 1/2) and p38MAPK<sup>113,114</sup>.

### **EPITHELIAL-MESENCHYMAL TRANSITION (EMT)**

Epithelial-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 production of ECM components<sup>115,116</sup>.

EMT is under tight control of multiple regulatory pathways. Transforming growth factor  $\beta$  (TGF- $\beta$ ) activity is usually enhanced during EMT<sup>117</sup>. Binding of TGF- $\beta$  to its cell surface receptors activates Smad family of transcription factors. These smad proteins translocate to the nucleus and cooperate with transcription factors from the Snail and Twist family<sup>118</sup>. TGF- $\beta$  can also activate Rho-like GTPases<sup>119</sup>. These effectors repress genes that are involved in cell polarity and cell-cell adhesion, such as RhoA and E-cadherin. At the same time, the expression of N-cadherin, another member of the cadherin family that allows for enhanced adhesion between mesenchymal cells, is upregulated. This change in cadherin expression is called cadherin switch and is a hallmark of EMT<sup>120</sup>.

## **CHAPTER 2**

### **MATERIALS AND METHODS**

#### **Cells and culture conditions**

Hep3B cells were obtained from the American Type Culture Collection (Manassas, VA); QGY-7703 cells were obtained from Fudan University, China; HepG3 and Huh7 cells were kindly provided by Dr. Paul Dent. QGY-7703, HepG3 and Huh7 cells were cultured in Dulbecco's Modified Eagle Media (DMEM). Hep3B cells were maintained in Minimum Essential Media (MEM) alpha. RPMI 1640 media was used for culturing SNU-182 and SNU-449. SK Hep-1 and PLC/PRF/5 cells were cultured in Eagle's Minimum Essential Media (EMEM). For generation of stable clones, Huh 7 and QGY-7703 cells transfected with luciferase gene (QGY-7703 luc cells) were used. All the above media were supplemented with 5% FBS and 0.5% penstrep.

#### **Generation of stable clones**

Huh7 and QGY-7703 luc cells were transfected with TAF2 shRNA (sc-77487-SH) or control shRNA (sc-108060). Transfection was performed using lipofectamine 2000 (Invitrogen). 15  $\mu$ l of lipofectamine reagent was diluted in 500  $\mu$ l of Opti-MEM I reduced serum medium. 8  $\mu$ g of TAF2 or control shRNA was diluted in 500  $\mu$ l of OPTI-MEM I reduced serum medium. They were incubated for 5 minutes. Diluted DNA was added to



diluted lipofectamine 2000 reagent (1:1 ratio). The mixture was incubated for 20 minutes and added to a 6cm dish containing cells at 70 % confluency. After 48 hours, single cell suspension was plated. Individual colonies were selected, expanded and maintained in media containing 1 µg/ml puromycin.

### **Western blotting**

Cells were washed twice with cold PBS and lysed on ice for 30 min in 250 µl of cold 1.5% n-dodecyl -D-maltoside (DDM). Cell debris were removed by centrifugation at 15,000 rpm for 15 min at 4°C. Protein concentrations were determined using the Bio-Rad protein assay system. Aliquots of cell extracts containing 30 µg of total protein were resolved in 10% SDS-PAGE and transferred to nitrocellulose blotting membranes. Membranes were blocked for 1 h at room temperature in Blocking buffer (5% nonfat milk powder in TBST: 10 mM TRIS-HCL (pH 8.0), 150 mM NaCl, 0.05% Tween 20), and then incubated overnight at 4 degree Celsius in Blocking buffer containing the respective primary antibody. After washing in TBST buffer thrice (10 min, RT), membranes were incubated for 1 h at room temperature in the respective secondary antibody diluted in suitable buffers. After washing in TBST, Enhanced Chemiluminescence(ECL) detection reagents were used to detect the protein of interest.

### **q-RT-PCR**

Cells were lysed with 700 µl of QIAzol lysis reagent and homogenized by vortexing. 140 µl chloroform was added to separate RNA from DNA and proteins. Centrifugation at 15,000 rpm for 15 minutes at 4°C was done to separate the solution into three phases. The RNA in the upper aqueous phase was transferred to a new collection tube and 1.5 volumes of 100% ethanol was added to precipitate the RNA out of the solution. The solution was pipetted into an RNeasy Mini column and centrifuged. The filter was then washed with wash buffers RWT and RPE (twice). The membrane was further dried by centrifuging at full speed for 1 minute. RNA was eluted using 50 µl RNase free water. Concentration of RNA was measured using the bio rad system. cDNA was synthesised from the extracted RNA by making a master mix composed of 2 µg RNA, 2 µl 10X buffer, 2 µl random primers, 6.4 µl dNTP, 1 µl RNase out and 1 µl reverse transcriptase. The volume was made upto 20 µl using RNase free water. cDNA synthesis was performed using PCR. The cDNA was then probed with TAF2 or GAPDH probe.

### **Colony formation assay**

500 cells were plated in a 6cm dish. After two weeks, the cells were fixed (with 3.7% formaldehyde) and stained (with 25% giemsa). The number of colonies (>50 cells) were counted. Colony forming ability was calculated as

$(\text{Number of colonies counted} / \text{Number of cells plated}) * 100$

### **MTT assay**

1000 cells were plated in a each well of a 96 well plate. At the desired time point, (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (MTT) reagent was added to the media. After 5 hours, equal volume of sodium dodecyl sulphate (SDS) was added. The plates were incubated at 37°C overnight. Absorbance was measured at 600 nm. Readings were obtained at 24, 48, 72 and 96 hours.

### **Migration assay**

A wound healing/scratch assay was used to evaluate cell migration. Confluent cells were scratched with a 200 µl micropipette tip to create a cell-free area. The scratched monolayers images were captured at 0, 24 and 48 hours after wounding. A mark was made on the plates to make sure that the same area is being imaged everyday. The wound width was measured at the specific time points and compared.

### **Invasion assay**

Invasion was measured by using 24-well BioCoat cell culture inserts with an 8-µ-porosity polyethyleneterephthalate membrane coated with Matrigel basement membrane matrix (100 µg/cm<sup>2</sup>). Briefly, the Matrigel was allowed to rehydrate for 2 h at 37 °C. The wells of the lower chamber were filled with medium containing 5% FBS. 25,000 cells were seeded in the upper compartment (6.25-mm membrane size) in serum-free medium. The invasion

assay was performed at 37 °C in a 5% CO<sub>2</sub> humidified incubator for 22 hours. At the end of the invasion assay, the filters were removed, fixed (with methanol), and stained (with 10% crystal violet). Cells on the upper surface of the filters were removed by wiping with a cotton swab, and invasion was determined by counting the number of cells that migrate to the lower side of the filter with a microscope at 100X magnification.

### **IMMUNOFLUORESCENCE AND IMMUNOHISTOCHEMISTRY**

Normal human liver and human HCC liver were fixed on a slide for immunohistochemistry. For IF, primary hepatocytes, QGY-7703 cells and SNU-182 cells were plated on a 4-chamber slide. The slides were blocked and incubated overnight in diluted TAF2 antibody. They were rinsed the following day and treated with fluorochrome conjugated secondary antibody. The amount of fluorescence was detected by confocal laser scanning microscopy (LSM700). For IHC, secondary antibodies bound to horseradish peroxidase was used and peroxidase was used to obtain staining. The distribution of the stain was observed under a microscope.

### **BIOINFORMATICS**

Pan-cancer analysis of gene expression data in multiple databases was done to determine the genes in 8q that correlate with AEG-1 expression in 38 different cancer types.

Liver cancer samples (LIHC, n = 371) from TCGA database were analyzed to determine the genomic alterations in these patients in the TAF2 gene. The “R2: Genomics Analysis and Visualization Platform” was used to determine the correlation between expression levels of AEG-1/MTDH and TAF2 genes in two independent datasets. Oncomine was used to compare TAF2 copy number differences between HCC samples and normal liver. Two independent datasets, TCGA and Guichard, were used for this analysis. RNA expression levels in three independent datasets (Roessler Liver 2, Wurmbach Liver, Chen Liver) were also compared using Oncomine.

To understand the relationship between TAF2 expression levels and patient survival, a Kaplan-Meier plot was obtained. Data from TCGA HCC database was used for this analysis.

TCGA HCC (LIHC, n = 371) dataset was divided into two groups: TAF2 high (top 25%; n = 93) and TAF2 low (bottom 25%; n = 93). R program was used to identify differentially expressed genes between patients expressing high and low TAF2. A heat map of top 50 differentially modulated genes was also obtained.

Canonical pathway analysis by ingenuity software was used to identify genes that are significantly upregulated and downregulated. These differentially changed genes were analysed using Ingenuity pathway analysis software to identify the upstream regulators that might lead to alterations in downstream genes. A z-score > 2 indicates activation and a z-score < -2 indicates inhibition.

## **STATISTICAL ANALYSIS**

All results were checked for statistical significance using the one-way ANOVA test.

## CHAPTER 3

### RESULTS

#### **TAF2 overexpression positively correlates with AEG-1/MTDH over expression**

Pan-cancer analysis of gene expression data in multiple databases which identified TATA-box binding protein associated factor 2 (TAF2; 8q24) as the most frequently co-expressed gene with AEG-1 in 36 out of 38 different cancer types (Pearson's correlation coefficient  $r > 0.5$ ;  $p < 3E-14$ ). Analysis of liver cancer samples (LIHC,  $n = 371$ ) from TCGA database confirmed our observations that TAF2 is co-expressed with AEG-1/MTDH ( $r = 0.569$ ,  $p < 1E-16$ ). Independent analysis using the "R2: Genomics Analysis and Visualization Platform" further confirmed strong positive correlation between expression levels of AEG-1/MTDH and TAF2 genes in two independent datasets (Fig. 2.1).

#### **TAF2 gene is amplified and over expressed in human HCC patients**

Analysis of TCGA database showed that TAF2 is amplified in approximately 20% of HCC patients (Fig. 2.2a). In OncoPrint, two independent datasets (TCGA and Guichard) on HCC showed TAF2 copy number gain in HCC samples compared to normal liver (Fig. 2.2b). At mRNA expression level three independent datasets in OncoPrint showed highly significant increase in TAF2 mRNA levels in HCC samples versus normal liver (Table 1).

### **Patient survival decreases with increasing TAF2 expression**

We then used a Kaplan-Meier plot to determine the relationship between TAF2 expression and patient survival. A significantly strong negative correlation ( $p = 0.0041$ ) was observed between TAF2 expression levels and HCC (stages T2-T4) patient survival in TCGA database (Fig. 2.3).

### **Genes involved in several oncogenic signaling pathways are upregulated in HCC patients with increased TAF2**

TCGA HCC dataset was divided into two groups-TAF2<sup>high</sup> and TAF2<sup>low</sup>. TAF2 showed log<sub>2</sub>fold change of 2.27 in TAF2<sup>high</sup> vs TAF2<sup>low</sup> ( $p = 4.77E-65$ ). A total of 8166 differentially expressed genes were identified at false discovery rate (FDR) corrected p-value of 0.01. 5327 genes were upregulated while 2839 were downregulated. A heat map of top 50 differentially modulated genes shows tight clustering (Fig. 2.4a). Canonical pathway analysis by ingenuity software identified significant upregulation of genes of several oncogenic signaling pathways (such as ERK5, IGF-1 and HGF signaling) and marked downregulation of genes in EIF signaling pathway and oxidative phosphorylation resulting in mitochondrial dysfunction in TAF2 high group (Fig. 2.4b). Molecules involved in oncogenic signaling pathways (CD24, MAP4K4, HGF, VEGF) and several miRNAs were found to be possible upstream regulators of the differentially expressed genes. Activation



of cyclin D1 (CCND1) and inhibition of RB1 were identified to cluster with TAF2 high further indicating a role of TAF2 in regulating cell cycle.

### **TAF2 is overexpressed in HCC**

Western blotting was done to detect TAF2 in primary hepatocytes and different HCC cell lines. It was found TAF2 present in primary hepatocytes was undetectable on a western blot. On the other hand, all the HCC lines that were used for analysis showed a much higher level of TAF2 protein expression (Fig 2.5b). Immunofluorescence assay was done to compare TAF2 expression levels between primary hepatocytes and HCC cell lines and showed similar results as western blotting (Fig 2.5a). Liver sections from normal subjects and HCC patients were used to compare the levels of TAF2 expression in vivo. It was found that normal liver section had much lesser staining of TAF2 compared to HCC liver (Fig 2.5c).

### **Knockdown of TAF2 does not affect proliferation**

TAF2 was knocked down using shRNA. The knockdown was confirmed using western blotting and q-RT-PCR. Two positive clones from Huh 7 cells showed a knockdown of 60 percent as compared to untreated cells (Fig 2.6a,c). The positive clone from QGY Luc 7703 cells showed a knockdown of 40 percent (Fig 2.6b,d). The clones were named shTAF2-1 and shTAF2-2 and shTAF2-3 and used for further assays. The cells treated

with control shRNA did not show any difference in TAF2 expression compared to untreated cells and were named shCon-1 (Huh7 cells) and shCon-2 (QGY Luc cells).

MTT and colony formation assays were done to determine the effect of the knockdown on proliferation. For MTT assay, the absorbance of the shCon and shTAF2 cells were measured at 24, 48, 72 and 96 hours and compared. There was no significant difference in proliferation between the cells (Fig 2.7 a,b).

For colony formation assay, the number of colonies were counted after 3 weeks for Huh7 cells and 2 weeks for QGY-7703 Luc cells. Representative plates after fixing and staining are shown in fig 2.8 c. The shCon and shTAF2 cells did not show any difference in their ability to form colonies (Fig 2.8 a,b).

### **Knockdown of TAF2 inhibits migration**

Wound healing assay was done to determine if there was any difference in ability to migrate between the cells. At 24 hours, the wound distance in both shTAF2-1 and shTAF2-2 cells was higher compared to shCon-1 cells ( $p=0.0009$  and  $p=0.0018$ ). At 48 hours, the wound distance remained higher in shTAF2-1 and shTAF2-2 cells compared to shCon cells ( $p=0.000385$  and  $p=0.028501$ ) (Fig 2.9 c). shTAF2-3 cells also showed a higher wound distance compared to shCon-2 cells at 24 and 48 hours ( $p=0.021$  and  $p=0.041$ ) (Fig 2.9 d).

### **Knockdown of TAF2 inhibits invasion**

Matrigel invasion assay was done to determine any difference between shCon and shTAF2 cells in their ability to invade the Matrigel. After fixing and staining, more cells per field were observed in the shCon cells compared to shTAF2 cells (Fig 2.10 a). Upon quantification, there was significant decrease in the number of shTAF2-1 ( $p=0.015727$ ) and shTAF2-2 ( $p=0.002568$ ) cells that had invaded the Matrigel compared to shCon-1 cells. This result was also observed in shTAF2-3 cells ( $p=0.027155$ ).

### **Knockdown of TAF2 does not affect cyclin A or cyclin B1 expression**

To find out the reason for not observing a change in cellular proliferation in the TAF2 knockdowns, the knockdowns and control were probed for cyclins A and B1. shTAF2-1 and shTAF2-2 cells did not show any difference in cyclin levels as compared to shCon-1 cells.

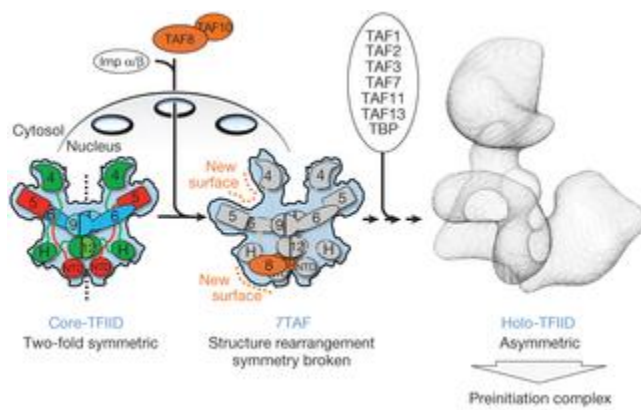
### **Knockdown of TAF2 increases E-Cadherin and decrease Snail protein expression**

The knockdowns and control were probed for EMT markers: E-Cadherin and N-Cadherin. An increase in E-Cadherin was observed in the shTAF2-1 and shTAF2-2 cells compared to shCon-1 cells. There was no change in N-Cadherin levels. The knockdowns were probed for known regulators of E-Cadherin. Snail, which is a known negative regulator of

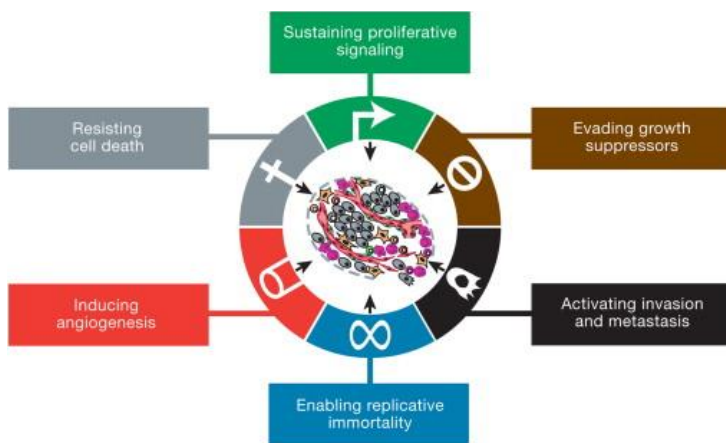
E-Cadherin showed decrease in shTAF2-1 and shTAF2-2 compared to shCon-1 cells.

## CHAPTER 4

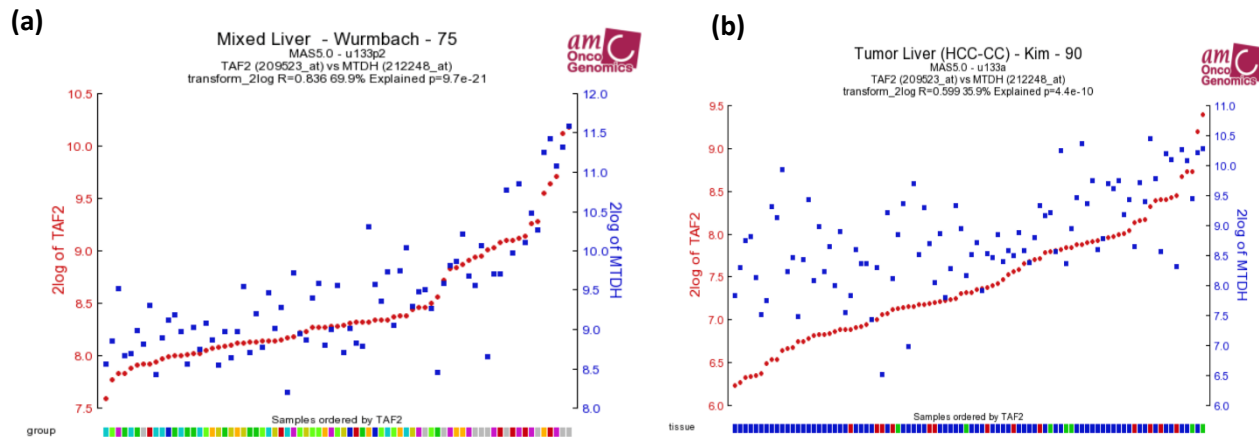
### FIGURES



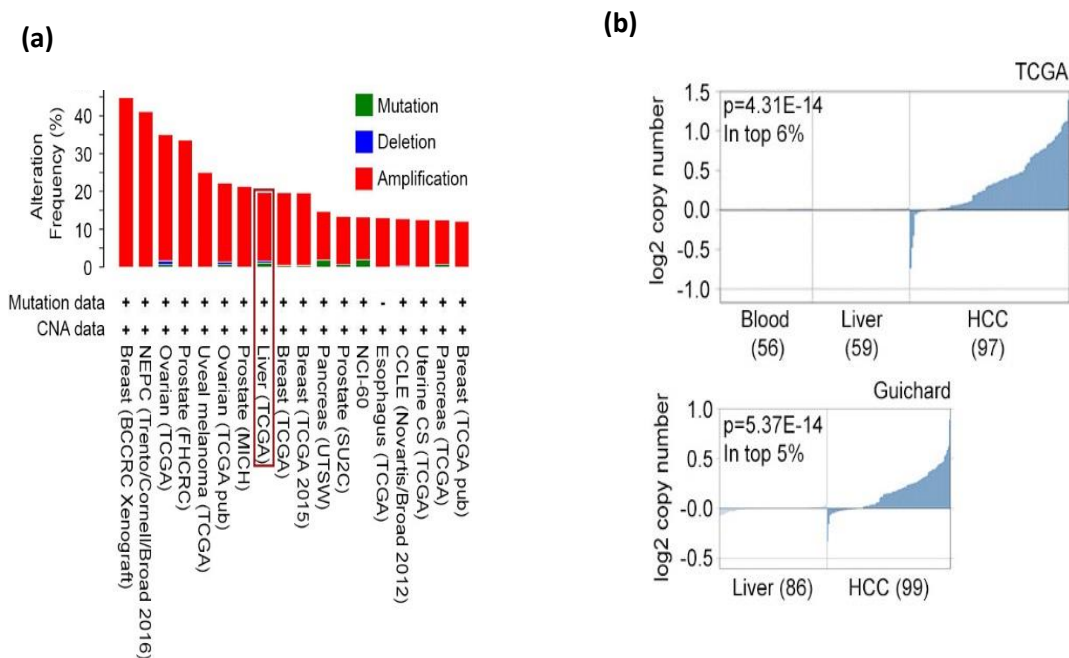
**Figure 1.1:** Model for holo-TFIID assembly



**Figure 1.2:** Hallmarks of cancer



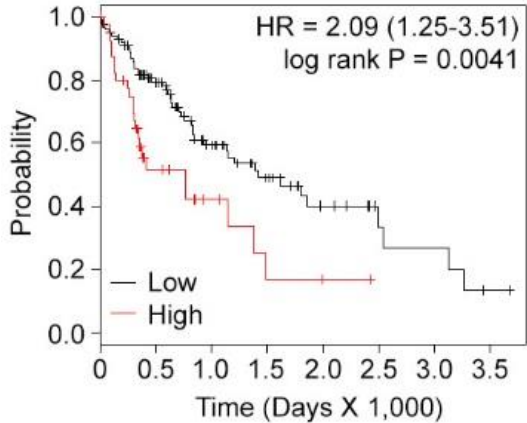
**Figure 2.1:** Positive correlation between TAF2 and AEG-1 expression in (a) tumor liver and (b) mixed liver sample



**Figure 2.2:** TAF2 is amplified and exhibits copy number gain in HCC patients  
 (a) Analysis of TCGA database shows the alteration frequency of TAF2 in liver cancer patients.  
 (b) Analysis of TAF2 copy number alteration in HCC patients in TCGA and Guichard databases shows increased TAF2 copy number in HCC samples

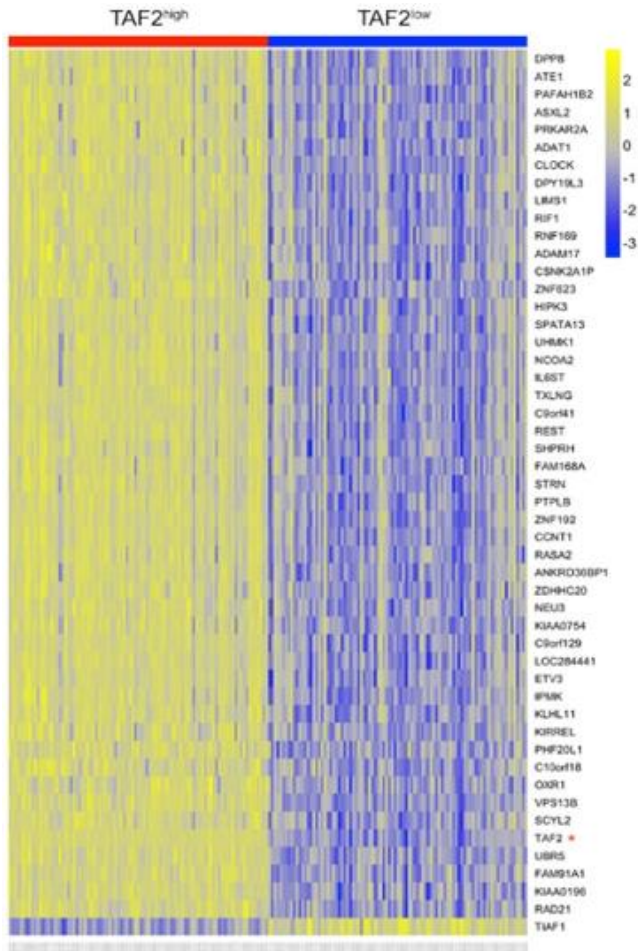
Database	Liver(n)	HCC(n)	Log2 fold	p-value	Rank
Roessler Liver 2	220	225	1.954	3.88E-55	In top 3%
Wurmbach Liver	10	35	1.866	3.57E-7	In top 2%
Chen Liver	75	104	1.457	4.33E-9	In top 7%

**Table 1:** TAF2 mRNA expression in HCC patients and normal subjects in three independent datasets

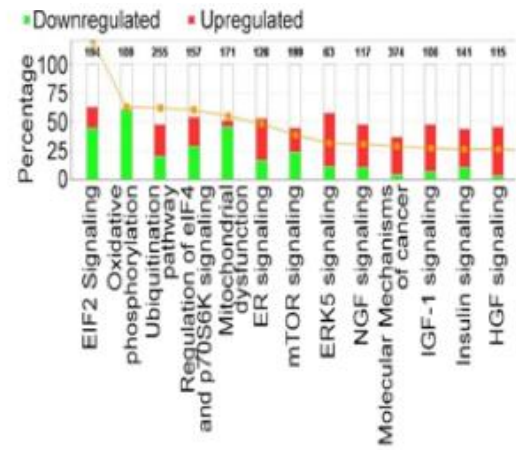


**Figure 2.3:** A Kaplan-Meier plot shows decreasing survival of HCC patients with increasing TAF2 expression

(a)



(b)

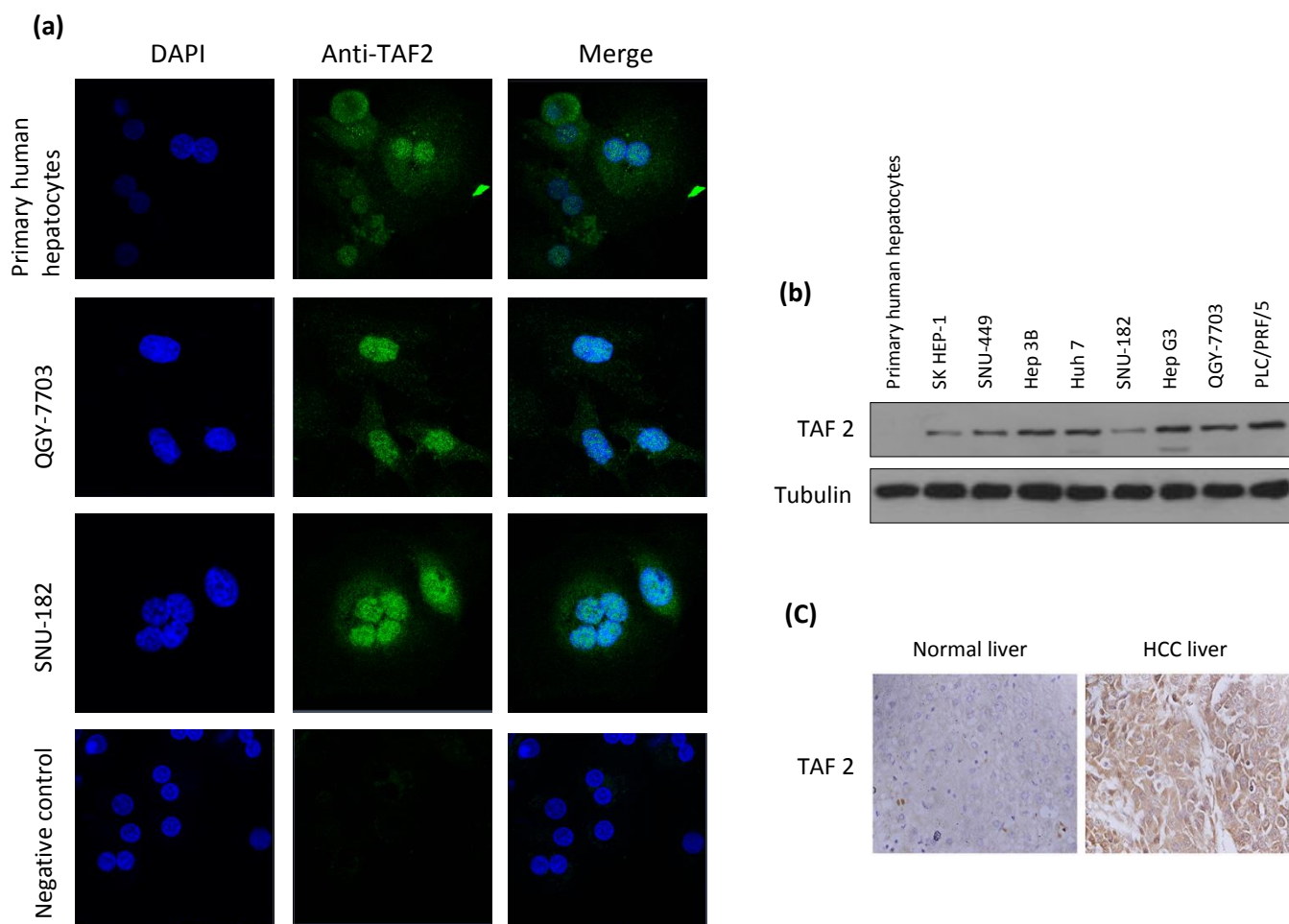


(c)

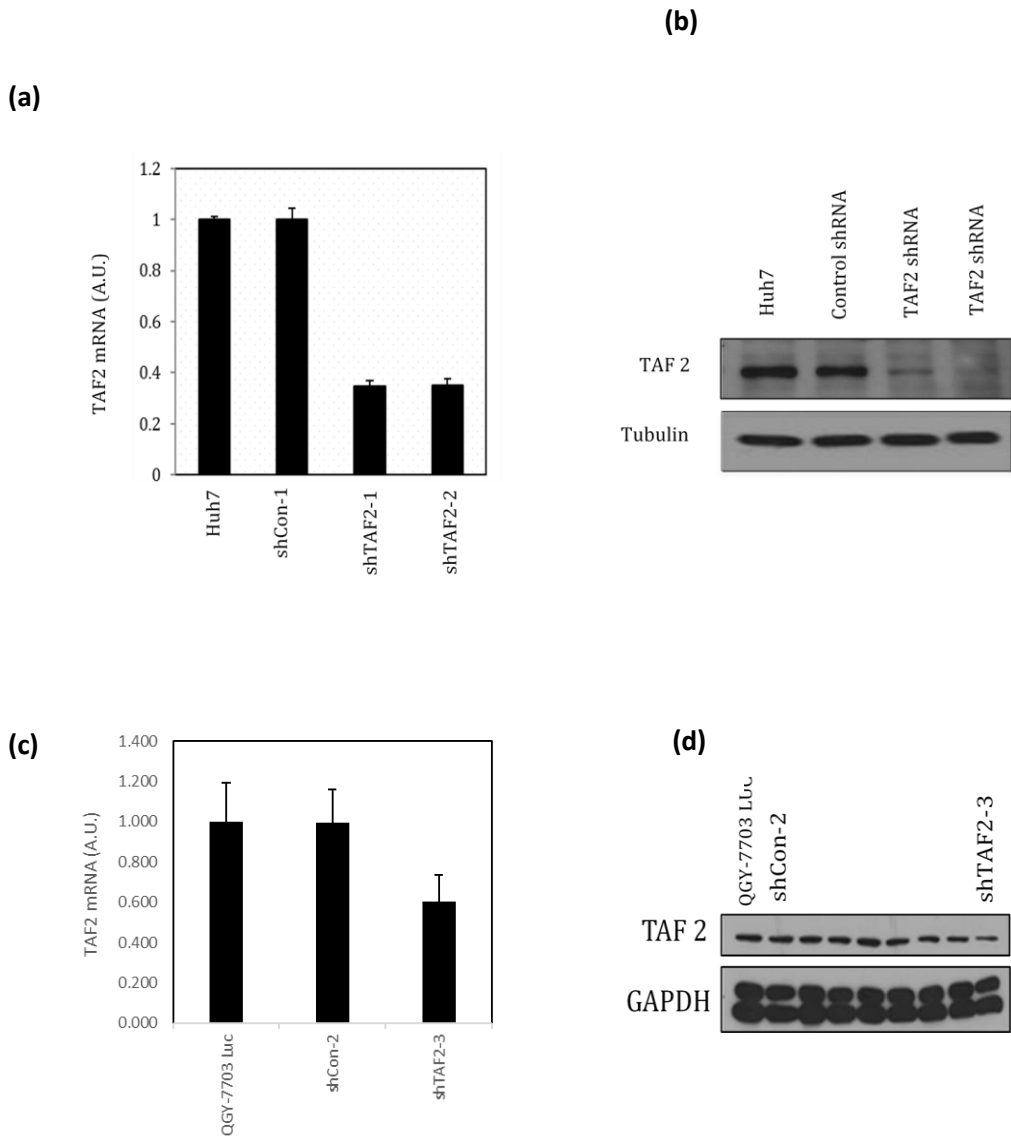
Upstream regulator	Z-score	p-value
RICTOR	9.87	1.66E-27
ER	9.74	8.33E-16
CD24	6.325	8.74E-07
MAP4K4	2.429	3.09E-06
TCF7L2	7.963	3.17E-06
ERG	4.555	7.78E-06
KDM5A	5.632	8.27E-06
HOKA9	6.522	1.02E-05
HGF	9.74	2.79E-05
SYVN1	3.636	2.91E-05
GNA12	2.984	6.88E-05
FLT1	3.413	9.61E-05
CCND1	3.325	1.45E-04
VEGF	8.64	1.79E-04
NUPR1	-6.633	1.83E-09
KLF3	-5.009	3.58E-09
miR-155-5p	-5.907	4.04E-08
EIF4E	-2.584	4.34E-08
miR-1-3p	-6.085	6.20E-08
miR-16-5p	-4.657	1.27E-07
RB1	-5.338	3.35E-06
miR-124-3p	-6.055	5.23E-06
Let-7a-5p	-2.336	9.95E-06
miR-8	-4.183	1.20E-05

**Figure 2.4:** (a) A heat map showing top 50 differentially expressed (DE) genes in TAF2<sup>high</sup> and TAF2<sup>low</sup> HCC patients. Canonical pathway analysis by ingenuity software showing genes that are (b) upregulated and downregulated in TAF2<sup>high</sup> group and (c) potential upstream regulators of the DE genes





**Figure 2.5:** TAF2 is over expressed in HCC cell lines and HCC liver (a) Representative fluorescent confocal micrographs showing TAF 2 protein expression in primary human hepatocytes and different HCC cell lines. (b) A western blot showing the levels of TAF 2 protein expression in primary human hepatocytes and different HCC cell lines.



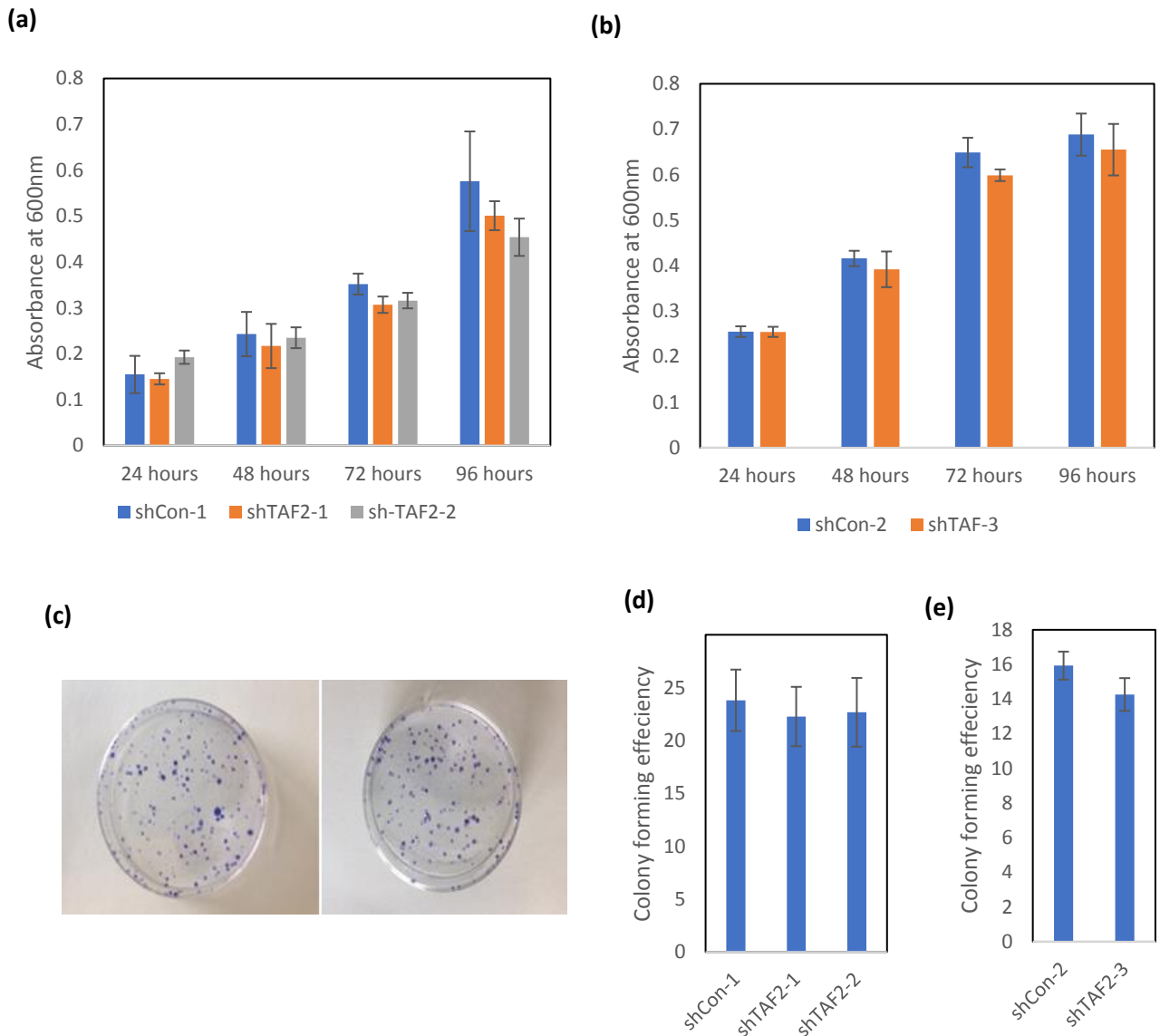
**Figure 2.6:** TAF2 is knocked down at mRNA and protein levels in cells treated with shRNA targeting TAF2 mRNA.

(a) A q-RT-PCR showing the levels of TAF2 mRNA expression in untreated Huh7 cells, Huh7 cells treated with control shRNA and Huh 7cells treated with TAF2 shRNA.

(b) A western blot showing the levels of TAF2 protein expression in untreated Huh7 cells, Huh7 cells treated with control shRNA and Huh 7 cells treated with TAF2 shRNA.

(c) A q-RT-PCR showing the levels of TAF2 mRNA expression in untreated QGY-7703 Luc cells, QGY-7703 Luc cells treated with control shRNA and QGY-7703 Luc cells treated with TAF2 shRNA.

(d) A western blot showing the levels of TAF2 protein expression in untreated QGY-7703 cells, QGY-7703 cells treated with control shRNA and QGY-7703 cells treated with TAF2 shRNA.

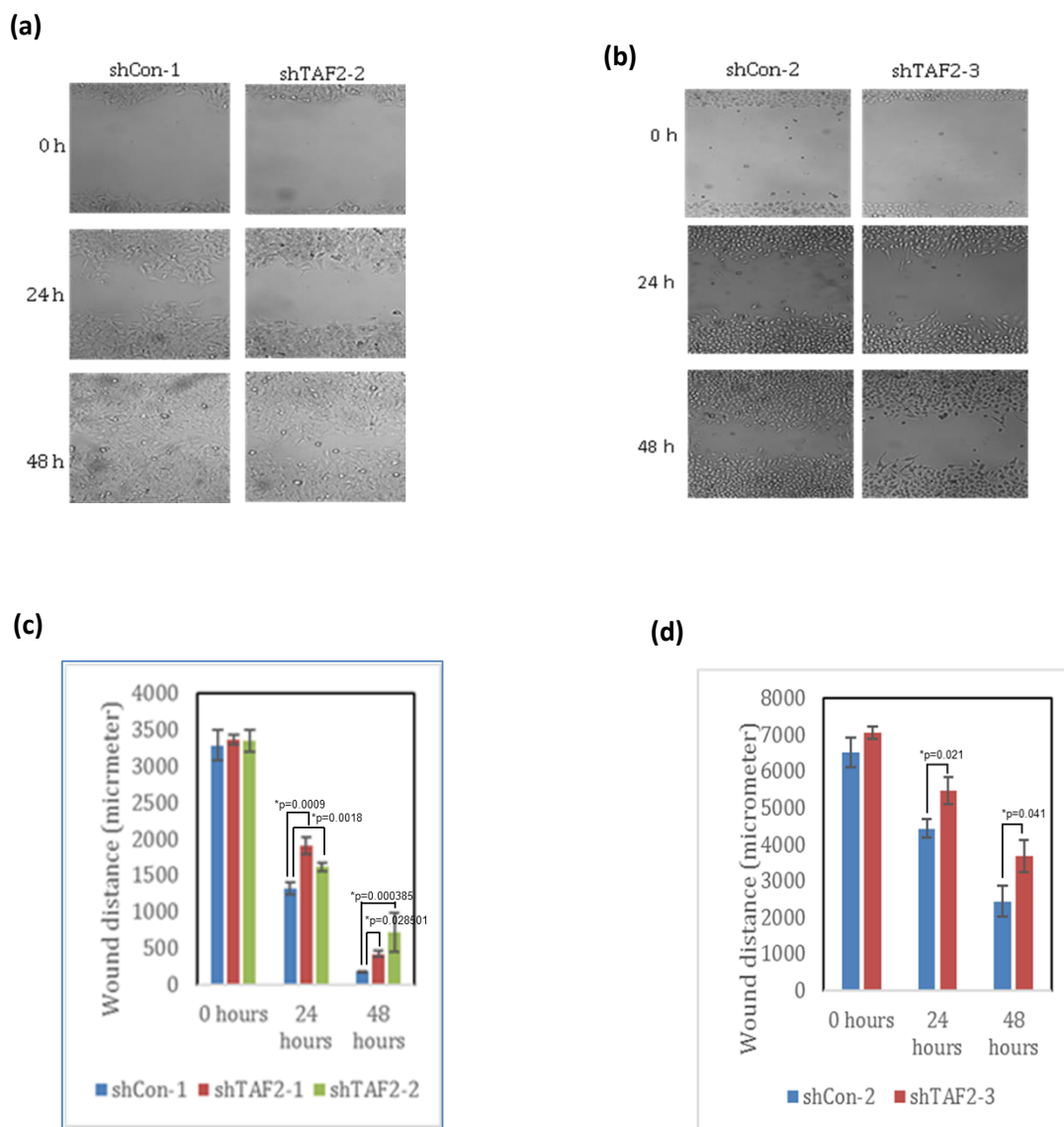


**Figure 2.7:** TAF2 knockdown does not affect cellular proliferation.

A bar graph showing absorbance at 600 nm for (a) shCon-1, shTAF2-1 and shTAF2-2 and (b) shCon-2 and shTAF2-3 cells at four time points (24, 48, 72, 96 hours).

(c) Representative colony formation plates (shCon-1 and shTAF2-2) after fixation and staining.

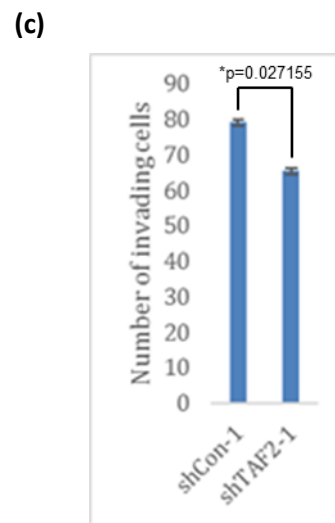
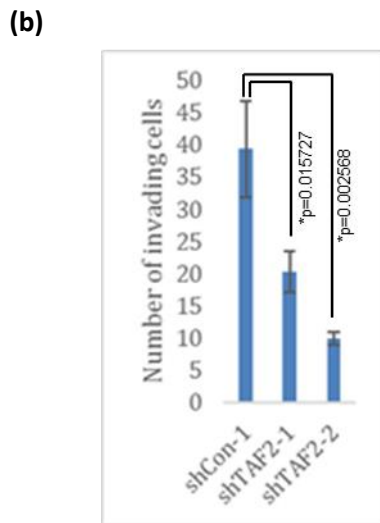
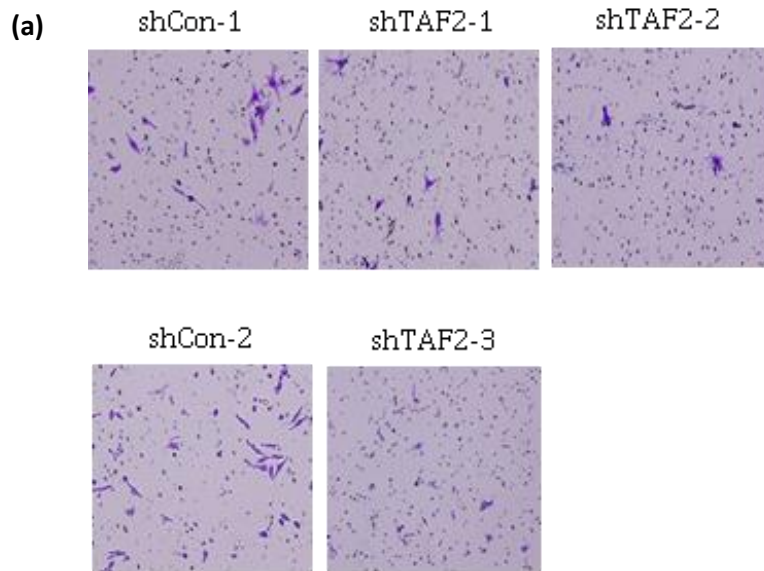
A bar graph showing colony forming efficiency of (d) shCon-1, shTAF2-1 and shTAF2-2 and (e) shCon-2 and shTAF2-3 cells.



**Figure 2.8:** Stable TAF2 knockdown results in decreased migration

Representative wound healing in (a) shCon-1 and shTAF2-2 and (b) shCon-2 and shTAF2-3 cells after 24 and 48 hours

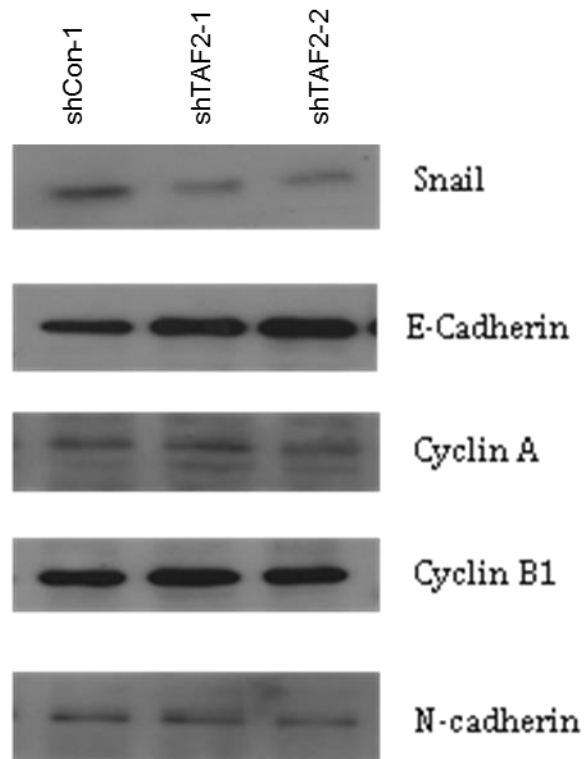
(c) A bar graph representing wound distance of (a) shCon-1, shTAF2-1 and shTAF2-2 and (b) shCon-2 and shTAF2-3 cells after 24 and 48 hours



**Figure 2.9:** Stable TAF2 knockdown results in decreased invasion.

(a) Pictomicrograph of invaded cells in shCon-1, shTAF2-1, shTAF2-2, shCon-2 and shTAF2-3.

(b) A bar graph showing number of invading cells in (b) shCon-1, shTAF2-1 and shTAF2-2 and (c) shCon-2 and shTAF2-3 cells.



**Figure 2.10:** A western blot showing E-Cadherin, N-Cadherin, Snail, cyclin A and cyclin B1 protein levels in shCon-1, shTAF2-1 and shTAF2-2 cells

## CHAPTER 5

### DISCUSSION AND FUTURE DIRECTION

AEG-1 is a well-established oncogene which is overexpressed in many common cancers. Its expression level negatively correlates with poor survival and overall adverse prognosis. A pan cancer analysis of gene expression data in multiple databases identified TATA-box binding protein associated factor 2 (TAF2) as the gene that is most frequently co-expressed with AEG1. Human AEG1 gene is present in the chromosomal location 8q22.1 and TAF2 is present in the chromosomal location 8q24.1. The 8q22-24 is a region that is amplified in many cancers. The fact that these two genes are overexpressed to the same level in cancer and present in the chromosomal region which is a mutational hotspot for amplifications indicates that these two genes could possibly be co-amplified during cancer.

Bioinformatic analyses have showed that levels of TAF2 overexpression positively correlates with AEG1 overexpression levels in different HCC datasets. These two genes could be working in cooperation with one another or have the same regulatory mechanism. It is known that AEG-1 is transcriptionally regulated by c-Myc, an oncogene frequently upregulated in HCC. It is possible that TAF2 is also a downstream target of c-Myc. Data analysis from TCGA database shows that TAF2 is altered in 20% of liver

cancer patients. As expected, majority of these alterations are amplifications. Comparison of copy number differences of the TAF2 gene between HCC samples and normal liver and blood samples in two independent datasets showed a significant increase in the copy number of the TAF2 gene in HCC samples. To determine if this increase in copy number results in mRNA overexpression, mRNA expression levels were compared between normal and HCC livers in three independent datasets. We found a significant increase in TAF2 mRNA levels in HCC liver compared to normal liver. TAF2 was also ranked high (top 2%, 3% and 7%) among the mRNAs that were differentially expressed between these samples. It was also observed that TAF2 mRNA overexpression was not observed in other liver abnormalities like liver injury or cirrhosis. This leads us to the conclusion that TAF2 overexpression is HCC specific and could be a driver event in HCC. Analysis of a survival plot generated from HCC dataset in TCGA showed negative correlation between TAF2 expression and overall survival of HCC patients.

The TCGA dataset was grouped into TAF2<sup>high</sup> and TAF2<sup>low</sup> and a heatmap showing the top 50 genes that are differentially regulated between the two groups was generated. This map showed that most of the genes are overexpressed in the TAF2<sup>high</sup> group as compared to the TAF2<sup>low</sup> group. Canonical pathway analysis by Ingenuity software showed the pathways that are affected by these differentially expressed genes. The analysis showed that several pro-oncogenic signaling pathways were upregulated (HGF, IGF-1 and NGF and ERK-5). Oxidative phosphorylation is a process that is downregulated in cancers since cancer cells are in a state of hypoxia and unable



to perform oxidative phosphorylation. Instead, they use glycolysis as their preferred source of energy. eIF2 is a signaling pathway that is required for successful translation of proteins. Approximately 50 percent of the genes involved in both of these pathways were found to be downregulated in TAF2<sup>high</sup> group. The possible upstream regulators of these differentially regulated genes were determined. Some of the notable activators of these genes include molecules involved in oncogenic pathways like CD24, MAP4K4, HGF, CCND1 and VEGF. Inhibition of the tumor suppressor gene RB1 and several miRNAs were identified to cluster with TAF2<sup>high</sup> group.

TAF2 is a protein that is involved in transcription of genes by RNA polymerase II. It is a factor that is dispensable for basal transcription but, required for activated transcription. The overexpression of TAF2 in HCC patients was confirmed *in silico*, *in vitro* (in cell lines) and *in vivo* (in liver sections). Stable cell lines with knockdown of TAF2 were created to check for the effect of the knockdown on various cancer hallmarks like proliferation, migration and invasion. Two different assays (MTT and colony formation assays) were performed to compare the proliferation capacity of TAF2 knockdown cell lines and control cell lines. TAF2 knockdown cell lines did not show a decrease in cellular proliferation. On performing a wound healing assay to compare the migration capacity of TAF2 knockdown and control cell lines, the knockdowns showed a delay in wound healing. A Matrigel invasion assay was then performed to compare the invasion capacity of TAF2 knockdown

and control cells. TAF2 knockdown cells showed a decreased ability to invade through the Matrigel. Thus, knockdown of TAF2 did not affect proliferation but, inhibited migration and invasion.

An earlier study found TAF2 to be involved in regulating cyclin levels (cyclin A and B1) and hence cell cycle progression. It was shown that transient knockout of TAF2 in HeLa cells resulted in a decrease in cyclin A and B1 levels and cell cycle arrest in G2/M phases. Also, cyclin D1 was one of the possible downstream targets of TAF2 identified through canonical pathway analysis that was performed as a part of this study. Given the above, it was contradicting that a decrease in cellular proliferation was not seen in the knockdowns. To explain this, lysates from the knockdowns were probed for cyclins A and B1. A decrease was not observed in cyclin A or cyclin B1 levels in the knockdowns. This gives a possible explanation for not seeing a decrease in cellular proliferation in the knockdowns. But, the reason for not seeing a decrease in cyclin levels in the knockdowns is yet to be determined. A possible explanation for this is that the stable cell lines used in this study are TAF2 knockdown cell lines and not TAF2 knockout cells lines. These cells have 40 to 60 percent TAF2 in them. This remaining TAF2 could be sufficient to keep cyclin levels normal and help cell cycle progress without any hindrance.

Epithelial Mesenchymal Transition (EMT) plays a key role in a tumor cell obtaining the

ability to migrate and invade the basement membrane. Inhibition of migration and invasion was observed in the knockdowns. To determine if there is a reversal of EMT transition occurring in the knockdowns, the cell lysates were probed for EMT markers E-Cadherin and N-Cadherin. An increase in E-Cadherin was observed. Unlike typical cadherin switching, this was not accompanied by a decrease in N-Cadherin. To find the molecule causing an increase in E-Cadherin levels, the lysates were probed for known regulators of E-cadherin. Snail, which is a known negative regulator of E-Cadherin was found to decrease in the knockdowns. Based on our observations, we can conclude that knockdown of TAF2 decreases Snail levels which causes increase in E-Cadherin thus making the cells more epithelial and hence unable to migrate and invade.

The phenotypic difference observed in the knockdowns is due to a decrease in Snail protein level. But, it is yet to be understood how TAF2 causes this decrease in Snail protein. The knockdowns need to be probed for known regulators of Snail like smad proteins. Since no decrease in the mesenchymal marker N-Cadherin is seen, it will be worthwhile probing for other known mesenchymal markers (Vimentin, Fibronectin, Type I collagen, Laminin 5 and certain integrins). There are other known EMT regulators (Slug, Snail, Twist, Zeb1 and Zeb2) that can be probed too. This will help reach a conclusion about the pathway that TAF2 is involved in to promote EMT. Given that we have obtained a difference in migration and invasion capacity of the cells, staining these cells will

determine if there is a difference in actin organization.

Also, bioinformatic data obtained from canonical pathway analysis performed during the beginning of this study helps explain this difference in migration and invasion. mTOR signaling, which is involved in tumor cell motility, invasion and metastasis<sup>121</sup>, was found to be deregulated in TAF2<sup>high</sup> group. Also, some of the possible downstream regulators of TAF2 included Rictor which is a component of the mTORC2 signaling complex and regulates cell migration<sup>122</sup>, CD24 which is a cell adhesion molecule, MAP4K4 which controls cell motility<sup>123</sup> HGF or scatter factor which is a cell motility inducer<sup>124</sup> and VEGF which is a cell migration inducer<sup>125</sup>. It is necessary to study if the observed decrease in invasion and migration is due to any of these factors.

With evidence from previous studies for involvement of TAF2 in cell cycle regulation and maintaining cyclin levels, it is important to determine if we did not see a difference in cellular proliferation due to the remaining TAF2 in the cells or because TAF2 does not regulate cyclins. It is therefore necessary to create a stable clone with complete knock out of TAF2 and study if a change in cell proliferation, migration and invasion is obtained. Another future direction of this project would be to create stable clones that over express TAF2 and study its effect on cellular proliferation, migration and invasion. Performing a Differential Expression (DE) analysis using RNA-seq to determine the difference in gene

expression between the knockout and overexpression cell lines will be help gain an insight into the genes that are affected by TAF2.

In summary, we have shown that TAF2 is important for tumor migration and invasion. It is a potential oncogene that needs to be further studied to establish its role in tumorigenesis.

## REFERENCES

- 1 Kadonaga, J. T. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307-313 (1998).
- 2 Griesenbeck, J., Tschochner, H. & Grohmann, D. Structure and Function of RNA Polymerases and the Transcription Machineries. *Subcell Biochem* **83**, 225-270, doi:10.1007/978-3-319-46503-6\_9 (2017).
- 3 Russell, J. & Zomerdijk, J. C. RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci* **30**, 87-96, doi:10.1016/j.tibs.2004.12.008 (2005).
- 4 Chandrasekharappa, S. C., Smith, J. H. & Eliceiri, G. L. Biosynthesis of small nuclear RNAs in human cells. *J Cell Physiol* **117**, 169-174, doi:10.1002/jcp.1041170206 (1983).
- 5 Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060, doi:10.1038/sj.emboj.7600385 (2004).
- 6 Dieci, G., Conti, A., Pagano, A. & Carnevali, D. Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim Biophys Acta* **1829**, 296-305, doi:10.1016/j.bbagr.2012.09.010 (2013).
- 7 Kerppola, T. K. & Kane, C. M. RNA polymerase: regulation of transcript elongation and termination. *FASEB J* **5**, 2833-2842 (1991).
- 8 Murthy, K. G. & Manley, J. L. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**, 2672-2683 (1995).
- 9 Shi, Y. & Manley, J. L. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* **29**, 889-897, doi:10.1101/gad.261974.115 (2015).
- 10 Proudfoot, N. J. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**, 105-110 (1989).
- 11 Riethoven, J. J. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol* **674**, 33-42, doi:10.1007/978-1-60761-854-6\_3 (2010).
- 12 Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5-8 (1994).
- 13 Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**, 449-479, doi:10.1146/annurev.biochem.72.121801.161520 (2003).
- 14 Smale, S. T. & Baltimore, D. The "initiator" as a transcription control element. *Cell* **57**, 103-113 (1989).
- 15 Butler, J. E. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* **16**, 2583-2592, doi:10.1101/gad.1026202 (2002).
- 16 Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the initiator

- over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52-65, doi:10.1016/j.gene.2006.09.029 (2007).
- 17 Flanagan, P. M., Kelleher, R. J., 3rd, Sayre, M. H., Tschochner, H. & Kornberg, R. D. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**, 436-438, doi:10.1038/350436a0 (1991).
- 18 Green, M. R. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci* **25**, 59-63 (2000).
- 19 Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).
- 20 Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**, 394-403, doi:10.1038/nsmb763 (2004).
- 21 Murakami, K. *et al.* Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem* **288**, 6325-6332, doi:10.1074/jbc.M112.433623 (2013).
- 22 Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724, doi:10.1126/science.1238724 (2013).
- 23 Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev* **10**, 2657-2683 (1996).
- 24 Davison, B. L., Egly, J. M., Mulvihill, E. R. & Chambon, P. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature* **301**, 680-686 (1983).
- 25 Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev* **7**, 1291-1308 (1993).
- 26 Dynlacht, B. D., Hoey, T. & Tjian, R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**, 563-576 (1991).
- 27 Tanese, N., Saluja, D., Vassallo, M. F., Chen, J. L. & Admon, A. Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. *Proc Natl Acad Sci U S A* **93**, 13611-13616 (1996).
- 28 Apone, L. M., Virbasius, C. M., Reese, J. C. & Green, M. R. Yeast TAF(II)90 is required for cell-cycle progression through G2/M but not for general transcription activation. *Genes Dev* **10**, 2368-2380 (1996).
- 29 Shen, W. C. *et al.* Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. *EMBO J* **22**, 3395-3402, doi:10.1093/emboj/cdg336 (2003).
- 30 Thomas, M. C. & Chiang, C. M. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**, 105-178,

- doi:10.1080/10409230600648736 (2006).
- 31 Poon, D. *et al.* Identification and characterization of a TFIID-like multiprotein complex from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **92**, 8224-8228 (1995).
- 32 Reese, J. C., Apone, L., Walker, S. S., Griffin, L. A. & Green, M. R. Yeast TAFIIS in a multisubunit complex required for activated transcription. *Nature* **371**, 523-527, doi:10.1038/371523a0 (1994).
- 33 Hoey, T. *et al.* Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell* **72**, 247-260 (1993).
- 34 Kokubo, T. *et al.* *Drosophila* 230-kD TFIID subunit, a functional homolog of the human cell cycle gene product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. *Genes Dev* **7**, 1033-1046 (1993).
- 35 Weinzierl, R. O., Ruppert, S., Dynlacht, B. D., Tanese, N. & Tjian, R. Cloning and expression of *Drosophila* TAFII60 and human TAFII70 reveal conserved interactions with other subunits of TFIID. *EMBO J* **12**, 5303-5309 (1993).
- 36 Chicca, J. J., 2nd, Auble, D. T. & Pugh, B. F. Cloning and biochemical characterization of TAF-172, a human homolog of yeast Mot1. *Mol Cell Biol* **18**, 1701-1710 (1998).
- 37 Ruppert, S., Wang, E. H. & Tjian, R. Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature* **362**, 175-179, doi:10.1038/362175a0 (1993).
- 38 Hisatake, K. *et al.* Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc Natl Acad Sci U S A* **92**, 8195-8199 (1995).
- 39 Ogryzko, V. V. *et al.* Histone-like TAFs within the PCAF histone acetylase complex. *Cell* **94**, 35-44 (1998).
- 40 Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699-702, doi:10.1038/nature11791 (2013).
- 41 Hoffmann, A. *et al.* A histone octamer-like structure within TFIID. *Nature* **380**, 356-359, doi:10.1038/380356a0 (1996).
- 42 Kaufmann, J., Ahrens, K., Koop, R., Smale, S. T. & Muller, R. CIF150, a human cofactor for transcription factor IID-dependent initiator function. *Mol Cell Biol* **18**, 233-239 (1998).
- 43 Verrijzer, C. P., Yokomori, K., Chen, J. L. & Tjian, R. *Drosophila* TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* **264**, 933-941 (1994).
- 44 Chalkley, G. E. & Verrijzer, C. P. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J* **18**, 4835-4845, doi:10.1093/emboj/18.17.4835 (1999).



- 45 Trowitzsch, S. *et al.* Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat Commun* **6**, 6011, doi:10.1038/ncomms7011 (2015).
- 46 Papai, G. *et al.* Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. *Structure* **17**, 363-373, doi:10.1016/j.str.2009.01.006 (2009).
- 47 Martin, J., Halenbeck, R. & Kaufmann, J. Human transcription factor hTAF(II)150 (CIF150) is involved in transcriptional regulation of cell cycle progression. *Mol Cell Biol* **19**, 5548-5556 (1999).
- 48 Walker, S. S., Shen, W. C., Reese, J. C., Apone, L. M. & Green, M. R. Yeast TAF(II)145 required for transcription of G1/S cyclin genes and regulated by the cellular growth state. *Cell* **90**, 607-614 (1997).
- 49 El-Serag, H. B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264-1273 e1261, doi:10.1053/j.gastro.2011.12.061 (2012).
- 50 Pinzani, M., Rosselli, M. & Zuckermann, M. Liver cirrhosis. *Best Pract Res Clin Gastroenterol* **25**, 281-290, doi:10.1016/j.bpg.2011.02.009 (2011).
- 51 Sanyal, A. J., Yoon, S. K. & Lencioni, R. The etiology of hepatocellular carcinoma and consequences for treatment. *Oncologist* **15 Suppl 4**, 14-22, doi:10.1634/theoncologist.2010-S4-14 (2010).
- 52 Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118**, 3030-3044, doi:10.1002/ijc.21731 (2006).
- 53 Ayub, A., Ashfaq, U. A. & Haque, A. HBV induced HCC: major risk factors from genetic to molecular level. *Biomed Res Int* **2013**, 810461, doi:10.1155/2013/810461 (2013).
- 54 Szabo, E., Paska, C., Kaposi Novak, P., Schaff, Z. & Kiss, A. Similarities and differences in hepatitis B and C virus induced hepatocarcinogenesis. *Pathol Oncol Res* **10**, 5-11, doi:PAOR.2004.10.1.0005 (2004).
- 55 Sheikh, M. Y., Choi, J., Qadri, I., Friedman, J. E. & Sanyal, A. J. Hepatitis C virus infection: molecular pathways to metabolic syndrome. *Hepatology* **47**, 2127-2133, doi:10.1002/hep.22269 (2008).
- 56 Hu, J. & Ludgate, L. HIV-HBV and HIV-HCV coinfection and liver cancer development. *Cancer Treat Res* **133**, 241-252 (2007).
- 57 Morgan, T. R., Mandayam, S. & Jamal, M. M. Alcohol and hepatocellular carcinoma. *Gastroenterology* **127**, S87-96 (2004).
- 58 Gholam, P. M., Flancbaum, L., Machan, J. T., Charney, D. A. & Kotler, D. P. Nonalcoholic fatty liver disease in severely obese subjects. *Am J Gastroenterol* **102**, 399-408, doi:10.1111/j.1572-0241.2006.01041.x (2007).
- 59 Brunt, E. M. Nonalcoholic steatohepatitis. *Semin Liver Dis* **24**, 3-20, doi:10.1055/s-2004-823098 (2004).
- 60 Montella, M., Crispo, A. & Giudice, A. HCC, diet and metabolic factors: Diet and HCC.

- Hepat Mon* **11**, 159-162 (2011).
- 61 Nouredin, M. & Rinella, M. E. Nonalcoholic Fatty liver disease, diabetes, obesity, and hepatocellular carcinoma. *Clin Liver Dis* **19**, 361-379, doi:10.1016/j.cld.2015.01.012 (2015).
- 62 VoPham, T. *et al.* Pesticide exposure and hepatocellular carcinoma risk: A case-control study using a geographic information system (GIS) to link SEER-Medicare and California pesticide data. *Environ Res* **143**, 68-82, doi:10.1016/j.envres.2015.09.027 (2015).
- 63 Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**, 2893-2917, doi:10.1002/ijc.25516 (2010).
- 64 Kao, J. H. & Chen, D. S. Changing disease burden of hepatocellular carcinoma in the Far East and Southeast Asia. *Liver Int* **25**, 696-703, doi:10.1111/j.1478-3231.2005.01139.x (2005).
- 65 Chang, M. H. *et al.* Decreased incidence of hepatocellular carcinoma in hepatitis B vaccinees: a 20-year follow-up study. *J Natl Cancer Inst* **101**, 1348-1355, doi:10.1093/jnci/djp288 (2009).
- 66 Altekruse, S. F., McGlynn, K. A. & Reichman, M. E. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. *J Clin Oncol* **27**, 1485-1491, doi:10.1200/JCO.2008.20.7753 (2009).
- 67 El-Serag, H. B. & Mason, A. C. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med* **340**, 745-750, doi:10.1056/NEJM199903113401001 (1999).
- 68 El-Serag, H. B., Davila, J. A., Petersen, N. J. & McGlynn, K. A. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med* **139**, 817-823 (2003).
- 69 El-Serag, H. B. & Mason, A. C. Risk factors for the rising rates of primary liver cancer in the United States. *Arch Intern Med* **160**, 3227-3230 (2000).
- 70 Mazzaferro, V. *et al.* Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* **334**, 693-699, doi:10.1056/NEJM199603143341104 (1996).
- 71 Bhoori, S., Schiavo, M., Russo, A. & Mazzaferro, V. First-line treatment for hepatocellular carcinoma: resection or transplantation? *Transplant Proc* **39**, 2271-2273, doi:10.1016/j.transproceed.2007.06.015 (2007).
- 72 Bruix, J. *et al.* Surgical resection of hepatocellular carcinoma in cirrhotic patients: prognostic value of preoperative portal pressure. *Gastroenterology* **111**, 1018-1022 (1996).
- 73 Nishikawa, H., Kimura, T., Kita, R. & Osaki, Y. Radiofrequency ablation for hepatocellular carcinoma. *Int J Hyperthermia* **29**, 558-568, doi:10.3109/02656736.2013.821528 (2013).

- 74 Forner, A., Llovet, J. M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **379**, 1245-1255, doi:10.1016/S0140-6736(11)61347-0 (2012).
- 75 Cheng, A. L. *et al.* Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol* **10**, 25-34, doi:10.1016/S1470-2045(08)70285-7 (2009).
- 76 Llovet, J. M. *et al.* Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* **359**, 378-390, doi:10.1056/NEJMoa0708857 (2008).
- 77 Matsui, A., Ihara, T., Suda, H., Mikami, H. & Semba, K. Gene amplification: mechanisms and involvement in cancer. *Biomol Concepts* **4**, 567-582, doi:10.1515/bmc-2013-0026 (2013).
- 78 Escot, C. *et al.* Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. *Proc Natl Acad Sci U S A* **83**, 4834-4838 (1986).
- 79 Garcia, I. *et al.* Genetic alterations of c-myc, c-erbB-2, and c-Ha-ras protooncogenes and clinical associations in human breast carcinomas. *Cancer Res* **49**, 6675-6679 (1989).
- 80 Guerin, M., Barrois, M., Terrier, M. J., Spielmann, M. & Riou, G. Overexpression of either c-myc or c-erbB-2/neu proto-oncogenes in human breast carcinomas: correlation with poor prognosis. *Oncogene Res* **3**, 21-31 (1988).
- 81 Abi-Ayad, N. *et al.* [Genomic profiling by comparative genomic hybridization: analysis of ten enucleated uveal melanoma cases]. *J Fr Ophthalmol* **34**, 17-23, doi:10.1016/j.jfo.2010.10.013 (2011).
- 82 Buffart, T. E. *et al.* High resolution analysis of DNA copy-number aberrations of chromosomes 8, 13, and 20 in gastric cancers. *Virchows Arch* **455**, 213-223, doi:10.1007/s00428-009-0814-y (2009).
- 83 de Krijger, R. R. *et al.* Gain of chromosome 8q is a frequent finding in pleuropulmonary blastoma. *Mod Pathol* **20**, 1191-1199, doi:10.1038/modpathol.3800953 (2007).
- 84 Kang, J. U. Chromosome 8q as the most frequent target for amplification in early gastric carcinoma. *Oncol Lett* **7**, 1139-1143, doi:10.3892/ol.2014.1849 (2014).
- 85 Klatte, T. *et al.* Gain of chromosome 8q is associated with metastases and poor survival of patients with clear cell renal cell carcinoma. *Cancer* **118**, 5777-5782, doi:10.1002/cncr.27607 (2012).
- 86 El Gammal, A. T. *et al.* Chromosome 8p deletions and 8q gains are associated with tumor progression and poor prognosis in prostate cancer. *Clin Cancer Res* **16**, 56-64, doi:10.1158/1078-0432.CCR-09-1423 (2010).
- 87 Domagk, D. *et al.* Expression analysis of pancreatic cancer cell lines reveals association of enhanced gene transcription and genomic amplifications at the 8q22.1 and 8q24.22 loci. *Oncol Rep* **17**, 399-407 (2007).

- 88 Baltaci, E., Karaman, E., Dalay, N. & Buyru, N. Analysis of gene copy number changes in head and neck cancer. *Clin Otolaryngol*, doi:10.1111/coa.12686 (2016).
- 89 Kornegoor, R. *et al.* Oncogene amplification in male breast cancer: analysis by multiplex ligation-dependent probe amplification. *Breast Cancer Res Treat* **135**, 49-58, doi:10.1007/s10549-012-2051-3 (2012).
- 90 Moelans, C. B. *et al.* Genomic evolution from primary breast carcinoma to distant metastasis: Few copy number changes of breast cancer related genes. *Cancer Lett* **344**, 138-146, doi:10.1016/j.canlet.2013.10.025 (2014).
- 91 Smith, S. C. *et al.* CIC-DUX sarcomas demonstrate frequent MYC amplification and ETS-family transcription factor expression. *Mod Pathol* **28**, 57-68, doi:10.1038/modpathol.2014.83 (2015).
- 92 Wang, K. *et al.* Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* **58**, 706-717, doi:10.1002/hep.26402 (2013).
- 93 Guo, F. *et al.* Astrocyte elevated gene-1 overexpression in histologically favorable Wilms tumor is related to poor prognosis. *J Pediatr Urol* **10**, 317-323, doi:10.1016/j.jpuro.2013.09.010 (2014).
- 94 Jian-bo, X. *et al.* Astrocyte-elevated gene-1 overexpression is associated with poor prognosis in gastric cancer. *Med Oncol* **28**, 455-462, doi:10.1007/s12032-010-9475-6 (2011).
- 95 Ke, Z. F. *et al.* Expression characteristics of astrocyte elevated gene-1 (AEG-1) in tongue carcinoma and its correlation with poor prognosis. *Cancer Epidemiol* **37**, 179-185, doi:10.1016/j.canep.2012.10.007 (2013).
- 96 Li, W. F., Wang, G., Zhao, Z. B. & Liu, C. A. High expression of metadherin correlates with malignant pathological features and poor prognostic significance in papillary thyroid carcinoma. *Clin Endocrinol (Oxf)* **83**, 572-580, doi:10.1111/cen.12683 (2015).
- 97 Lee, S. G., Su, Z. Z., Emdad, L., Sarkar, D. & Fisher, P. B. Astrocyte elevated gene-1 (AEG-1) is a target gene of oncogenic Ha-ras requiring phosphatidylinositol 3-kinase and c-Myc. *Proc Natl Acad Sci U S A* **103**, 17390-17395, doi:10.1073/pnas.0608386103 (2006).
- 98 He, X. X. *et al.* MicroRNA-375 targets AEG-1 in hepatocellular carcinoma and suppresses liver cancer cell growth in vitro and in vivo. *Oncogene* **31**, 3357-3369, doi:10.1038/onc.2011.500 (2012).
- 99 Hu, G., Wei, Y. & Kang, Y. The multifaceted role of MTDH/AEG-1 in cancer progression. *Clin Cancer Res* **15**, 5615-5620, doi:10.1158/1078-0432.CCR-09-0049 (2009).
- 100 Ribeiro, J. R., Lovasco, L. A., Vanderhyden, B. C. & Freiman, R. N. Targeting TBP-Associated Factors in Ovarian Cancer. *Front Oncol* **4**, 45,

- doi:10.3389/fonc.2014.00045 (2014).
- 101 Parris, T. Z. *et al.* Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. *Oncogenesis* **3**, e95, doi:10.1038/oncsis.2014.8 (2014).
- 102 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
- 103 McClatchey, A. I. & Yap, A. S. Contact inhibition (of proliferation) redux. *Curr Opin Cell Biol* **24**, 685-694, doi:10.1016/j.ceb.2012.06.009 (2012).
- 104 Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol Pathol* **35**, 495-516, doi:10.1080/01926230701320337 (2007).
- 105 Greenberg, R. A. Telomeres, crisis and cancer. *Curr Mol Med* **5**, 213-218 (2005).
- 106 Bergers, G. & Benjamin, L. E. Tumorigenesis and the angiogenic switch. *Nat Rev Cancer* **3**, 401-410, doi:10.1038/nrc1093 (2003).
- 107 van Zijl, F., Krupitza, G. & Mikulits, W. Initial steps of metastasis: cell invasion and endothelial transmigration. *Mutat Res* **728**, 23-34, doi:10.1016/j.mrrev.2011.05.002 (2011).
- 108 Alizadeh, A. M., Shiri, S. & Farsinejad, S. Metastasis review: from bench to bedside. *Tumour Biol* **35**, 8483-8523, doi:10.1007/s13277-014-2421-z (2014).
- 109 McGee, S. F., Lanigan, F., Gilligan, E. & Groner, B. Mammary gland biology and breast cancer. Conference on Common Molecular Mechanisms of Mammary Gland Development and Breast Cancer Progression. *EMBO Rep* **7**, 1084-1088, doi:10.1038/sj.embor.7400839 (2006).
- 110 Adams, J. C. Cell-matrix contact structures. *Cell Mol Life Sci* **58**, 371-392, doi:10.1007/PL00000864 (2001).
- 111 Weber, G. F., Bjerke, M. A. & DeSimone, D. W. Integrins and cadherins join forces to form adhesive networks. *J Cell Sci* **124**, 1183-1193, doi:10.1242/jcs.064618 (2011).
- 112 Mierke, C. T., Rosel, D., Fabry, B. & Brabek, J. Contractile forces in tumor cell migration. *Eur J Cell Biol* **87**, 669-676, doi:10.1016/j.ejcb.2008.01.002 (2008).
- 113 Stetler-Stevenson, W. G., Aznavoorian, S. & Liotta, L. A. Tumor cell interactions with the extracellular matrix during invasion and metastasis. *Annu Rev Cell Biol* **9**, 541-573, doi:10.1146/annurev.cb.09.110193.002545 (1993).
- 114 Bozzuto, G., Ruggieri, P. & Molinari, A. Molecular aspects of tumor cell migration and invasion. *Ann Ist Super Sanita* **46**, 66-80, doi:10.4415/ANN\_10\_01\_09 (2010).
- 115 Kalluri, R. & Neilson, E. G. Epithelial-mesenchymal transition and its implications for fibrosis. *J Clin Invest* **112**, 1776-1784, doi:10.1172/JCI20530 (2003).
- 116 Steinestel, K., Eder, S., Schrader, A. J. & Steinestel, J. Clinical significance of epithelial-mesenchymal transition. *Clin Transl Med* **3**, 17, doi:10.1186/2001-1326-3-17 (2014).
- 117 Zhang, H. *et al.* KLF8 involves in TGF-beta-induced EMT and promotes invasion and

- migration in gastric cancer cells. *J Cancer Res Clin Oncol* **139**, 1033-1042, doi:10.1007/s00432-012-1363-3 (2013).
- 118 Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).
- 119 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196, doi:10.1038/nrm3758 (2014).
- 120 Hazan, R. B., Qiao, R., Keren, R., Badano, I. & Suyama, K. Cadherin switch in tumor progression. *Ann N Y Acad Sci* **1014**, 155-163 (2004).
- 121 Zhou, H. & Huang, S. Role of mTOR signaling in tumor cell motility, invasion and metastasis. *Curr Protein Pept Sci* **12**, 30-42 (2011).
- 122 Agarwal, N. K. *et al.* Rictor regulates cell migration by suppressing RhoGDI2. *Oncogene* **32**, 2521-2526, doi:10.1038/onc.2012.287 (2013).
- 123 Vitorino, P. *et al.* MAP4K4 regulates integrin-FERM binding to control endothelial cell motility. *Nature* **519**, 425-430, doi:10.1038/nature14323 (2015).
- 124 Jiang, W. G. *et al.* Hepatocyte growth factor, its receptor, and their potential value in cancer therapies. *Crit Rev Oncol Hematol* **53**, 35-69, doi:10.1016/j.critrevonc.2004.09.004 (2005).
- 125 Oommen, S., Gupta, S. K. & Vlahakis, N. E. Vascular endothelial growth factor A (VEGF-A) induces endothelial and cancer cell migration through direct binding to integrin  $\alpha_9\beta_1$ : identification of a specific  $\alpha_9\beta_1$  binding site. *J Biol Chem* **286**, 1083-1092, doi:10.1074/jbc.M110.175158 (2011).

## **VITA**

Saranya Chidambaranathan Reghupaty was born in Tamilnadu, India on the 17<sup>th</sup> of August 1992. She graduated high school in 2008. She received her Bachelor of Technology (B.Tech) degree in Biotechnology from Sathyabama University, Tamilnadu, India in 2014. She worked as an intern in Rajiv Gandhi Center for Biotechnology in Kerala, India for six months before starting Graduate school at Virginia Commonwealth University in Fall 2015.